Cours d'analyse numérique

Laurent Seppecher & Grégory Vial

École Centrale de Lyon

Année 2022-2023 – Semestre 5

ÉCOLE CENTRALE LYON

Informations générales

- Monitorat (romain.meaux@ecl21.ec-lyon.fr)

Informations générales

- Monitorat (romain.meaux@ecl21.ec-lyon.fr)
- Cours : pré-requis à regarder avant.

Informations générales

- Monitorat (romain.meaux@ecl21.ec-lyon.fr)
- Cours : pré-requis à regarder avant.
- TD : à préparer ! Utilisation des ordinateurs portables (Matlab).

Informations générales

- Monitorat (romain.meaux@ecl21.ec-lyon.fr)
- Cours : pré-requis à regarder avant.
- TD : à préparer ! Utilisation des ordinateurs portables (Matlab).
- Un tutoriel Matlab succinct est disponible (et à apprendre par cœur)

Informations générales

- Monitorat (romain.meaux@ecl21.ec-lyon.fr)
- Cours : pré-requis à regarder avant.
- TD : à préparer ! Utilisation des ordinateurs portables (Matlab).
- Un tutoriel Matlab succinct est disponible (et à apprendre par cœur)
- Contrôle des connaissances :

Informations générales

- Monitorat (romain.meaux@ecl21.ec-lyon.fr)
- Cours : pré-requis à regarder avant.
- TD : à préparer ! Utilisation des ordinateurs portables (Matlab).
- Un tutoriel Matlab succinct est disponible (et à apprendre par cœur)
- Contrôle des connaissances :
  - Savoir-faire 25% : mini-test (1H lors du TD4) à confirmer

    Autorisés : documents cours/TD, programmes matlab.
  - Savoir 75% : examen terminal (1H30).

    Autorisé : un A4 recto-verso manuscrit.

Introduction

Plan du cours

- **Cours 1.** Systèmes linéaires et calcul de valeurs propres
- **Cours 2.** Interpolation et intégration numérique
- **Cours 3.** Optimisation numérique
- **Cours 4.** Approximation numérique des équations différentielles
- **Cours 5.** Discrétisation des EDP : l'équation de Laplace
- **Cours 6.** Discrétisation des EDP : l'équation de transport

Un conseil

Un conseil

« On ne peut pas faire de physique sans un bon niveau en mathématiques. Mais c'est vrai quelle que soit la science. Je vais même être caricatural : si vous voulez trouver du travail facilement, faites des maths. »

Alain Aspect

COURS 1

Systèmes linéaires

Calcul de valeurs propres

Le modèle de Léontiev en comptabilité nationale (version production)

- d branches : B_i produit P_i.
- 1 bien P_j nécessite a_{ij} biens P_i,
- B_i produit une quantité q_i.

- d branches : B_i produit P_i.
- 1 bien P_j nécessite a_{ij} biens P_i,
- B_i produit une quantité q_i.

Bilan pour la branche B_i :

$$q_i - \sum_{j=1}^{d} a_{ij} q_j = (Mq)_i \quad \text{avec} \quad M = I - A.$$

Le modèle de Léontiev en comptabilité nationale (version production)

- d branches : B_i produit P_i.
- 1 bien P_j nécessite a_{ij} biens P_i,
- B_i produit une quantité q_i.

Bilan pour la branche B_i :

$$q_i - \sum_{j=1}^{d} a_{ij} q_j = (Mq)_i \quad \text{avec} \quad M = I - A.$$

Équilibre offre-demande :

$$Mq = \delta.$$

Le modèle de Léontiev en comptabilité nationale (version production)

- d branches : B_i produit P_i.
- 1 bien P_j nécessite a_{ij} biens P_i,
- B_i produit une quantité q_i.

Bilan pour la branche B_i :

$$q_i - \sum_{j=1}^{d} a_{ij} q_j = (Mq)_i \quad \text{avec} \quad M = I - A.$$

Équilibre offre-demande :

$$Mq = \delta.$$

Question : quelle influence de la demande sur les quantités produites ?

$$\boxed{\Longrightarrow \text{ système linéaire } Mq = \delta.}$$

Le modèle de Léontiev en comptabilité nationale (version bénéfice)

- d branches : B_i produit P_i.
- 1 bien P_i nécessite
  - a_{ji} biens P_j,
  - ℓ_i travailleurs au salaire w.
- B_i vend au prix π_i.
- salaire fonction linéaire des prix : $w = \sum c_j \pi_j$.

Le modèle de Léontiev en comptabilité nationale (version bénéfice)

- d branches : B_i produit P_i.
- 1 bien P_i nécessite
  - a_{ji} biens P_j,
  - ℓ_i travailleurs au salaire w.
- B_i vend au prix π_i.
- salaire fonction linéaire des prix : $w = \sum c_j \pi_j$.

Coût de production de la branche B_i :

$$\sum_{j=1}^{d} a_{ji}\pi_j + \ell_i w = \sum_{j=1}^{d} (a_{ji} + \ell_i c_j)\pi_j = (N\pi)_i \quad \text{avec} \quad N = A^T + \ell\, c^T.$$

Le modèle de Léontiev en comptabilité nationale (version bénéfice)

- d branches : B_i produit P_i.
- 1 bien P_i nécessite
  - a_{ji} biens P_j,
  - ℓ_i travailleurs au salaire w.
- B_i vend au prix π_i.
- salaire fonction linéaire des prix : $w = \sum c_j \pi_j$.

Coût de production de la branche B_i :

$$\sum_{j=1}^{d} a_{ji}\pi_j + \ell_i w = \sum_{j=1}^{d} (a_{ji} + \ell_i c_j)\pi_j = (N\pi)_i \quad \text{avec} \quad N = A^T + \ell\, c^T.$$

Profit relatif de la branche B_i : $\tau_i = \dfrac{\pi_i - (N\pi)_i}{(N\pi)_i}$.

Le modèle de Léontiev en comptabilité nationale (version bénéfice)

- d branches : B_i produit P_i.
- 1 bien P_i nécessite
  - a_{ji} biens P_j,
  - ℓ_i travailleurs au salaire w.
- B_i vend au prix π_i.
- salaire fonction linéaire des prix : $w = \sum c_j \pi_j$.

Coût de production de la branche B_i :

$$\sum_{j=1}^{d} a_{ji}\pi_j + \ell_i w = \sum_{j=1}^{d} (a_{ji} + \ell_i c_j)\pi_j = (N\pi)_i \quad \text{avec} \quad N = A^T + \ell\, c^T.$$

Profit relatif de la branche B_i : $\tau_i = \dfrac{\pi_i - (N\pi)_i}{(N\pi)_i}$.

Question : profit équilibré parmi les branches ?

Le modèle de Léontiev en comptabilité nationale (version bénéfice)

- d branches : B_i produit P_i.
- 1 bien P_i nécessite
  - a_{ji} biens P_j,
  - ℓ_i travailleurs au salaire w.
- B_i vend au prix π_i.
- salaire fonction linéaire des prix : $w = \sum c_j \pi_j$.

Coût de production de la branche B_i :

$$\sum_{j=1}^{d} a_{ji}\pi_j + \ell_i w = \sum_{j=1}^{d} (a_{ji} + \ell_i c_j)\pi_j = (N\pi)_i \quad \text{avec} \quad N = A^T + \ell\, c^T.$$

Profit relatif de la branche B_i : $\tau_i = \dfrac{\pi_i - (N\pi)_i}{(N\pi)_i}$.

Question : profit équilibré parmi les branches ?

$$\boxed{\Longrightarrow \text{Problème aux valeurs propres } N\pi = \frac{1}{1+\tau}\pi}$$

Analyse numérique matricielle

Résolution de systèmes linéaires

- En théorie... Soit A est inversible. $\exists! x$ t.q. $Ax = b$. On a $x = A^{-1}b$.

Analyse numérique matricielle

Résolution de systèmes linéaires

- En théorie... Soit A est inversible. $\exists! x$ t.q. $Ax = b$. On a $x = A^{-1}b$.
- En pratique...
  - Soit $H \in \mathbb{R}^{d \times d}$ t.q. $H_{ij} = \frac{1}{i+j-1}$. [*hilbert.m*]

    Rq. H est inversible et $H^{-1} \in \mathbb{Z}^{d \times d}$ *(formule explicite)*.

- En théorie... Soit A est inversible. $\exists! x$ t.q. $Ax = b$. On a $x = A^{-1}b$.
- En pratique...
  - Soit $H \in \mathbb{R}^{d \times d}$ t.q. $H_{ij} = \frac{1}{i+j-1}$. *[hilbert.m]*

    Rq. H est inversible et $H^{-1} \in \mathbb{Z}^{d \times d}$ *(formule explicite)*.

    $\Longrightarrow$ **comment expliquer le phénomène ?**

- En théorie. . . Soit A est inversible. $\exists ! x$ t.q. $Ax = b$. On a $x = A^{-1}b$.
- En pratique. . .
  - Soit $H \in \mathbb{R}^{d \times d}$ t.q. $H_{ij} = \frac{1}{i+j-1}$. [*hilbert.m*]

    Rq. H est inversible et $H^{-1} \in \mathbb{Z}^{d \times d}$ *(formule explicite)*.

    $\Longrightarrow$ **comment expliquer le phénomène ?**

  - Formules de Cramer :

$$x_i = \frac{\det(A_i)}{\det(A)}, \quad \text{avec } A_i = \big(C_1 | \dots | C_{i-1} | b | C_{i+1} | \dots | C_d\big).$$

Analyse numérique matricielle

Résolution de systèmes linéaires

- En théorie... Soit A est inversible. $\exists! x$ t.q. $Ax = b$. On a $x = A^{-1}b$.
- En pratique...
  - Soit $H \in \mathbb{R}^{d\times d}$ t.q. $H_{ij} = \frac{1}{i+j-1}$. [*hilbert.m*]

    Rq. H est inversible et $H^{-1} \in \mathbb{Z}^{d\times d}$ *(formule explicite)*.

    $\Longrightarrow$ **comment expliquer le phénomène ?**
  - Formules de Cramer :

    $$x_i = \frac{\det(A_i)}{\det(A)}, \quad \text{avec } A_i = \big(C_1|\ldots|C_{i-1}|b|C_{i+1}|\ldots|C_d\big).$$

    Coût si les det sont calculés par dvpt $^p/_r$ lignes : $\simeq (d+1)!$.

- En théorie... Soit A est inversible. $\exists! x$ t.q. $Ax = b$. On a $x = A^{-1}b$.
- En pratique...
  - Soit $H \in \mathbb{R}^{d\times d}$ t.q. $H_{ij} = \frac{1}{i+j-1}$. [*hilbert.m*]

    Rq. H est inversible et $H^{-1} \in \mathbb{Z}^{d\times d}$ *(formule explicite)*.

    $\Longrightarrow$ **comment expliquer le phénomène ?**
  - Formules de Cramer :

    $$x_i = \frac{\det(A_i)}{\det(A)}, \quad \text{avec } A_i = \left(C_1|\dots|C_{i-1}|b|C_{i+1}|\dots|C_d\right).$$

    Coût si les det sont calculés par dvpt $^p/_r$ lignes : $\simeq (d+1)!$.

    $\Longrightarrow$ **quels algorithmes ? quelles performances ?**

Question : comment quantifier la propagation des erreurs ?

Question : comment quantifier la propagation des erreurs ?

Proposition. Soient $A \in \mathbb{R}^{d \times d}$ inversible, $b, \delta b \in \mathbb{R}^d$.
On note $x \in \mathbb{R}^d$ et $x + \delta x \in \mathbb{R}^d$ les solutions de

$$Ax = b, \qquad A(x + \delta x) = b + \delta b.$$

Alors

$$\frac{\|\delta x\|}{\|x\|} \leqslant \mathrm{cond}(A) \frac{\|\delta b\|}{\|b\|},$$

avec $\mathrm{cond}(A) = |||A||| \times |||A^{-1}|||$.

Question : comment quantifier la propagation des erreurs ?

Proposition. Soient $A \in \mathbb{R}^{d\times d}$ inversible, $b, \delta b \in \mathbb{R}^d$.
On note $x \in \mathbb{R}^d$ et $x + \delta x \in \mathbb{R}^d$ les solutions de

$$Ax = b, \qquad A(x + \delta x) = b + \delta b.$$

Alors

$$\frac{\|\delta x\|}{\|x\|} \leqslant \mathrm{cond}(A)\frac{\|\delta b\|}{\|b\|},$$

avec $\mathrm{cond}(A) = |||A||| \times |||A^{-1}|||$.

- $|||A|||$ est la norme subordonnée à $\|x\|$.
- $\mathrm{cond}(A)$ mesure la propagation des erreurs relatives.
- Si $\|\cdot\| = \|\cdot\|_p$, on note cond_p.

Question : comment quantifier la propagation des erreurs ?

Proposition. Soient $A \in \mathbb{R}^{d\times d}$ inversible, $b, \delta b \in \mathbb{R}^d$.
On note $x \in \mathbb{R}^d$ et $x + \delta x \in \mathbb{R}^d$ les solutions de

$$Ax = b, \qquad A(x + \delta x) = b + \delta b.$$

Alors

$$\frac{\|\delta x\|}{\|x\|} \leqslant \mathrm{cond}(A)\frac{\|\delta b\|}{\|b\|},$$

avec $\mathrm{cond}(A) = |||A||| \times |||A^{-1}|||$.

Preuve. Par différence $A(\delta x) = \delta b$ d'où $\|\delta x\| \leqslant |||A^{-1}||| \times \|\delta b\|$.

Par ailleurs $b = Ax$ d'où $\|b\| \leqslant |||A||| \times \|x\|$, soit $\dfrac{1}{\|x\|} \leqslant \dfrac{|||A|||}{\|b\|}$.

D'où le résultat par produit. $\square$

Proposition. Soit $A \in \mathbb{R}^{d\times d}$ inversible.

- $\forall \|\cdot\|, \quad \mathrm{cond}(A) \geqslant 1.$
- Si A est symétrique,
$$\mathrm{cond}_2(A) = \frac{\lambda_{\max}(A)}{\lambda_{\min}(A)}.$$
- Si A est orthogonale, $\mathrm{cond}_2(A) = 1$.

Proposition. Soit $A \in \mathbb{R}^{d\times d}$ inversible.

- $\forall \|\cdot\|, \quad \mathrm{cond}(A) \geqslant 1.$
- Si A est symétrique,
$$\mathrm{cond}_2(A) = \frac{\lambda_{\max}(A)}{\lambda_{\min}(A)}.$$
- Si A est orthogonale, $\mathrm{cond}_2(A) = 1$.

Rmq. Si $\mathrm{cond}(A)$ est « grand », on dit que A est *mal conditionnée*.

Proposition. Soit $A \in \mathbb{R}^{d\times d}$ inversible.

- $\forall \|\cdot\|, \quad \mathrm{cond}(A) \geqslant 1$.
- Si A est symétrique,
$$\mathrm{cond}_2(A) = \frac{\lambda_{\max}(A)}{\lambda_{\min}(A)}.$$
- Si A est orthogonale, $\mathrm{cond}_2(A) = 1$.

Rmq. Si $\mathrm{cond}(A)$ est « grand », on dit que A est *mal conditionnée*.

d	3	5	10	20
$\mathrm{cond}_2(H)$	524	4.8×10^5	1.6×10^{13}	2.5×10^{18}

Table – *Conditionnement de la matrice de Hilbert.*

- Principe.
  - éliminer les inconnues successivement.
  - on aboutit à un système triangulaire.

- Principe.
  - éliminer les inconnues successivement.
  - on about à un système triangulaire.
- Description.

- Principe.
  - éliminer les inconnues successivement.
  - on aboutit à un système triangulaire.
- Description.

Initialisation

$$\begin{array}{|cccc|c}
a_{1,1} & a_{1,2} & \dots & a_{1,d} & b_1 \\
a_{2,1} & a_{2,2} & \dots & a_{2,d} & b_2 \\
\vdots & & & & \vdots \\
a_{d,1} & a_{d,2} & \dots & a_{d,d} & b_d
\end{array}$$

- Principe.
  - éliminer les inconnues successivement.
  - on aboutit à un système triangulaire.
- Description.

  Initialisation

$$
\left|\begin{array}{cccc|c}
a_{1,1}^{[0]} & a_{1,2}^{[0]} & \dots & a_{1,d}^{[0]} & b_1^{[0]} \\
a_{2,1}^{[0]} & a_{2,2}^{[0]} & \dots & a_{2,d}^{[0]} & b_2^{[0]} \\
\vdots & & & & \vdots \\
a_{d,1}^{[0]} & a_{d,2}^{[0]} & \dots & a_{d,d}^{[0]} & b_d^{[0]}
\end{array}\right.
$$

- Principe.
  - éliminer les inconnues successivement.
  - on aboutit à un système triangulaire.
- Description.

Étape 1.1 : si $a_{1,1}^{[0]} = 0$, on permute $\boxed{L_1 \leftrightarrow L_k}$ t.q. $a_{k,1}^{[0]} \neq 0$.

$$\left|\begin{array}{cccc} a_{1,1}^{[0]} & a_{1,2}^{[0]} & \dots & a_{1,d}^{[0]} \\ a_{2,1}^{[0]} & a_{2,2}^{[0]} & \dots & a_{2,d}^{[0]} \\ \vdots & & & \\ a_{d,1}^{[0]} & a_{d,2}^{[0]} & \dots & a_{d,d}^{[0]} \end{array}\right| \begin{array}{c} b_1^{[0]} \\ b_2^{[0]} \\ \vdots \\ b_d^{[0]} \end{array}$$

- Principe.
  - éliminer les inconnues successivement.
  - on aboutit à un système triangulaire.
- Description.

Étape 1.2 : $\tilde{a}_{1,1}^{[0]} \neq 0$; on effectue $\boxed{L_k \leftarrow L_k - r_k L_1}$ avec $r_k = \frac{\tilde{a}_{k,1}^{[0]}}{\tilde{a}_{1,1}^{[0]}}$.

$$\left|\begin{array}{cccc} \tilde{a}_{1,1}^{[0]} & \tilde{a}_{1,2}^{[0]} & \dots & \tilde{a}_{1,d}^{[0]} \\ \tilde{a}_{2,1}^{[0]} & \tilde{a}_{2,2}^{[0]} & \dots & \tilde{a}_{2,d}^{[0]} \\ \vdots & & & \\ \tilde{a}_{d,1}^{[0]} & \tilde{a}_{d,2}^{[0]} & \dots & \tilde{a}_{d,d}^{[0]} \end{array}\right| \begin{array}{c} \tilde{b}_{1}^{[0]} \\ \tilde{b}_{2}^{[0]} \\ \vdots \\ \tilde{b}_{d}^{[0]} \end{array}$$

- Principe.
  - éliminer les inconnues successivement.
  - on aboutit à un système triangulaire.
- Description.

Étape 1.2 : $\tilde{a}_{1,1}^{[0]} \neq 0$; on effectue $\boxed{L_k \leftarrow L_k - r_k L_1}$ avec $r_k = \frac{\tilde{a}_{k,1}^{[0]}}{\tilde{a}_{1,1}^{[0]}}$.

$$\left|\begin{array}{cccc} \tilde{a}_{1,1}^{[0]} & \tilde{a}_{1,2}^{[0]} & \dots & \tilde{a}_{1,d}^{[0]} \\ 0 & a_{2,2}^{[1]} & \dots & a_{2,d}^{[1]} \\ \vdots & & & \\ 0 & a_{d,2}^{[1]} & \dots & a_{d,d}^{[1]} \end{array}\right| \begin{array}{c} \tilde{b}_1^{[0]} \\ b_2^{[1]} \\ \vdots \\ b_d^{[1]} \end{array}$$

- Principe.
  - éliminer les inconnues successivement.
  - on aboutit à un système triangulaire.
- Description.

  Étape 2 : On recommence avec la sous-matrice !

$$
\left|\begin{array}{cccc}
\tilde{a}_{1,1}^{[0]} & \tilde{a}_{1,2}^{[0]} & \dots & \tilde{a}_{1,d}^{[0]} \\
0 & a_{2,2}^{[1]} & \dots & a_{2,d}^{[1]} \\
0 & \vdots & & \\
\vdots & & & \\
0 & a_{d,2}^{[1]} & \dots & a_{d,d}^{[1]}
\end{array}\right|
\quad
\left|\begin{array}{c}
\tilde{b}_1^{[0]} \\
b_2^{[1]} \\
\vdots \\
\\
b_d^{[1]}
\end{array}\right.
$$

- Principe.
  - éliminer les inconnues successivement.
  - on aboutit à un système triangulaire.
- Synthèse.
  - Après $d-1$ étapes, on est ramené à $Uy=\bar{c}$ avec U tri. sup.
  - $Uy=\bar{c}$ est résolu par « remontée ».

- Principe.
  - éliminer les inconnues successivement.
  - on aboutit à un système triangulaire.
- Synthèse.
  - Après $d-1$ étapes, on est ramené à $Uy = \bar{c}$ avec U tri. sup.
  - $Uy = \bar{c}$ est résolu par « remontée ».
- Coût de calcul.
  - L'étape k affecte $(d-k)^2$ coefficients de la matrice.
  - Coût total $\simeq \sum_{k=1}^{d-1}(d-k)^2 = \mathcal{O}(d^3)$.
  - *Les opérations sur le second membre et la remontée sont négligeables.*

- Interprétation matricielle des opérations élémentaires
  - Permutation $\boxed{L_i \leftrightarrow L_k} \iff \boxed{A \mapsto PA}$
  - Transvection $\boxed{L_k \leftarrow L_k - rL_i} \iff \boxed{A \mapsto TA}$ (pour $k > i$)

$$P = \begin{bmatrix} 1 & & & & & \\ & \ddots & & & & \\ & & 0 & & 1 & \\ & & & \ddots & & \\ & & 1 & & 0 & \\ & & & & & \ddots & \\ & & & & & & 1 \end{bmatrix}, \qquad T = \begin{bmatrix} 1 & & & & & \\ & \ddots & & & & \\ & & 1 & & & \\ & & & \ddots & & \\ & & -r & & 1 & \\ & & & & & \ddots & \\ & & & & & & 1 \end{bmatrix}$$

- Interprétation matricielle des opérations élémentaires
  - Permutation $\boxed{L_i \leftrightarrow L_k} \iff \boxed{A \mapsto PA}$
  - Transvection $\boxed{L_k \leftarrow L_k - rL_i} \iff \boxed{A \mapsto TA}$ *(pour $k > i$)*

$$P = \begin{bmatrix} 1 & & & & & \\ & \ddots & & & & \\ & & 0 & & 1 & & \\ & & & \ddots & & & \\ & & 1 & & 0 & & \\ & & & & & \ddots & \\ & & & & & & 1 \end{bmatrix}, \qquad T = \begin{bmatrix} 1 & & & & & \\ & \ddots & & & & \\ & & 1 & & & & \\ & & & \ddots & & & \\ & & -r & & 1 & & \\ & & & & & \ddots & \\ & & & & & & 1 \end{bmatrix}$$

- Écriture matricielle de l'algorithme de Gauss

$\boxed{A \longmapsto M_1 \times M_2 \times \cdots \times M_q A = U}$ avec les M_p de type P ou T.

- Interprétation matricielle des opérations élémentaires
  - Permutation $\boxed{L_i \leftrightarrow L_k} \iff \boxed{A \mapsto PA}$
  - Transvection $\boxed{L_k \leftarrow L_k - rL_i} \iff \boxed{A \mapsto TA}$ *(pour $k > i$)*

$$P^{-1} = \begin{bmatrix} 1 & & & & & \\ & \ddots & & & & \\ & & 0 & & 1 & \\ & & & \ddots & & \\ & & 1 & & 0 & \\ & & & & & \ddots & \\ & & & & & & 1 \end{bmatrix}, \qquad T^{-1} = \begin{bmatrix} 1 & & & & & \\ & \ddots & & & & \\ & & 1 & & & \\ & & & \ddots & & \\ & & +r & & 1 & \\ & & & & & \ddots & \\ & & & & & & 1 \end{bmatrix}$$

- Écriture matricielle de l'algorithme de Gauss

$\boxed{A \longmapsto M_1 \times M_2 \times \cdots \times M_q A = U}$ avec les M_p de type P ou T.

- Interprétation matricielle des opérations élémentaires
  - Permutation $\boxed{L_i \leftrightarrow L_k} \iff \boxed{A \mapsto PA}$
  - Transvection $\boxed{L_k \leftarrow L_k - rL_i} \iff \boxed{A \mapsto TA}$ $(pour\ k > i)$

$$P^{-1} = \begin{bmatrix} 1 & & & & & & \\ & \ddots & & & & & \\ & & 0 & & 1 & & \\ & & & \ddots & & & \\ & & 1 & & 0 & & \\ & & & & & \ddots & \\ & & & & & & 1 \end{bmatrix}, \quad T^{-1} = \begin{bmatrix} 1 & & & & & & \\ & \ddots & & & & & \\ & & 1 & & & & \\ & & & \ddots & & & \\ & & +r & & 1 & & \\ & & & & & \ddots & \\ & & & & & & 1 \end{bmatrix}$$

- Écriture matricielle de l'algorithme de Gauss

$\boxed{A = N_1 \times N_2 \times \cdots \times N_q \times U}$ avec les N_p de type P ou T !

- Interprétation matricielle des opérations élémentaires
  - Permutation $\boxed{L_i \leftrightarrow L_k} \iff \boxed{A \mapsto PA}$
  - Transvection $\boxed{L_k \leftarrow L_k - rL_i} \iff \boxed{A \mapsto TA}$ *(pour $k > i$)*

- Écriture matricielle de l'algorithme de Gauss

$$\boxed{A = N_1 \times N_2 \times \cdots \times N_q \times U}$$ avec les N_p de type P ou T !

  - Si aucune permutation, les N_p sont tri. inf.

- Interprétation matricielle des opérations élémentaires
  - Permutation $\boxed{L_i \leftrightarrow L_k} \iff \boxed{A \mapsto PA}$
  - Transvection $\boxed{L_k \leftarrow L_k - rL_i} \iff \boxed{A \mapsto TA}$ *(pour $k > i$)*
- Écriture matricielle de l'algorithme de Gauss

$$\boxed{A = N_1 \times N_2 \times \cdots \times N_q \times U} \quad \text{avec les } N_p \text{ de type } P \text{ ou } T\,!$$

  - Si aucune permutation, les N_p sont tri. inf.
  - Dans ce cas,

$$A = LU \quad \text{avec } L \text{ tri. inf. et } U \text{ tri. sup.}$$

- Interprétation matricielle des opérations élémentaires
  - Permutation $\boxed{L_i \leftrightarrow L_k} \iff \boxed{A \mapsto PA}$
  - Transvection $\boxed{L_k \leftarrow L_k - rL_i} \iff \boxed{A \mapsto TA}$ *(pour $k > i$)*
- Écriture matricielle de l'algorithme de Gauss

$$\boxed{A = N_1 \times N_2 \times \cdots \times N_q \times U}$$ avec les N_p de type P ou T !

  - Si aucune permutation, les N_p sont tri. inf.
  - Dans ce cas,

$$A = LU \quad \text{avec } L \text{ tri. inf. et } U \text{ tri. sup.}$$

  - **Comment caractériser ce cas ?**

Définition. Soit $A \in \mathbb{R}^{d \times d}$. On appelle mineur fondamental de taille $k \leqslant d$ le déterminant

$$\Delta_k = \det\left[(A_{ij})_{1 \leqslant i,j \leqslant k}\right].$$

Théorème. Soit $A \in \mathbb{R}^{d \times d}$ telle que

$$\forall k \leqslant d, \quad \Delta_k \neq 0.$$

Alors $\exists!(L, U)$ t.q.

- L tri. inf. avec diagonale unité.
- U tri. sup.
- $A = LU$.

Intérêt algorithmique de la décomposition LU.

Intérêt algorithmique de la décomposition LU.

- On doit résoudre $Ax_p = b_p$ pour de nombreux $p = 1, \ldots, P$.
  - 1 fois : $A = LU$.
  - P fois : $Ly_p = b_p$ et $Ux_p = y_p$.
  - Coût total : $\mathcal{O}(d^3 + Pd^2)$.

Intérêt algorithmique de la décomposition LU.

- On doit résoudre $Ax_p = b_p$ pour de nombreux $p = 1, \dots, P$.
  - 1 fois : $A = LU$.
  - P fois : $Ly_p = b_p$ et $Ux_p = y_p$.
  - Coût total : $\mathcal{O}(d^3 + Pd^2)$.
- Mieux que P fois Gauss.

Intérêt algorithmique de la décomposition LU.

- On doit résoudre $Ax_p = b_p$ pour de nombreux $p = 1, \ldots, P$.
  - 1 fois : $A = LU$.
  - P fois : $Ly_p = b_p$ et $Ux_p = y_p$.
  - Coût total : $\mathcal{O}(d^3 + Pd^2)$.
- Mieux que P fois Gauss.
- Vraiment utile seulement lorsque les b_p sont connus seulement successivement...

Intérêt algorithmique de la décomposition LU.

- On doit résoudre $Ax_p = b_p$ pour de nombreux $p = 1, \ldots, P$.
  - 1 fois : $A = LU$.
  - P fois : $Ly_p = b_p$ et $Ux_p = y_p$.
  - Coût total : $\mathcal{O}(d^3 + Pd^2)$.
- Mieux que P fois Gauss.
- Vraiment utile seulement lorsque les b_p sont connus seulement successivement...

Extensions, optimisations

- Si A est symétrique, définie positive, $A = BB^\mathsf{T}$ avec B tri.inf.
- Si A a une structure particulière, les calculs peuvent être simplifiés.

Intérêt algorithmique de la décomposition LU.

- On doit résoudre $Ax_p = b_p$ pour de nombreux $p = 1, \ldots, P$.
  - 1 fois : $A = LU$.
  - P fois : $Ly_p = b_p$ et $Ux_p = y_p$.
  - Coût total : $\mathcal{O}(d^3 + Pd^2)$.
- Mieux que P fois Gauss.
- Vraiment utile seulement lorsque les b_p sont connus seulement successivement. . .

Extensions, optimisations

- Si A est symétrique, définie positive, $A = BB^{\mathsf{T}}$ avec B tri.inf.
- Si A a une structure particulière, les calculs peuvent être simplifiés.

Méthode utilisée par matlab ? Voir `help \`

- Un système linéaire...

$$\begin{cases}
\alpha x_1 + x_2 + \dots \dots + x_d = 1 \\
x_1 + \alpha x_2 = 1 \\
x_1 + \alpha x_3 = 1 \\
\vdots \qquad \dots \qquad \vdots \\
x_1 + \alpha x_d = 1
\end{cases}$$

- Un système linéaire...

$$\begin{cases} \alpha x_1 + x_2 + \dots \dots + x_d = 1 \\ x_1 + \alpha x_2 = 1 \\ x_1 + \alpha x_3 = 1 \\ \vdots \quad \dots \quad \vdots \\ x_1 + \alpha x_d = 1 \end{cases} \qquad \begin{cases} \alpha y_d + y_{d-1} + \dots \dots + y_1 = 1 \\ y_d + \alpha y_{d-1} = 1 \\ y_d + \alpha y_{d-2} = 1 \\ \vdots \quad \dots \quad \vdots \\ y_d + \alpha y_1 = 1 \end{cases}$$

- Un système linéaire...

$$
\begin{cases}
\alpha x_1 + x_2 + \dots \dots + x_d = 1 \\
x_1 + \alpha x_2 = 1 \\
x_1 + \alpha x_3 = 1 \\
\vdots \qquad \dots \qquad \vdots \\
x_1 + \alpha x_d = 1
\end{cases}
\qquad
\begin{cases}
y_1 + y_2 + \dots \dots + \alpha y_d = 1 \\
\alpha y_{d-1} + y_d = 1 \\
\alpha y_{d-2} + y_d = 1 \\
\vdots \qquad \dots \qquad \vdots \\
\alpha y_1 + y_d = 1
\end{cases}
$$

- Un système linéaire...

$$\begin{cases} \alpha x_1 + x_2 + \dots \dots + x_d = 1 \\ x_1 + \alpha x_2 = 1 \\ x_1 + \alpha x_3 = 1 \\ \vdots \qquad \dots \qquad \vdots \\ x_1 + \alpha x_d = 1 \end{cases} \qquad \begin{cases} \alpha y_1 + y_d = 1 \\ \vdots \qquad \dots \qquad \vdots \\ \alpha y_{d-2} + y_d = 1 \\ \alpha y_{d-1} + y_d = 1 \\ y_1 + y_2 + \dots \dots + \alpha y_d = 1 \end{cases}$$

- Un système linéaire...

$$
\left\{
\begin{array}{llllll}
\alpha x_1 + & x_2 & + \dots & \dots + & x_d & = 1 \\
x_1 & + \alpha x_2 & & & & = 1 \\
x_1 & & + \alpha x_3 & & & = 1 \\
\vdots & & \dots & & & \vdots \\
x_1 & & & & + \alpha x_d & = 1
\end{array}
\right.
\qquad
\left\{
\begin{array}{lllll}
\alpha y_1 & & & + \; y_d & = 1 \\
\vdots & \dots & & & \vdots \\
& \alpha y_{d-2} & & + \; y_d & = 1 \\
& & \alpha y_{d-1} & + \; y_d & = 1 \\
y_1 \; + y_2 + & \dots & \dots & + \alpha y_d & = 1
\end{array}
\right.
$$

- ... deux matrices...

$$
A_1 = \begin{pmatrix}
\alpha & 1 & \cdots & 1 \\
1 & \ddots & & \\
\vdots & & \ddots & \\
1 & & & \alpha
\end{pmatrix}
\qquad
A_2 = \begin{pmatrix}
\alpha & & & 1 \\
& \ddots & & \vdots \\
& & \ddots & 1 \\
1 & \cdots & 1 & \alpha
\end{pmatrix}
$$

- Un système linéaire… ($\alpha > d$)

$$
\left\{\begin{array}{lllllr}
\alpha x_1 + & x_2 & + \ldots & \ldots + & x_d & = 1 \\
x_1 & + \alpha x_2 & & & & = 1 \\
x_1 & & + \alpha x_3 & & & = 1 \\
\vdots & & \ldots & & & \vdots \\
x_1 & & & & + \alpha x_d & = 1
\end{array}\right.
\qquad
\left\{\begin{array}{llllr}
\alpha y_1 & & & + \; y_d & = 1 \\
\vdots & \ldots & & & \vdots \\
& \alpha y_{d-2} & & + \; y_d & = 1 \\
& & \alpha y_{d-1} & + \; y_d & = 1 \\
y_1 \; + y_2 + & \ldots & \ldots & + \alpha y_d & = 1
\end{array}\right.
$$

- … deux matrices…

$$
A_1 = \begin{pmatrix} \alpha & 1 & \cdots & 1 \\ 1 & \ddots & & \\ \vdots & & \ddots & \\ 1 & & & \alpha \end{pmatrix}
\qquad
A_2 = \begin{pmatrix} \alpha & & & 1 \\ & \ddots & & \vdots \\ & & \ddots & 1 \\ 1 & \cdots & 1 & \alpha \end{pmatrix}
$$

- … deux coûts si on utilise LU ! [*LUfleches.m*]

Analyse numérique matricielle

D'autres méthodes de résolution de systèmes linéaires

- **But :** construire une suite $\left(x^{(n)}\right)$ qui converge vers x t.q. $Ax = b$.

- But : construire une suite $\left(x^{(n)}\right)$ qui converge vers x t.q. $Ax = b$.
- Idée : si $A = M - N$ avec M inversible, alors

$$Ax = b \Longleftrightarrow x = \underbrace{M^{-1}(Nx + b)}_{\varphi(x)}.$$

- But : construire une suite $\left(x^{(n)}\right)$ qui converge vers x t.q. $Ax = b$.
- Idée : si $A = M - N$ avec M inversible, alors

$$Ax = b \Longleftrightarrow x = \underbrace{M^{-1}(Nx + b)}_{\varphi(x)}.$$

Problème de point fixe : soit $x^{(0)} \in \mathbb{R}^d$, on définit

$$x^{(n+1)} = \varphi\left(x^{(n)}\right).$$

- But : construire une suite $\left(x^{(n)}\right)$ qui converge vers x t.q. $Ax = b$.
- Idée : si $A = M - N$ avec M inversible, alors

$$Ax = b \Longleftrightarrow x = \underbrace{M^{-1}(Nx + b)}_{\varphi(x)}.$$

Problème de point fixe : soit $x^{(0)} \in \mathbb{R}^d$, on définit

$$x^{(n+1)} = \varphi\left(x^{(n)}\right).$$

> **Théorème.** La suite $\left(x^{(n)}\right)$ converge vers la solution de $Ax = b$ pour b quelconque **ssi** $\rho(M^{-1}N) < 1$.

- But : construire une suite $\left(x^{(n)}\right)$ qui converge vers x t.q. $Ax = b$.
- Idée : si $A = M - N$ avec M inversible, alors

$$Ax = b \Longleftrightarrow x = \underbrace{M^{-1}(Nx + b)}_{\varphi(x)}.$$

Problème de point fixe : soit $x^{(0)} \in \mathbb{R}^d$, on définit

$$x^{(n+1)} = \varphi\left(x^{(n)}\right).$$

> **Théorème.** La suite $\left(x^{(n)}\right)$ converge vers la solution de $Ax = b$ pour b quelconque **ssi** $\rho(M^{-1}N) < 1$.

Preuve. On pose $e^{(n)} = x - x^{(n)}$: $e^{(n+1)} = M^{-1}N\, e^{(n)}$ donc

$$e^{(n)} = (M^{-1}N)^n e^{(0)}.$$

Le cas où $M^{-1}N$ est diagonalisable est clair. $\square$

- But : construire une suite $\left(x^{(n)}\right)$ qui converge vers x t.q. $Ax = b$.
- Idée : si $A = M - N$ avec M inversible, alors

$$Ax = b \Longleftrightarrow x = \underbrace{M^{-1}(Nx + b)}_{\varphi(x)}.$$

Problème de point fixe : soit $x^{(0)} \in \mathbb{R}^d$, on définit

$$x^{(n+1)} = \varphi\left(x^{(n)}\right).$$

> **Théorème.** La suite $\left(x^{(n)}\right)$ converge vers la solution de $Ax = b$ pour b quelconque **ssi** $\rho(M^{-1}N) < 1$.

Remarque. Si $\rho(M^{-1}N) < 1$ alors $\exists ||| \cdot |||$ t.q. $|||M^{-1}N||| < 1$.

Rappel. $\rho(M^{-1}N) \leqslant |||M^{-1}N|||$ *pour toute norme subordonnée.*

$$x^{(n+1)} = M^{-1}(Nx^{(n)} + b).$$

En pratique.

$$x^{(n+1)} = M^{-1}(Nx^{(n)} + b).$$

En pratique.

- Système linéaire $Mx^{(n+1)} = Nx^{(n)} + b$ à chaque itération.

$$x^{(n+1)} = M^{-1}(Nx^{(n)} + b).$$

En pratique.

- Système linéaire $Mx^{(n+1)} = Nx^{(n)} + b$ à chaque itération.
- M doit être « facile à inverser ».

$$x^{(n+1)} = M^{-1}(Nx^{(n)} + b).$$

En pratique.

- Système linéaire $Mx^{(n+1)} = Nx^{(n)} + b$ à chaque itération.
- M doit être « facile à inverser ».
    - Méthode de Jacobi
      $M =$ diagonale de A *si non nulle !*

$$x^{(n+1)} = M^{-1}(Nx^{(n)} + b).$$

En pratique.

- Système linéaire $Mx^{(n+1)} = Nx^{(n)} + b$ à chaque itération.
- M doit être « facile à inverser ».
  - Méthode de Jacobi
    $M =$ diagonale de A *si non nulle !*
  - Méthode de Gauss-Seidel
    $M =$ partie triangulaire inférieure de A *si diagonale non nulle !*

$$x^{(n+1)} = M^{-1}(Nx^{(n)} + b).$$

En pratique.

- Système linéaire $Mx^{(n+1)} = Nx^{(n)} + b$ à chaque itération.
- M doit être « facile à inverser ».
  - Méthode de Jacobi
    $M =$ diagonale de A *si non nulle !*
  - Méthode de Gauss-Seidel
    $M =$ partie triangulaire inférieure de A *si diagonale non nulle !*
- Si $\rho(M^{-1}N) < 1$, convergence en $\mathcal{O}\left(\rho(M^{-1}N)^n\right)$.

Problématique : Rechercher quelques *vp* de A, et les $\overrightarrow{vp}$ associés.

Problématique : Rechercher quelques *vp* de A, et les $\overrightarrow{vp}$ associés.

- Constat. si $A = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}$ avec $\lambda_1 > \lambda_2 > 0$, et $x^0 = \begin{pmatrix} a \\ b \end{pmatrix}$,

Problématique : Rechercher quelques *vp* de A, et les $\overrightarrow{vp}$ associés.

- Constat. si $A = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}$ avec $\lambda_1 > \lambda_2 > 0$, et $x^0 = \begin{pmatrix} a \\ b \end{pmatrix}$,

$$A^n x^0 \sim a\lambda_1^n \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \qquad \text{si } a \neq 0.$$

Problématique : Rechercher quelques vp de A, et les $\overrightarrow{vp}$ associés.

- Constat. si $A = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}$ avec $\lambda_1 > \lambda_2 > 0$, et $x^0 = \begin{pmatrix} a \\ b \end{pmatrix}$,

$$A^n x^0 \sim a\lambda_1^n \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \qquad \text{si } a \neq 0.$$

Ainsi,

$$\frac{A^n x^0}{\|A^n x^0\|} \sim \mathrm{e}_1 \quad \text{avec} \quad A\mathrm{e}_1 = \lambda_1 \mathrm{e}_1 \ !$$

Problématique : Rechercher quelques *vp* de A, et les $\overrightarrow{vp}$ associés.

- Constat. si $A = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}$ avec $\lambda_1 > \lambda_2 > 0$, et $x^0 = \begin{pmatrix} a \\ b \end{pmatrix}$,

$$A^n x^0 \sim a\lambda_1^n \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \qquad \text{si } a \neq 0.$$

Ainsi,

$$\frac{A^n x^0}{\|A^n x^0\|} \sim e_1 \quad \text{avec} \quad Ae_1 = \lambda_1 e_1 \text{ !}$$

- Algorithme. Soit $x^0 \in \mathbb{R}^d$, on construit

$$\boxed{y^{n+1} = Ax^n \quad \text{et} \quad x^{n+1} = \frac{y^{n+1}}{\|y^{n+1}\|}}.$$

Algorithme. Soit $x^0 \in \mathbb{R}^d$, on construit

$$\boxed{y^{n+1} = Ax^n \quad \text{et} \quad x^{n+1} = \frac{y^{n+1}}{\|y^{n+1}\|}}.$$

Algorithme. Soit $x^0 \in \mathbb{R}^d$, on construit

$$\boxed{y^{n+1} = Ax^n \quad \text{et} \quad x^{n+1} = \frac{y^{n+1}}{\|y^{n+1}\|}}.$$

Théorème. Soit A de *vp* (complexes) $|\lambda_1| \leq |\lambda_2| \leq \cdots \leq |\lambda_{d-1}| < |\lambda_d|$

Pour presque† tout $x^0 \in \mathbb{R}^n$, la suite $(Ax^n|x^n)$ converge vers λ_d. De plus, la suite

$$(\tilde{x}^n) := \left(\frac{|\lambda_d|}{\lambda_d}\right)^n x^n$$

converge vers un vecteur propre associé à λ_d.

Algorithme. Soit $x^0 \in \mathbb{R}^d$, on construit

$$\boxed{y^{n+1} = Ax^n \quad \text{et} \quad x^{n+1} = \frac{y^{n+1}}{\|y^{n+1}\|}}.$$

Théorème. Soit A de *vp* (complexes) $|\lambda_1| \leq |\lambda_2| \leq \cdots \leq |\lambda_{d-1}| < |\lambda_d|$

Pour presque[†] tout $x^0 \in \mathbb{R}^n$, la suite $(Ax^n|x^n)$ converge vers λ_d. De plus, la suite

$$(\tilde{x}^n) := \left(\frac{|\lambda_d|}{\lambda_d}\right)^n x^n$$

converge vers un vecteur propre associé à λ_d.

[†] *La condition est : la projection de x^0 sur le sous-espace propre associé à λ_1 n'est pas 0.*

Attention à l'hypothèse $|\lambda_{d-1}| < |\lambda_d|$ qui est nécessaire.

[*puissance1.m*] [*puissance2.m*]

COURS 2

Intégration numérique

Intégration numérique

Modélisation : calcul d'un temps de parcours

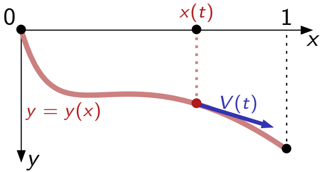
Deux toboggans... deux vitesses...

Quel temps de descente

pour chaque toboggan ?

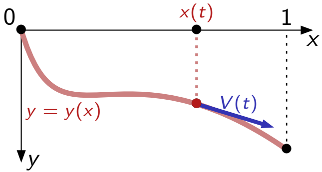
Intégration numérique

Modélisation : calcul d'un temps de parcours



Intégration numérique

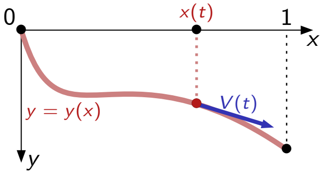
Modélisation : calcul d'un temps de parcours



- $x \mapsto y(x)$: forme du toboggan,

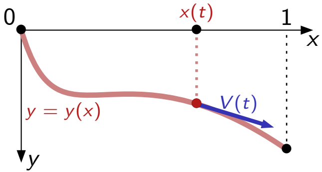
Intégration numérique

Modélisation : calcul d'un temps de parcours



- $x \mapsto y(x)$: forme du toboggan,
- $t \mapsto \big(x(t), y(x(t))\big)$: position,

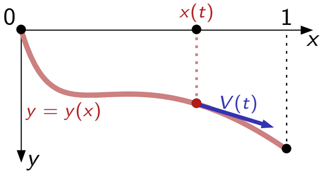
Modélisation : calcul d'un temps de parcours



- $x \mapsto y(x)$: forme du toboggan,
- $t \mapsto \big(x(t), y(x(t))\big)$: position,
- T : temps de descente,

Intégration numérique

Modélisation : calcul d'un temps de parcours

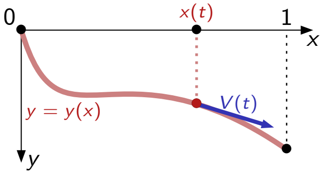


- $x \mapsto y(x)$: forme du toboggan,
- $t \mapsto \big(x(t), y(x(t))\big)$: position,
- T : temps de descente,

Vitesse : $V(t) = \begin{bmatrix} x'(t) \\ y'(x(t))x'(t) \end{bmatrix}, \qquad |V(t)| = |x'(t)|\sqrt{1 + y'^2(x(t))}$

Intégration numérique

Modélisation : calcul d'un temps de parcours

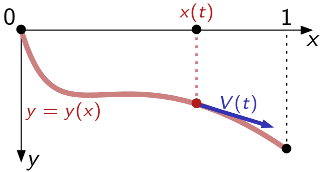


- $x \mapsto y(x)$: forme du toboggan,
- $t \mapsto \big(x(t), y(x(t))\big)$: position,
- T : temps de descente,

Vitesse : $V(t) = \begin{bmatrix} x'(t) \\ y'(x(t))x'(t) \end{bmatrix}, \qquad |V(t)| = |x'(t)|\sqrt{1 + y'^2(x(t))}$

Énergies : $E_c(t) = \dfrac{1}{2} m x'(t)^2 \left[1 + y'(x(t))^2\right], \quad E_p(t) = -mgy(x(t)).$

Modélisation : calcul d'un temps de parcours



- $x \mapsto y(x)$: forme du toboggan,
- $t \mapsto \big(x(t), y(x(t))\big)$: position,
- T : temps de descente,

Vitesse : $V(t) = \begin{bmatrix} x'(t) \\ y'(x(t))x'(t) \end{bmatrix}$, $\quad |V(t)| = |x'(t)|\sqrt{1 + y'^2(x(t))}$

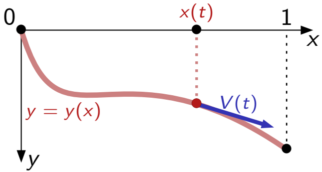
Énergies : $E_c(t) = \dfrac{1}{2} m x'(t)^2 \left[1 + y'(x(t))^2\right]$, $\quad E_p(t) = -mgy(x(t))$.

Bilan entre $t = 0$ et $t = T$: $\dfrac{1}{2} m \left(x'(t)\right)^2 \left[1 + y'(x(t))^2\right] - mgy(x(t)) = 0$.

$$\Longrightarrow x'(t) = \sqrt{2gy(x(t))}/\sqrt{1 + y'(x(t))^2}.$$

Intégration numérique

Modélisation : calcul d'un temps de parcours



- $x \mapsto y(x)$: forme du toboggan,
- $t \mapsto \big(x(t), y(x(t))\big)$: position,
- T : temps de descente,

Vitesse : $V(t) = \begin{bmatrix} x'(t) \\ y'(x(t))x'(t) \end{bmatrix}$, $\quad |V(t)| = |x'(t)|\sqrt{1 + y'^2(x(t))}$

Énergies : $E_c(t) = \dfrac{1}{2} m x'(t)^2 \left[1 + y'(x(t))^2\right]$, $\quad E_p(t) = -mgy(x(t))$.

Bilan entre $t = 0$ et $t = T$: $\dfrac{1}{2} m \left(x'(t)\right)^2 \left[1 + y'(x(t))^2\right] - mgy(x(t)) = 0$.

$$\Longrightarrow x'(t) = \sqrt{2gy(x(t))}/\sqrt{1 + y'(x(t))^2}.$$

Temps de descente : $T = \displaystyle\int_0^T \mathrm{d}t = \int_0^1 \frac{\sqrt{1 + y'(x)^2}}{\sqrt{2gy(x)}}\, \mathrm{d}x.$

Problématique. La forme $x \mapsto y(x)$ du toboggan étant donnée, comment calculer le temps de descente T ?

$$T = \int_0^1 \frac{\sqrt{1 + y'(x)^2}}{\sqrt{2gy(x)}} \, \mathrm{d}x.$$

Problématique. La forme $x \mapsto y(x)$ du toboggan étant donnée, comment calculer le temps de descente T ?

$$T = \int_0^1 \frac{\sqrt{1 + y'(x)^2}}{\sqrt{2gy(x)}} \, dx.$$

- Problème général : calcul de $\displaystyle\int_a^b f(x) \, dx$.

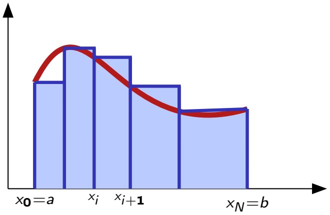
Problématique. La forme $x \mapsto y(x)$ du toboggan étant donnée, comment calculer le temps de descente T ?

$$T = \int_0^1 \frac{\sqrt{1 + y'(x)^2}}{\sqrt{2gy(x)}}\,\mathrm{d}x.$$

- Problème général : calcul de $\displaystyle\int_a^b f(x)\,\mathrm{d}x$.
- Enjeux :
  - Mise en place d'une méthode numérique,
  - Étude de la précision,
  - Analyse du temps de calcul,
  - Quelle méthode pour quelle fonction f ?

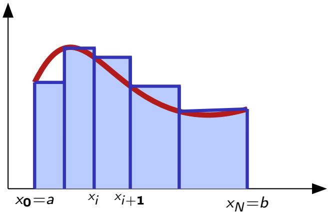
Intégration numérique

Une première idée : la méthode des rectangles



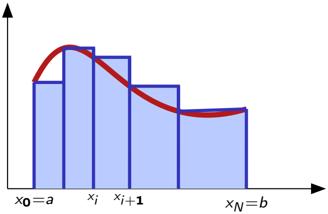
Intégration numérique

Une première idée : la méthode des rectangles



$$\int_a^b f(x)\,\mathrm{d}x \simeq \sum_{i=0}^{N-1} (x_{i+1} - x_i) f(x_i).$$

Une première idée : la méthode des rectangles



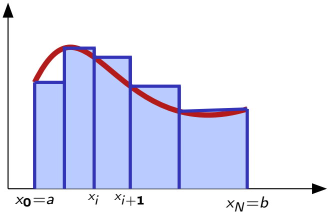
$$\int_a^b f(x)\,\mathrm{d}x \simeq \sum_{i=0}^{N-1} (x_{i+1} - x_i) f(x_i).$$

- Cas équidistant : $x_{i+1} - x_i = h = \frac{b-a}{N}$,

$$\int_a^b f(x)\,\mathrm{d}x \simeq h \sum_{i=0}^{N-1} f(x_i).$$

Intégration numérique

Une première idée : la méthode des rectangles



$$\int_a^b f(x)\,\mathrm{d}x \simeq \sum_{i=0}^{N-1} (x_{i+1} - x_i) f(x_i).$$

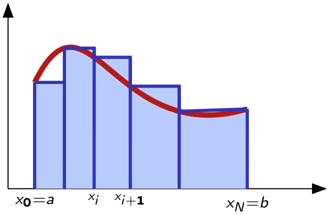
- Cas équidistant : $x_{i+1} - x_i = h = \frac{b-a}{N}$,

$$\int_a^b f(x)\,\mathrm{d}x \simeq h \sum_{i=0}^{N-1} f(x_i).$$

- Est-ce que l'approximation converge lorsque $N \to +\infty$?

Intégration numérique

Une première idée : la méthode des rectangles



$$\int_a^b f(x)\,\mathrm{d}x \simeq \sum_{i=0}^{N-1} (x_{i+1} - x_i) f(x_i).$$

- Cas équidistant : $x_{i+1} - x_i = h = \frac{b-a}{N}$,

$$\int_a^b f(x)\,\mathrm{d}x \simeq h \sum_{i=0}^{N-1} f(x_i).$$

- Est-ce que l'approximation converge lorsque $N \to +\infty$?
- Si oui, avec quelle précision ?

Intégration numérique

Convergence de la méthode des rectangles

- Cas équidistant (rappel : $Nh = b - a$)

$$I_N(f) = h \sum_{i=0}^{N-1} f(x_i).$$

- Cas équidistant (rappel : $Nh = b - a$)

$$I_N(f) = h \sum_{i=0}^{N-1} f(x_i).$$

- Si f est continue, $I_N(f)$ est une *somme de Riemann*.

$$\Longrightarrow I_N(f) \xrightarrow[N \to +\infty]{} \int_a^b f(x)\,\mathrm{d}x.$$

Convergence de la méthode des rectangles

- Cas équidistant (rappel : $Nh = b - a$)

$$I_N(f) = h \sum_{i=0}^{N-1} f(x_i).$$

- Si f est continue, $I_N(f)$ est une *somme de Riemann*.

$$\implies I_N(f) \xrightarrow[N \to +\infty]{} \int_a^b f(x)\,\mathrm{d}x.$$

C'est lié à l'approximation uniforme des fonctions continues par des fonctions en escalier / à la continuité uniforme des fonctions continues sur un segment.

- Cas équidistant (rappel : $Nh = b - a$)

$$I_N(f) = h \sum_{i=0}^{N-1} f(x_i).$$

- Si f est continue, $I_N(f)$ est une *somme de Riemann*.

$$\Longrightarrow I_N(f) \xrightarrow[N \to +\infty]{} \int_a^b f(x)\,\mathrm{d}x.$$

C'est lié à l'approximation uniforme des fonctions continues par des fonctions en escalier / à la continuité uniforme des fonctions continues sur un segment.

- Peut-on quantifier la précision de l'approximation ?

Précision de la méthode des rectangles

$$I = \int_a^b f(x)\,\mathrm{d}x, \qquad I_N = h\sum_{i=0}^{N-1} f(a+ih).$$

Précision de la méthode des rectangles

$$I = \int_a^b f(x)\,\mathrm{d}x, \qquad I_N = h\sum_{i=0}^{N-1} f(a+ih).$$

- Erreur :

$$|I - I_N| = \left|\sum_{i=0}^{N-1}\int_{x_i}^{x_{i+1}}\left[f(x)-f(x_i)\right]\mathrm{d}x\right|.$$

Précision de la méthode des rectangles

$$I = \int_a^b f(x)\,\mathrm{d}x, \qquad I_N = h\sum_{i=0}^{N-1} f(a+ih).$$

- Erreur :

$$|I - I_N| = \left|\sum_{i=0}^{N-1} \int_{x_i}^{x_{i+1}} \left[f(x) - f(x_i)\right] \mathrm{d}x\right|.$$

- Inégalité des accroissements finis : si $f \in \mathscr{C}^1([a,b])$,

$$|f(x) - f(x_i)| \leqslant M_1|x - x_i| \quad \text{avec} \quad M_1 = \sup_{t\in[a,b]} |f'(t)|.$$

Précision de la méthode des rectangles

$$I = \int_a^b f(x)\,\mathrm{d}x, \qquad I_N = h\sum_{i=0}^{N-1} f(a+ih).$$

- Erreur :

$$|I - I_N| = \left|\sum_{i=0}^{N-1}\int_{x_i}^{x_{i+1}} \left[f(x) - f(x_i)\right]\mathrm{d}x\right|.$$

- Inégalité des accroissements finis : si $f \in \mathscr{C}^1([a,b])$,

$$|f(x) - f(x_i)| \leqslant M_1|x - x_i| \quad \text{avec} \quad M_1 = \sup_{t\in[a,b]} |f'(t)|.$$

- Estimation d'erreur

$$|I - I_N| \leqslant \sum_{i=0}^{N-1}\int_{x_i}^{x_{i+1}} |f(x) - f(x_i)|\,\mathrm{d}x$$

Précision de la méthode des rectangles

$$I = \int_a^b f(x)\,\mathrm{d}x, \qquad I_N = h\sum_{i=0}^{N-1} f(a+ih).$$

- Erreur :

$$|I - I_N| = \left| \sum_{i=0}^{N-1} \int_{x_i}^{x_{i+1}} \left[f(x) - f(x_i) \right] \mathrm{d}x \right|.$$

- Inégalité des accroissements finis : si $f \in \mathscr{C}^1([a,b])$,

$$|f(x) - f(x_i)| \leqslant M_1 |x - x_i| \quad \text{avec} \quad M_1 = \sup_{t\in[a,b]} |f'(t)|.$$

- Estimation d'erreur

$$|I - I_N| \leqslant M_1 \sum_{i=0}^{N-1} \int_{x_i}^{x_{i+1}} |x - x_i|\,\mathrm{d}x$$

Précision de la méthode des rectangles

$$I = \int_a^b f(x)\,\mathrm{d}x, \qquad I_N = h \sum_{i=0}^{N-1} f(a+ih).$$

- Erreur :

$$|I - I_N| = \left| \sum_{i=0}^{N-1} \int_{x_i}^{x_{i+1}} \Big[f(x) - f(x_i) \Big] \,\mathrm{d}x \right|.$$

- Inégalité des accroissements finis : si $f \in \mathscr{C}^1([a,b])$,

$$|f(x) - f(x_i)| \leqslant M_1 |x - x_i| \quad \text{avec} \quad M_1 = \sup_{t\in[a,b]} |f'(t)|.$$

- Estimation d'erreur

$$|I - I_N| \leqslant M_1 \sum_{i=0}^{N-1} \int_{x_i}^{x_{i+1}} (x - x_i)\,\mathrm{d}x$$

Intégration numérique

Précision de la méthode des rectangles

$$I = \int_a^b f(x)\,\mathrm{d}x, \qquad I_N = h\sum_{i=0}^{N-1} f(a+ih).$$

- Erreur :

$$|I - I_N| = \left|\sum_{i=0}^{N-1}\int_{x_i}^{x_{i+1}} \left[f(x) - f(x_i)\right]\mathrm{d}x\right|.$$

- Inégalité des accroissements finis : si $f \in \mathscr{C}^1([a,b])$,

$$|f(x) - f(x_i)| \leqslant M_1 |x - x_i| \quad \text{avec} \quad M_1 = \sup_{t\in[a,b]} |f'(t)|.$$

- Estimation d'erreur

$$|I - I_N| \leqslant M_1 \sum_{i=0}^{N-1} \left[\frac{(x-x_i)^2}{2}\right]_{x_i}^{x_{i+1}} \mathrm{d}x$$

Précision de la méthode des rectangles

$$I = \int_a^b f(x)\,\mathrm{d}x, \qquad I_N = h \sum_{i=0}^{N-1} f(a+ih).$$

- Erreur :

$$|I - I_N| = \left| \sum_{i=0}^{N-1} \int_{x_i}^{x_{i+1}} \left[f(x) - f(x_i) \right] \mathrm{d}x \right|.$$

- Inégalité des accroissements finis : si $f \in \mathscr{C}^1([a,b])$,

$$|f(x) - f(x_i)| \leqslant M_1 |x - x_i| \quad \text{avec} \quad M_1 = \sup_{t \in [a,b]} |f'(t)|.$$

- Estimation d'erreur

$$|I - I_N| \leqslant M_1 \sum_{i=0}^{N-1} \frac{h^2}{2}$$

Précision de la méthode des rectangles

$$I = \int_a^b f(x)\,\mathrm{d}x, \qquad I_N = h \sum_{i=0}^{N-1} f(a+ih).$$

- Erreur :

$$|I - I_N| = \left| \sum_{i=0}^{N-1} \int_{x_i}^{x_{i+1}} \left[f(x) - f(x_i) \right] \mathrm{d}x \right|.$$

- Inégalité des accroissements finis : si $f \in \mathscr{C}^1([a,b])$,

$$|f(x) - f(x_i)| \leqslant M_1 |x - x_i| \quad \text{avec} \quad M_1 = \sup_{t\in[a,b]} |f'(t)|.$$

- Estimation d'erreur

$$|I - I_N| \leqslant M_1 N \frac{h^2}{2}$$

Précision de la méthode des rectangles

$$I = \int_a^b f(x)\,\mathrm{d}x, \qquad I_N = h \sum_{i=0}^{N-1} f(a+ih).$$

- Erreur :

$$|I - I_N| = \left| \sum_{i=0}^{N-1} \int_{x_i}^{x_{i+1}} \left[f(x) - f(x_i) \right] \mathrm{d}x \right|.$$

- Inégalité des accroissements finis : si $f \in \mathscr{C}^1([a,b])$,

$$|f(x) - f(x_i)| \leqslant M_1 |x - x_i| \quad \text{avec} \quad M_1 = \sup_{t \in [a,b]} |f'(t)|.$$

- Estimation d'erreur

$$|I - I_N| \leqslant \frac{M_1(b-a)}{2} h$$

Précision de la méthode des rectangles

$$I = \int_a^b f(x)\,\mathrm{d}x, \qquad I_N = h \sum_{i=0}^{N-1} f(a+ih).$$

- Erreur :

$$|I - I_N| = \left| \sum_{i=0}^{N-1} \int_{x_i}^{x_{i+1}} \Big[f(x) - f(x_i) \Big] \,\mathrm{d}x \right|.$$

- Inégalité des accroissements finis : si $f \in \mathscr{C}^1([a,b])$,

$$|f(x) - f(x_i)| \leqslant M_1 |x - x_i| \quad \text{avec} \quad M_1 = \sup_{t \in [a,b]} |f'(t)|.$$

- Estimation d'erreur

$$|I - I_N| \leqslant \frac{M_1 (b-a)^2}{2N}$$

Intégration numérique

Précision de la méthode des rectangles

Si $f \in \mathscr{C}^1([a,b])$, alors, pour la méthode des rectangles,

$$|I - I_N| \leqslant \frac{(b-a)^2}{2N} \sup_{t \in [a,b]} |f'(t)|$$

Précision de la méthode des rectangles

> Si $f \in \mathscr{C}^1([a,b])$, alors, pour la méthode des rectangles,
>
> $$|I - I_N| \leqslant \frac{(b-a)^2}{2N} \sup_{t \in [a,b]} |f'(t)|$$

À retenir :

- Hypothèse de régularité : $f \in \mathscr{C}^1([a,b])$.
- Convergence en $\mathcal{O}\left(\frac{1}{N}\right)$.

> Si $f \in \mathscr{C}^1([a,b])$, alors, pour la méthode des rectangles,
>
> $$|I - I_N| \leqslant \frac{(b-a)^2}{2N} \sup_{t\in[a,b]} |f'(t)|$$

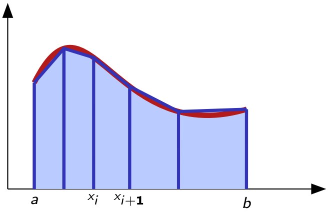
À retenir :

- Hypothèse de régularité : $f \in \mathscr{C}^1([a,b])$.
- Convergence en $\mathcal{O}\left(\frac{1}{N}\right)$.

Question : que se passe-t-il si $f \notin \mathscr{C}^1([a,b])$?

- Simulation pour diverses fonctions.
- Analyse théorique : *cf. TD*.

- Méthode des trapèzes

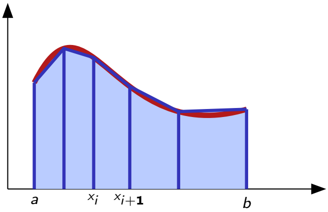


$$\int_a^b f(x)\,\mathrm{d}x \simeq \sum_{i=0}^{N-1} (x_{i+1} - x_i)\frac{f(x_i) + f(x_{i+1})}{2}.$$

Intégration numérique

Une autre méthode

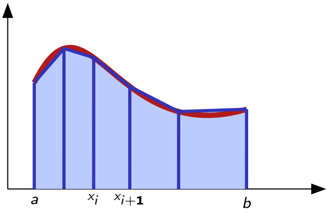
- Méthode des trapèzes



$$\int_a^b f(x)\,\mathrm{d}x \simeq \sum_{i=0}^{N-1} (x_{i+1} - x_i) \frac{f(x_i) + f(x_{i+1})}{2}.$$

- Erreur observée dans le cas général : $\mathcal{O}\left(N^{-2}\right)$.

- Méthode des trapèzes



$$\int_a^b f(x)\,\mathrm{d}x \simeq \sum_{i=0}^{N-1} (x_{i+1} - x_i)\frac{f(x_i) + f(x_{i+1})}{2}.$$

- Erreur observée dans le cas général : $\mathcal{O}\left(N^{-2}\right)$.
- Estimation d'erreur théorique ?

  $\rightsquigarrow$ *estimation sur $[x_i, x_{i+1}]$ de la différence courbe/corde.*

- Estimation d'erreur (subdivision uniforme de pas h)

- Estimation d'erreur (subdivision uniforme de pas h)

$$\left|I - I_N\right| \leqslant \sum_{i=0}^{N-1} \int_{x_i}^{x_{i+1}} |f(x) - p(x)| \, \mathrm{d}x,$$

où $p \in \mathbb{P}_1$ satisfait $p(x_i) = f(x_i)$ et $p(x_{i+1}) = f(x_{i+1})$.

- Estimation d'erreur (subdivision uniforme de pas h)

$$\left|I - I_N\right| \leqslant \sum_{i=0}^{N-1} \int_{x_i}^{x_{i+1}} |f(x) - p(x)| \, \mathrm{d}x,$$

où $p \in \mathbb{P}_1$ satisfait $p(x_i) = f(x_i)$ et $p(x_{i+1}) = f(x_{i+1})$.

- On pose $\psi = f - p$. Th. Rolle : $\exists \xi_i \in [x_i, x_{i+1}]$, $\psi'(\xi_i) = 0$.

$$\forall x \in [x_i, x_{i+1}], \ \left|\psi'(x)\right| \leqslant |x - \xi_i| \sup_{[x_i, x_{i+1}]} \left|\psi''\right| \leqslant h \sup_{[a,b]} \left|f''\right|.$$

- Estimation d'erreur (subdivision uniforme de pas h)

$$\left|I - I_N\right| \leqslant \sum_{i=0}^{N-1} \int_{x_i}^{x_{i+1}} |f(x) - p(x)| \, \mathrm{d}x,$$

où $p \in \mathbb{P}_1$ satisfait $p(x_i) = f(x_i)$ et $p(x_{i+1}) = f(x_{i+1})$.

  - On pose $\psi = f - p$. Th. Rolle : $\exists \xi_i \in [x_i, x_{i+1}]$, $\psi'(\xi_i) = 0$.

$$\forall x \in [x_i, x_{i+1}], \ \left|\psi'(x)\right| \leqslant |x - \xi_i| \sup_{[x_i, x_{i+1}]} \left|\psi''\right| \leqslant h \sup_{[a,b]} \left|f''\right|.$$

  - Donc $\forall x \in [x_i, x_{i+1}]$,

$$|\psi(x)| \leqslant \int_{x_i}^{x} |\psi'(x)| \leqslant h^2 \sup_{[a,b]} \left|f''\right|.$$

- Estimation d'erreur (subdivision uniforme de pas h)

$$\left|I - I_N\right| \leqslant \sum_{i=0}^{N-1} \int_{x_i}^{x_{i+1}} |f(x) - p(x)| \,\mathrm{d}x,$$

où $p \in \mathbb{P}_1$ satisfait $p(x_i) = f(x_i)$ et $p(x_{i+1}) = f(x_{i+1})$.

  - On pose $\psi = f - p$. Th. Rolle : $\exists \xi_i \in [x_i, x_{i+1}]$, $\psi'(\xi_i) = 0$.

$$\forall x \in [x_i, x_{i+1}],\ \left|\psi'(x)\right| \leqslant |x - \xi_i| \sup_{[x_i, x_{i+1}]} \left|\psi''\right| \leqslant h \sup_{[a,b]} \left|f''\right|.$$

  - Donc $\forall x \in [x_i, x_{i+1}]$,

$$|\psi(x)| \leqslant \int_{x_i}^{x} |\psi'(x)| \leqslant h^2 \sup_{[a,b]} \left|f''\right|.$$

  - Conclusion : $\boxed{|I - I_N| \leqslant h^2 (b-a) \sup_{[a,b]} \left|f''\right| = \frac{(b-a)^3}{N^2} M_2.}$

D'autres méthodes

- Idée : remplacer φ par un polynôme π_φ et
$$\int_0^1 \varphi(t)\,\mathrm{d}t \simeq \int_0^1 \pi_\varphi(t)\,\mathrm{d}t.$$

- **Idée** : remplacer φ par un polynôme π_φ et

$$\int_0^1 \varphi(t)\,\mathrm{d}t \simeq \int_0^1 \pi_\varphi(t)\,\mathrm{d}t.$$

- **Rappel** : pour $0 \leqslant t_0 < t_1 < \cdots < t_k \leqslant 1$ et $b_0, b_1, \ldots, b_k \in \mathbb{R}$ fixés,

$$\exists! p \in \mathbb{P}_k, \quad \forall q = 0, 1, \ldots, k, \quad p(t_q) = b_q.$$

  Dans la base de Lagrange,

$$p(t) = \sum_{q=0}^{k} b_q L_q(t), \quad \text{avec} \quad L_q(t) = \prod_{\ell \neq q} \frac{t - t_\ell}{t_q - t_\ell}.$$

- Idée : remplacer φ par un polynôme π_φ et

$$\int_0^1 \varphi(t)\,\mathrm{d}t \simeq \int_0^1 \pi_\varphi(t)\,\mathrm{d}t.$$

- Rappel : pour $0 \leqslant t_0 < t_1 < \cdots < t_k \leqslant 1$ et $b_0, b_1, \ldots, b_k \in \mathbb{R}$ fixés,

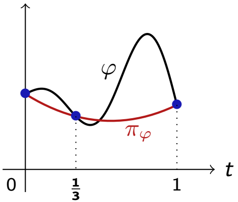
$$\exists! p \in \mathbb{P}_k, \quad \forall q = 0, 1, \ldots, k, \quad p(t_q) = b_q.$$

Dans la base de Lagrange,

$$p(t) = \sum_{q=0}^{k} b_q L_q(t), \quad \text{avec} \quad L_q(t) = \prod_{\ell \neq q} \frac{t - t_\ell}{t_q - t_\ell}.$$

- Exemple :

π_φ = polynôme d'interpolation de Lagrange de φ en 0, $\frac{1}{3}$, 1.



$$\int_0^1 \varphi(t)\,\mathrm{d}t \simeq \int_0^1 \pi_\varphi(t)\,\mathrm{d}t.$$

$$\int_0^1 \varphi(t)\,\mathrm{d}t \simeq \int_0^1 \pi_\varphi(t)\,\mathrm{d}t.$$

- Il n'est pas nécessaire d'expliciter π_φ pour approcher l'intégrale :

$$\pi_\varphi(t) = \sum_{q=0}^{k} \varphi(t_q) L_q(t) \quad \Longrightarrow \quad \int_0^1 \pi_\varphi(t)\,\mathrm{d}t = \sum_{q=0}^{k} \underbrace{\left(\int_0^1 L_q(t)\,\mathrm{d}t\right)}_{=w_q} \varphi(t_q)$$

$$\int_0^1 \varphi(t)\,\mathrm{d}t \simeq \int_0^1 \pi_\varphi(t)\,\mathrm{d}t.$$

- Il n'est pas nécessaire d'expliciter π_φ pour approcher l'intégrale :

$$\pi_\varphi(t) = \sum_{q=0}^{k} \varphi(t_q) L_q(t) \quad \Longrightarrow \quad \int_0^1 \pi_\varphi(t)\,\mathrm{d}t = \sum_{q=0}^{k} \underbrace{\left(\int_0^1 L_q(t)\,\mathrm{d}t \right)}_{=w_q} \varphi(t_q)$$

d'où l'approximation

$$\boxed{\int_0^1 \varphi(t)\,\mathrm{d}t \simeq \sum_{q=0}^{k} w_q \varphi(t_q).}$$

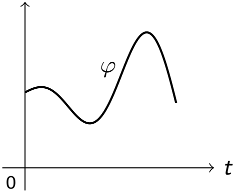
Intermède : interpolation de Lagrange

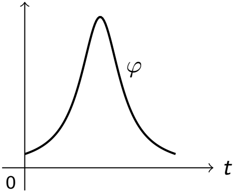
- Que se passe-t-il lorsque $k \to \infty$? *(points (t_q) équidistants)*

- Que se passe-t-il lorsque $k \to \infty$? *(points (t_q) équidistants)*

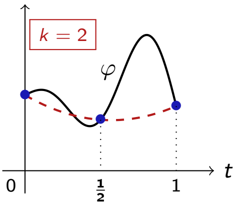
$$\varphi(t) = 1 + \frac{t}{2}\cos(8t) \qquad\qquad \varphi(t) = \frac{2}{1 + 10(2t-1)^2}$$

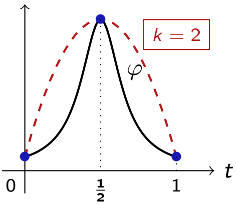
- Que se passe-t-il lorsque $k \to \infty$? *(points (t_q) équidistants)*



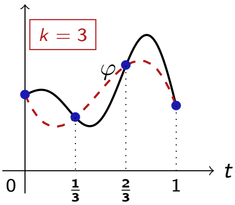


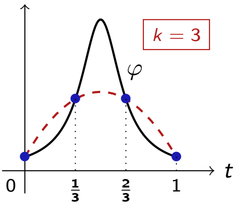
- Que se passe-t-il lorsque $k \to \infty$? *(points (t_q) équidistants)*



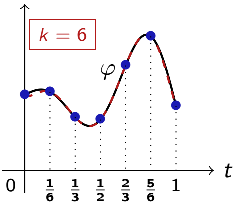


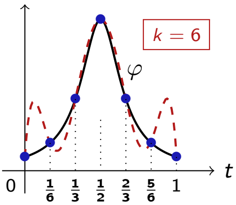
- Que se passe-t-il lorsque $k \to \infty$? *(points (t_q) équidistants)*



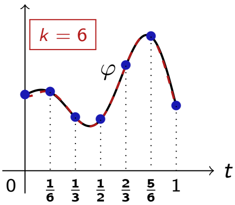


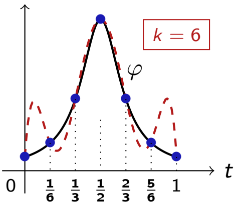
- Que se passe-t-il lorsque $k \to \infty$? *(points (t_q) équidistants)*





- Que se passe-t-il lorsque $k \to \infty$? *(points (t_q) équidistants)*





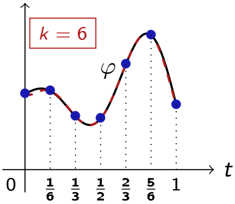
Non convergence lorsque $k \to +\infty$

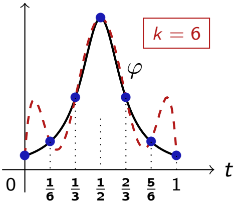
(phénomène de Runge)

Intégration numérique

Intermède : interpolation de Lagrange

- Que se passe-t-il lorsque $k \to \infty$? *(points (t_q) équidistants)*

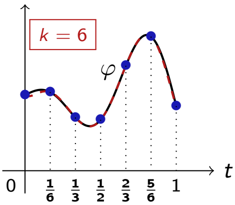


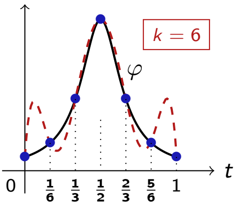


Non convergence lorsque $k \to +\infty$
(phénomène de Runge)

- Pas d'espoir d'obtenir une approximation qui converge lorsque $k \to +\infty$.

- Que se passe-t-il lorsque $k \to \infty$? *(points (t_q) équidistants)*





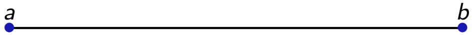
Non convergence lorsque $k \to +\infty$
(phénomène de Runge)

- Pas d'espoir d'obtenir une approximation qui converge lorsque $k \to +\infty$.
- On va conserver k fixé et travailler sur une subdivision de l'intervalle.

Intégration numérique

Principe des méthodes composées

$$\int_a^b f(x)\,\mathrm{d}x$$

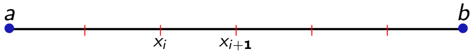


Intégration numérique

Principe des méthodes composées

$$\int_a^b f(x)\,\mathrm{d}x = \sum_{i=0}^{N-1} \int_{x_i}^{x_{i+1}} f(x)\,\mathrm{d}x$$

- On subdivise l'intervalle $[a, b]$ (pas uniforme h).

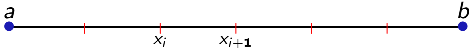


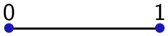
Intégration numérique

Principe des méthodes composées

$$\int_a^b f(x)\,\mathrm{d}x = \sum_{i=0}^{N-1} \int_{x_i}^{x_{i+1}} f(x)\,\mathrm{d}x$$

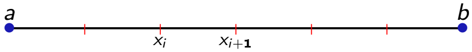
- On subdivise l'intervalle $[a, b]$ (pas uniforme h).
- On choisit un modèle élémentaire sur $[0, 1]$

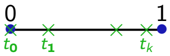




$$\int_a^b f(x)\,\mathrm{d}x = \sum_{i=0}^{N-1} \int_{x_i}^{x_{i+1}} f(x)\,\mathrm{d}x$$

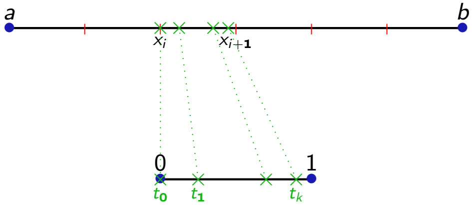
- On subdivise l'intervalle $[a, b]$ (pas uniforme h).
- On choisit un modèle élémentaire sur $[0, 1]$: $\displaystyle\int_0^1 \varphi(t)\,\mathrm{d}t \simeq \sum_{q=0}^{k} w_q \varphi(t_q)$





$$\int_a^b f(x)\,\mathrm{d}x = \sum_{i=0}^{N-1} \int_{x_i}^{x_{i+1}} f(x)\,\mathrm{d}x = \sum_{i=0}^{N-1} h \int_0^1 f(x_i + th)\,\mathrm{d}t \simeq h \sum_{i=0}^{N-1} \sum_{q=0}^{k} w_q f(x_i + t_q h).$$

- On subdivise l'intervalle $[a, b]$ (pas uniforme h).
- On choisit un modèle élémentaire sur $[0, 1]$: $\displaystyle\int_0^1 \varphi(t)\,\mathrm{d}t \simeq \sum_{q=0}^{k} w_q \varphi(t_q)$
- On transporte sur chaque sous-intervalle.



Intégration numérique

Méthodes composées

Exemple :

- $(k = 1)$. $t_0 = 0,\ t_1 = 1,$

Exemple :

- $(k = 1)$. $t_0 = 0$, $t_1 = 1$,
- $w_0 = \frac{1}{2}$, $w_1 = \frac{1}{2}$.

Exemple :

- $(k = 1)$. $t_0 = 0$, $t_1 = 1$,
- $w_0 = \frac{1}{2}$, $w_1 = \frac{1}{2}$.

$$\int_a^b f(x)\,\mathrm{d}x \simeq h \sum_{i=0}^{N-1} \sum_{q=0}^{k} w_q f(x_i + t_q h).$$

$$= h \sum_{i=0}^{N-1} \frac{f(x_i) + f(x_{i+1})}{2}$$

Exemple :

- $(k = 1)$. $t_0 = 0$, $t_1 = 1$,
- $w_0 = \frac{1}{2}$, $w_1 = \frac{1}{2}$.

$$\int_a^b f(x)\,\mathrm{d}x \simeq h \sum_{i=0}^{N-1} \sum_{q=0}^{k} w_q f(x_i + t_q h).$$

$$= h \sum_{i=0}^{N-1} \frac{f(x_i) + f(x_{i+1})}{2}$$

$\Longrightarrow$ Méthode des trapèzes !

- Bilan
  - Modèle élémentaire sur $[0, 1]$:

$$(M_e) \qquad \int_0^1 \varphi(t)\,\mathrm{d}t \simeq \sum_{q=0}^{k} w_q \varphi(t_q).$$

- Bilan
  - Modèle élémentaire sur $[0, 1]$:

$$(M_e) \qquad \int_0^1 \varphi(t)\,\mathrm{d}t \simeq \sum_{q=0}^{k} w_q \varphi(t_q).$$

  - Méthode composée sur $[a, b]$:

$$(FQ_c) \qquad I = \int_a^b f(x)\,\mathrm{d}x \simeq I_N = h \sum_{i=0}^{N-1} \sum_{q=0}^{k} w_q f(x_i + t_q h).$$

- Bilan
  - Modèle élémentaire sur $[0,1]$:

$$(M_e) \qquad \int_0^1 \varphi(t)\,\mathrm{d}t \simeq \sum_{q=0}^{k} w_q \varphi(t_q).$$

  - Méthode composée sur $[a,b]$:

$$(FQ_c) \qquad I = \int_a^b f(x)\,\mathrm{d}x \simeq I_N = h \sum_{i=0}^{N-1} \sum_{q=0}^{k} w_q f(x_i + t_q h).$$

  - *A priori* : $N(k+1)$ évaluations de f.

- Bilan
  - Modèle élémentaire sur $[0, 1]$:

$$(M_e) \qquad \int_0^1 \varphi(t)\,\mathrm{d}t \simeq \sum_{q=0}^{k} w_q \varphi(t_q).$$

  - Méthode composée sur $[a, b]$:

$$(FQ_c) \qquad I = \int_a^b f(x)\,\mathrm{d}x \simeq I_N = h \sum_{i=0}^{N-1} \sum_{q=0}^{k} w_q f(x_i + t_q h).$$

  - *A priori* : $N(k+1)$ évaluations de f.
- Quelle précision pour (**??**) ?

- Une remarque

$$I - I_N = h \sum_{i=0}^{N-1} \left(\int_0^1 \underbrace{f(x_i + th)}_{\varphi_i(t)} \, \mathrm{d}t - \int_0^1 \pi_{\varphi_i}(t) \, \mathrm{d}t \right),$$

  - $\varphi_i : t \mapsto f(x_i + th)$,
  - $\forall \varphi$, $\pi_\varphi \in \mathbb{P}_k$ satisfait $\pi_\varphi(t_q) = \varphi(t_q)$ $(0 \leqslant q \leqslant k)$.

- Une remarque

$$I - I_N = h \sum_{i=0}^{N-1} \left(\int_0^1 \underbrace{f(x_i + th)}_{\varphi_i(t)} \, \mathrm{d}t - \int_0^1 \pi_{\varphi_i}(t) \, \mathrm{d}t \right),$$

  - $\varphi_i : t \mapsto f(x_i + th)$,
  - $\forall \varphi$, $\pi_\varphi \in \mathbb{P}_k$ satisfait $\pi_\varphi(t_q) = \varphi(t_q)$ $(0 \leqslant q \leqslant k)$.

Théorème. pour $\varphi \in \mathscr{C}^{k+1}([0,1])$,

$$|\varphi(t) - \pi_\varphi(t)| \leqslant \frac{\sup_{t\in[0,1]} \left|\varphi^{(k+1)}(t)\right|}{(k+1)!}.$$

- Une remarque

$$I - I_N = h \sum_{i=0}^{N-1} \left(\int_0^1 \underbrace{f(x_i + th)}_{\varphi_i(t)} \,\mathrm{d}t - \int_0^1 \pi_{\varphi_i}(t) \,\mathrm{d}t \right),$$

  - $\varphi_i : t \mapsto f(x_i + th)$, $\boxed{\varphi_i^{(k+1)}(t) = h^{k+1} f^{(k+1)}(x_i + th).}$
  - $\forall \varphi$, $\pi_\varphi \in \mathbb{P}_k$ satisfait $\pi_\varphi(t_q) = \varphi(t_q)$ $(0 \leqslant q \leqslant k)$.

Théorème. pour $\varphi \in \mathscr{C}^{k+1}([0,1])$,

$$|\varphi(t) - \pi_\varphi(t)| \leqslant \frac{\sup_{t\in[0,1]} \left|\varphi^{(k+1)}(t)\right|}{(k+1)!}.$$

- Conséquence : si $f \in \mathscr{C}^{k+1}[a,b]$,

$$\boxed{|I - I_N| \leqslant \frac{h^{k+1}(b-a)}{(k+1)!} \sup_{x\in[a,b]} \left|f^{(k+1)}(x)\right|.}$$

$$(M_e) \qquad \int_0^1 \varphi(t)\,\mathrm{d}t \simeq \sum_{q=0}^{k} w_q \varphi(t_q).$$

$$(FQ_c) \qquad I = \int_a^b f(x)\,\mathrm{d}x \simeq I_N = h \sum_{i=0}^{N-1} \sum_{q=0}^{k} w_q f(x_i + t_q h).$$

> **Théorème.** Si (??) est exacte pour tout $\varphi \in \mathbb{P}_\ell$ et si $f \in \mathscr{C}^{\ell+1}([a, b])$, alors
>
> $$|I - I_N| \leqslant \frac{h^{\ell+1}(b-a)}{(\ell+1)!} \sup_{x \in [a,b]} \left| f^{(\ell+1)}(x) \right|.$$

$$(M_e) \qquad \int_0^1 \varphi(t)\,\mathrm{d}t \simeq \sum_{q=0}^{k} w_q \varphi(t_q).$$

$$(FQ_c) \qquad I = \int_a^b f(x)\,\mathrm{d}x \simeq I_N = h \sum_{i=0}^{N-1} \sum_{q=0}^{k} w_q f(x_i + t_q h).$$

Théorème. Si (??) est exacte pour tout $\varphi \in \mathbb{P}_\ell$ et si $f \in \mathscr{C}^{\ell+1}([a,b])$, alors

$$|I - I_N| \leqslant \frac{h^{\ell+1}(b-a)}{(\ell+1)!} \sup_{x\in[a,b]} \left| f^{(\ell+1)}(x) \right|.$$

Remarque. La constante n'est pas optimale. . .

$$(M_e) \qquad \int_0^1 \varphi(t)\,\mathrm{d}t \simeq \sum_{q=0}^{k} w_q \varphi(t_q).$$

$$(FQ_c) \qquad I = \int_a^b f(x)\,\mathrm{d}x \simeq I_N = h \sum_{i=0}^{N-1} \sum_{q=0}^{k} w_q f(x_i + t_q h).$$

Théorème. Si (??) est exacte pour tout $\varphi \in \mathbb{P}_\ell$ et si $f \in \mathscr{C}^{\ell+1}([a, b])$, alors

$$|I - I_N| \leqslant \frac{h^{\ell+1}(b-a)}{(\ell+1)!} \sup_{x \in [a,b]} \left| f^{(\ell+1)}(x) \right|.$$

Remarque. La constante n'est pas optimale. . .

Question. $\boxed{\text{A priori, } \ell = k}$ $N(k+1)$ évaluations de f, précision en $\mathcal{O}\left(\frac{1}{N^{k+1}}\right)$.

$$(M_e) \qquad \int_0^1 \varphi(t)\,\mathrm{d}t \simeq \sum_{q=0}^{k} w_q \varphi(t_q).$$

$$(FQ_c) \qquad I = \int_a^b f(x)\,\mathrm{d}x \simeq I_N = h \sum_{i=0}^{N-1} \sum_{q=0}^{k} w_q f(x_i + t_q h).$$

Théorème. Si (??) est exacte pour tout $\varphi \in \mathbb{P}_\ell$ et si $f \in \mathscr{C}^{\ell+1}([a,b])$, alors

$$|I - I_N| \leqslant \frac{h^{\ell+1}(b-a)}{(\ell+1)!} \sup_{x\in[a,b]} \left| f^{(\ell+1)}(x) \right|.$$

Remarque. La constante n'est pas optimale. . .

Question. $\boxed{\text{A priori, } \ell = k}$ $N(k+1)$ évaluations de f, précision en $\mathcal{O}\left(\frac{1}{N^{k+1}}\right)$.

Peut-on faire mieux ?

- $\boxed{\text{Points de quadrature fixés : comment calculer les poids ?}}$

$$(M_e) \qquad \int_0^1 \varphi(t)\,\mathrm{d}t \simeq \sum_{q=0}^{k} w_q \varphi(t_q).$$

Optimalité des choix des points/poids de quadrature ?

- $\boxed{\text{Points de quadrature fixés : comment calculer les poids ?}}$

$$(M_e) \qquad \int_0^1 \varphi(t)\,\mathrm{d}t \simeq \sum_{q=0}^{k} w_q \varphi(t_q).$$

  - k et $(t_q)_{q=0,\dots,k}$ donnés.

Optimalité des choix des points/poids de quadrature ?

- $\boxed{\text{Points de quadrature fixés : comment calculer les poids ?}}$

$$(M_e) \qquad \int_0^1 \varphi(t)\,\mathrm{d}t \simeq \sum_{q=0}^{k} w_q \varphi(t_q).$$

- k et $(t_q)_{q=0,\dots,k}$ donnés.
- (**??**) exact sur $\mathbb{P}_k \iff$ (**??**) exact pour la base $(1, t, \dots, t^k)$,

$$\iff \forall i = 0, 1, \dots, k, \quad \sum_{q=0}^{k} w_q t_q^i = \frac{1}{i+1}.$$

(système linéaire de Vandermonde)

- $\boxed{\text{Points de quadrature fixés : comment calculer les poids ?}}$

$$(M_e) \qquad \int_0^1 \varphi(t)\,\mathrm{d}t \simeq \sum_{q=0}^{k} w_q \varphi(t_q).$$

  - k et $(t_q)_{q=0,\dots,k}$ donnés.
  - (**??**) exact sur $\mathbb{P}_k \iff$ (**??**) exact pour la base $(1, t, \dots, t^k)$,

$$\iff \forall i = 0, 1, \dots, k, \quad \sum_{q=0}^{k} w_q t_q^i = \frac{1}{i+1}.$$

(système linéaire de Vandermonde)

$\Longrightarrow$ les poids (w_q) sont uniquement déterminés !

Optimalité des choix des points/poids de quadrature ?

- Points de quadrature fixés : comment calculer les poids ?

$$(M_e) \qquad \int_0^1 \varphi(t)\,\mathrm{d}t \simeq \sum_{q=0}^{k} w_q \varphi(t_q).$$

  - k et $(t_q)_{q=0,\dots,k}$ donnés.
  - (**??**) exact sur $\mathbb{P}_k \iff$ (**??**) exact pour la base $(1, t, \dots, t^k)$,

$$\iff \forall i = 0, 1, \dots, k, \quad \sum_{q=0}^{k} w_q t_q^i = \frac{1}{i+1}.$$

(système linéaire de Vandermonde)

$\Longrightarrow$ les poids (w_q) sont uniquement déterminés !

Conclusion. Sauf « miracle », (**??**) est exact sur $\mathbb{P}_k$ seulement…

Optimalité des choix des points/poids de quadrature ?

- $\boxed{\text{Avec } k+1 \text{ points de quadrature‡® : peut-on être exact au delà de } \mathbb{P}_k \text{ ?}}$

$$(M_e) \qquad \int_0^1 \varphi(t)\,\mathrm{d}t \simeq \sum_{q=0}^{k} w_q \varphi(t_q).$$

Optimalité des choix des points/poids de quadrature ?

- $\boxed{\text{Avec } k+1 \text{ points de quadrature‡® : peut-on être exact au delà de } \mathbb{P}_k \text{ ?}}$

$$(M_e) \qquad \int_0^1 \varphi(t)\,\mathrm{d}t \simeq \sum_{q=0}^{k} w_q \varphi(t_q).$$

- **Remarque.** pour $p(t) = \prod_{q=0}^{k} (t - t_q)^2 \in \mathbb{P}_{2k+2}$,

$$\int_0^1 p(t)\,\mathrm{d}t > 0, \quad \sum_{q=0}^{k} w_q p(t_q) = 0.$$

Optimalité des choix des points/poids de quadrature ?

- $\boxed{\text{Avec } k+1 \text{ points de quadrature‡® : peut-on être exact au delà de } \mathbb{P}_k \text{ ?}}$

$$(M_e) \qquad \int_0^1 \varphi(t)\,\mathrm{d}t \simeq \sum_{q=0}^{k} w_q \varphi(t_q).$$

- **Remarque.** pour $p(t) = \prod_{q=0}^{k} (t - t_q)^2 \in \mathbb{P}_{2k+2}$,

$$\int_0^1 p(t)\,\mathrm{d}t > 0, \quad \sum_{q=0}^{k} w_q p(t_q) = 0.$$

$\Longrightarrow$ **(??)** ne peut pas être exact $\mathbb{P}_{2k+2}$!

Optimalité des choix des points/poids de quadrature ?

- $\boxed{\text{Avec } k+1 \text{ points de quadrature‡® : peut-on être exact au delà de } \mathbb{P}_k \text{ ?}}$

$$(M_e) \qquad \int_0^1 \varphi(t)\,\mathrm{d}t \simeq \sum_{q=0}^{k} w_q \varphi(t_q).$$

- **Remarque.** pour $p(t) = \prod_{q=0}^{k} (t - t_q)^2 \in \mathbb{P}_{2k+2}$,

$$\int_0^1 p(t)\,\mathrm{d}t > 0, \quad \sum_{q=0}^{k} w_q p(t_q) = 0.$$

$\Longrightarrow$ (??) ne peut pas être exact $\mathbb{P}_{2k+2}$!

- Peut-on espérer être exact $\mathbb{P}_{2k+1}$?

Intégration numérique

Optimalité des choix des points/poids de quadrature ?

$$(M_e) \qquad \int_0^1 \varphi(t)\,\mathrm{d}t \simeq \sum_{q=0}^{k} w_q \varphi(t_q).$$

- (**??**) exact sur $\mathbb{P}_{2k+1}$?

ÉCOLE CENTRALE LYON

Intégration numérique

Optimalité des choix des points/poids de quadrature ?

$$(M_e) \qquad \int_0^1 \varphi(t)\,\mathrm{d}t \simeq \sum_{q=0}^{k} w_q \varphi(t_q).$$

- (??) exact sur $\mathbb{P}_{2k+1}$? Soit $p \in \mathbb{P}_{2k+1}$ et $\pi(t) = \prod_{q=0}^{k}(t - t_q) \in \mathbb{P}_{k+1}$.

Intégration numérique

Optimalité des choix des points/poids de quadrature ?

$$(M_e) \qquad \int_0^1 \varphi(t)\,\mathrm{d}t \simeq \sum_{q=0}^{k} w_q \varphi(t_q).$$

- (??) exact sur $\mathbb{P}_{2k+1}$? Soit $p \in \mathbb{P}_{2k+1}$ et $\pi(t) = \prod_{q=0}^{k}(t - t_q) \in \mathbb{P}_{k+1}$.
- Division euclidienne : $p = \pi u + r$, avec $\mathrm{d}^\circ u \leqslant k$ et $\mathrm{d}^\circ r \leqslant k$.

Intégration numérique

Optimalité des choix des points/poids de quadrature ?

$$(M_e) \qquad \int_0^1 \varphi(t)\,\mathrm{d}t \simeq \sum_{q=0}^{k} w_q \varphi(t_q).$$

- (??) exact sur $\mathbb{P}_{2k+1}$? Soit $p \in \mathbb{P}_{2k+1}$ et $\pi(t) = \prod_{q=0}^{k}(t - t_q) \in \mathbb{P}_{k+1}$.
- Division euclidienne : $p = \pi u + r$, avec $d^\circ u \leqslant k$ et $d^\circ r \leqslant k$.

$$\sum_{q=0}^{k} w_q p(t_q) = \sum_{q=0}^{k} w_q r(t_q). \qquad \text{car}\, \pi(t_q) = 0.$$

Optimalité des choix des points/poids de quadrature ?

$$(M_e) \qquad \int_0^1 \varphi(t)\,\mathrm{d}t \simeq \sum_{q=0}^{k} w_q \varphi(t_q).$$

- (??) exact sur $\mathbb{P}_{2k+1}$? Soit $p \in \mathbb{P}_{2k+1}$ et $\pi(t) = \prod_{q=0}^{k}(t - t_q) \in \mathbb{P}_{k+1}$.
- Division euclidienne : $p = \pi u + r$, avec $\mathrm{d}^\circ u \leqslant k$ et $\mathrm{d}^\circ r \leqslant k$.

$$\sum_{q=0}^{k} w_q p(t_q) = \sum_{q=0}^{k} w_q r(t_q). \qquad \text{car} \, \pi(t_q) = 0.$$

$$\int_0^1 p(t)\,\mathrm{d}t = \int_0^1 \pi(t)u(t)\,\mathrm{d}t + \int_0^1 r(t)\,\mathrm{d}t.$$

Intégration numérique

Optimalité des choix des points/poids de quadrature ?

$$(M_e) \qquad \int_0^1 \varphi(t)\,\mathrm{d}t \simeq \sum_{q=0}^{k} w_q \varphi(t_q).$$

- (??) exact sur $\mathbb{P}_{2k+1}$? Soit $p \in \mathbb{P}_{2k+1}$ et $\pi(t) = \prod_{q=0}^{k}(t - t_q) \in \mathbb{P}_{k+1}$.
- Division euclidienne : $p = \pi u + r$, avec $\mathrm{d}^\circ u \leqslant k$ et $\mathrm{d}^\circ r \leqslant k$.

$$\sum_{q=0}^{k} w_q p(t_q) = \sum_{q=0}^{k} w_q r(t_q). \qquad \text{car} \pi(t_q) = 0.$$

$$\int_0^1 p(t)\,\mathrm{d}t = \int_0^1 \pi(t)u(t)\,\mathrm{d}t + \int_0^1 r(t)\,\mathrm{d}t.$$

- **Si** (??) est exact sur $\mathbb{P}_k$ **et** $\pi \perp \mathbb{P}_k$, **alors** (??) est exact sur $\mathbb{P}_{2k+1}$!

Optimalité des choix des points/poids de quadrature ?

$$(M_e) \qquad \int_0^1 \varphi(t)\,\mathrm{d}t \simeq \sum_{q=0}^{k} w_q \varphi(t_q).$$

- (??) exact sur $\mathbb{P}_{2k+1}$? Soit $p \in \mathbb{P}_{2k+1}$ et $\pi(t) = \prod_{q=0}^{k}(t - t_q) \in \mathbb{P}_{k+1}$.
- Division euclidienne : $p = \pi u + r$, avec $\mathrm{d}^\circ u \leqslant k$ et $\mathrm{d}^\circ r \leqslant k$.

$$\sum_{q=0}^{k} w_q p(t_q) = \sum_{q=0}^{k} w_q r(t_q). \qquad \text{car}\, \pi(t_q) = 0.$$

$$\int_0^1 p(t)\,\mathrm{d}t = \int_0^1 \pi(t)u(t)\,\mathrm{d}t + \int_0^1 r(t)\,\mathrm{d}t.$$

- **Si** (??) est exact sur $\mathbb{P}_k$ **et** $\pi \perp \mathbb{P}_k$, **alors** (??) est exact sur $\mathbb{P}_{2k+1}$!

$\Longrightarrow$ $\boxed{(t_q) = \text{racines du } (k+1)^{\mathrm{e}} \text{ polynôme orthogonal pour } \mathrm{L}^2(0,1).}$

ÉCOLE CENTRALE LYON

Théorème. Soit $(P_0, P_1, \ldots)$ famille de polynômes orthogonaux (avec $d^\circ P_i = i$) pour le produit scalaire

$$(\varphi, \psi) = \int_0^1 \varphi(t)\psi(t)\,\mathrm{d}t.$$

On note $(t_q)_{q=0,\ldots,k}$ les racines de P_{k+1}. On détermine les poids $(w_q)_{q=0,\ldots,k}$ t.q. (**??**) soit exact $\mathbb{P}_k$, i.e. solution de

$$\forall i = 0, 1, \ldots, k, \quad \sum_{q=0}^{k} w_q t_q^i = \frac{1}{i+1}.$$

Alors la formule

$$(M_e) \qquad \int_0^1 \varphi(t)\,\mathrm{d}t \simeq \sum_{q=0}^{k} w_q \varphi(t_q).$$

est exacte $\mathbb{P}_{2k+1}$. *(Méthode de Gauss-Legendre)*.

Intégration numérique

- **Exemple.** Méthode de Gauss-Legendre exacte $\mathbb{P}_5$.

$$\int_0^1 \varphi(t)\,\mathrm{d}t \simeq \frac{5}{18}\varphi\left(\frac{1}{2}-\frac{\sqrt{15}}{10}\right)+\frac{4}{9}\varphi\left(\frac{1}{2}\right)+\frac{5}{18}\varphi\left(\frac{1}{2}+\frac{\sqrt{15}}{10}\right).$$

La méthode composée correspondante est en $\mathcal{O}\left(N^{-6}\right)$.

Pour k grand, le calcul des points n'est pas explicite.

Intégration numérique

- **Exemple.** Méthode de Gauss-Legendre exacte $\mathbb{P}_5$.

$$\int_0^1 \varphi(t)\,\mathrm{d}t \simeq \frac{5}{18}\varphi\left(\frac{1}{2}-\frac{\sqrt{15}}{10}\right)+\frac{4}{9}\varphi\left(\frac{1}{2}\right)+\frac{5}{18}\varphi\left(\frac{1}{2}+\frac{\sqrt{15}}{10}\right).$$

La méthode composée correspondante est en $\mathcal{O}\left(N^{-6}\right)$.

Pour k grand, le calcul des points n'est pas explicite.

- **Remarque.** Si le *poids* $\omega \geqslant 0$ est tel que

$$\forall p \in \mathbb{P}, \quad \int_0^1 |p(t)|\omega(t)\,\mathrm{d}t < +\infty,$$

alors on peut faire de même pour le produit scalaire $\mathrm{L}^2(\omega(t)\,\mathrm{d}t)$.

Culture : la méthode de Monte-Carlo

Rappel. Si $X \hookrightarrow \mathbb{U}([a,b])$ et $f : [a,b] \to \mathbb{R}$, alors

$$\mathbb{E}\left[f(X)\right] = \frac{1}{b-a}\int_a^b f(x)\,\mathrm{d}x.$$

Si $(x_0, x_1, \ldots, x_{N-1})$ est une réalisation d'un échantillon issu de la loi de X, alors

$$\frac{b-a}{N}\sum_{i=0}^{N-1} f(x_i) \quad \text{approche} \quad \int_a^b f(x)\,\mathrm{d}x.$$

Intégration numérique

Culture : la méthode de Monte-Carlo

Rappel. Si $X \hookrightarrow \mathbb{U}([a,b])$ et $f : [a,b] \to \mathbb{R}$, alors

$$\mathbb{E}\left[f(X)\right] = \frac{1}{b-a} \int_a^b f(x)\,\mathrm{d}x.$$

Si $(x_0, x_1, \ldots, x_{N-1})$ est une réalisation d'un échantillon issu de la loi de X, alors

$$\frac{b-a}{N} \sum_{i=0}^{N-1} f(x_i) \quad \text{approche} \quad \int_a^b f(x)\,\mathrm{d}x.$$

- Sous `matlab` :

```matlab
% Rectangles a gauche
x=a+(b-a)*(0:N-1)/N;
Irect=(b-a)/N*sum(f(x));
```

```matlab
% Monte-Carlo
x=a+(b-a)*rand(N,1)/N;
Imc=(b-a)/N*sum(f(x));
```

- Méthode en $\mathcal{O}\left(N^{-\frac{1}{2}}\right)$, surtout utile en grande dimension.

COURS 3

Optimisation numérique

ÉCOLE CENTRALE LYON

Optimisation numérique

Optimiser : une démarche universelle

- Mécanique :

Optimiser : une démarche universelle

- Mécanique :
  - Équilibre d'un système minimisant une énergie

- Mécanique :
  - Équilibre d'un système minimisant une énergie
  - Structure qui maximise la résistance

Optimiser : une démarche universelle

- Mécanique :
  - Équilibre d'un système minimisant une énergie
  - Structure qui maximise la résistance
- Transport :

Optimiser : une démarche universelle

- Mécanique :
  - Équilibre d'un système minimisant une énergie
  - Structure qui maximise la résistance
- Transport :
  - Minimiser la distance, le temps de trajet, le coût,...

Optimiser : une démarche universelle

- Mécanique :
  - Équilibre d'un système minimisant une énergie
  - Structure qui maximise la résistance
- Transport :
  - Minimiser la distance, le temps de trajet, le coût,...
- Economie, gestion :

Optimiser : une démarche universelle

- Mécanique :
  - Équilibre d'un système minimisant une énergie
  - Structure qui maximise la résistance
- Transport :
  - Minimiser la distance, le temps de trajet, le coût,...
- Economie, gestion :
  - Maximiser le profit d'une entreprise

Optimiser : une démarche universelle

- Mécanique :
  - Équilibre d'un système minimisant une énergie
  - Structure qui maximise la résistance
- Transport :
  - Minimiser la distance, le temps de trajet, le coût,...
- Economie, gestion :
  - Maximiser le profit d'une entreprise
  - Minimiser les risques d'un placement boursier

Optimiser : une démarche universelle

- Mécanique :
  - Équilibre d'un système minimisant une énergie
  - Structure qui maximise la résistance
- Transport :
  - Minimiser la distance, le temps de trajet, le coût,...
- Economie, gestion :
  - Maximiser le profit d'une entreprise
  - Minimiser les risques d'un placement boursier
- Médecine :

Optimiser : une démarche universelle

- Mécanique :
  - Équilibre d'un système minimisant une énergie
  - Structure qui maximise la résistance
- Transport :
  - Minimiser la distance, le temps de trajet, le coût,...
- Economie, gestion :
  - Maximiser le profit d'une entreprise
  - Minimiser les risques d'un placement boursier
- Médecine :
  - Optimiser une thérapie sous contraintes de dosage

Optimisation numérique

Optimiser : une démarche universelle

- Mécanique :
  - Équilibre d'un système minimisant une énergie
  - Structure qui maximise la résistance
- Transport :
  - Minimiser la distance, le temps de trajet, le coût,...
- Economie, gestion :
  - Maximiser le profit d'une entreprise
  - Minimiser les risques d'un placement boursier
- Médecine :
  - Optimiser une thérapie sous contraintes de dosage
- Imagerie, problèmes inverses :

Optimiser : une démarche universelle

- Mécanique :
  - Équilibre d'un système minimisant une énergie
  - Structure qui maximise la résistance
- Transport :
  - Minimiser la distance, le temps de trajet, le coût,...
- Economie, gestion :
  - Maximiser le profit d'une entreprise
  - Minimiser les risques d'un placement boursier
- Médecine :
  - Optimiser une thérapie sous contraintes de dosage
- Imagerie, problèmes inverses :
  - Restaurer une image abimée par minimisation de variation

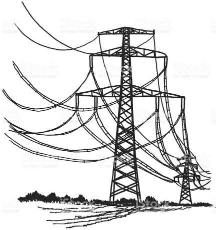
Optimiser : une démarche universelle

- Mécanique :
  - Équilibre d'un système minimisant une énergie
  - Structure qui maximise la résistance
- Transport :
  - Minimiser la distance, le temps de trajet, le coût,...
- Economie, gestion :
  - Maximiser le profit d'une entreprise
  - Minimiser les risques d'un placement boursier
- Médecine :
  - Optimiser une thérapie sous contraintes de dosage
- Imagerie, problèmes inverses :
  - Restaurer une image abimée par minimisation de variation
  - Résoudre des équations de forme $\varphi(x) = y$ en minimisant

$$f(x) = \|\varphi(x) - y\|^2$$

Optimisation numérique

Position d'un câble soumis à la gravité

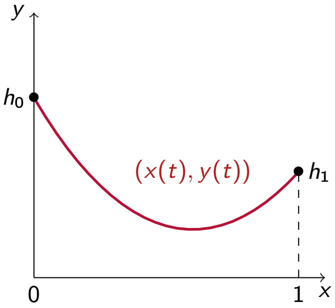


Quelle est la position exacte, la tension (mécanique), la distance au sol... des lignes HT ?

Idem pour les cables d'un téléphérique ?

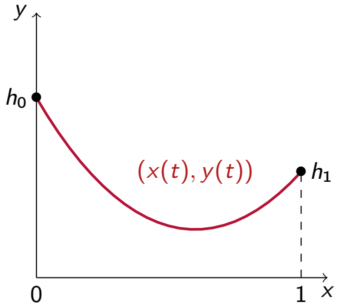
Les cables **minimisent** *leur énergie potentielle*

Modélisation : câbles HT



- Inconnue : arc paramètré $(x(t), y(t))$
- Contraintes :
  - longueur totale L,
  - extrémités fixées.

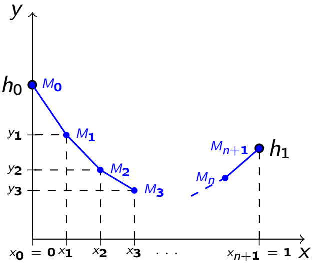
Modélisation : câbles HT



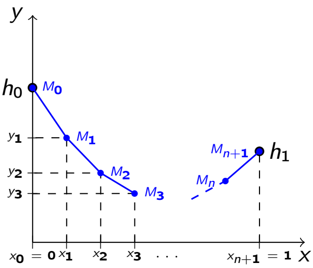
- Inconnue : arc paramètré $(x(t), y(t))$
- Contraintes :
  - longueur totale L,
  - extrémités fixées.

Théorème de l'énergie potentielle.

Le cable est à l'équilibre s'il minimise son énergie potentielle

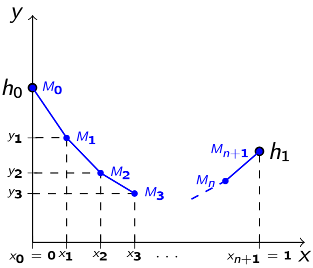


Discrétisation de l'espace



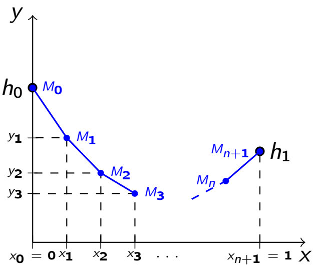
Discrétisation de l'espace

- Câble affine par morceaux



Discrétisation de l'espace

- Câble affine par morceaux
- $M_i = (x_i, y_i)$.



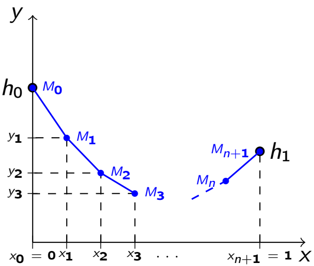
Discrétisation de l'espace

- Câble affine par morceaux
- $M_i = (x_i, y_i)$.
- On note

$$\mathrm{x} = (x_1, \dots, x_n), \;\; \mathrm{y} = (y_1, \dots, y_n).$$

$$\text{et } y_0 = h_0, \;\; y_{n+1} = h_1.$$

Modélisation du câble HT : discrétisation



Discrétisation de l'espace

- Câble affine par morceaux
- $M_i = (x_i, y_i)$.
- On note

$$\mathrm{x} = (x_1, \dots, x_n), \ \ \mathrm{y} = (y_1, \dots, y_n).$$

$$\text{et } y_0 = h_0, \ \ y_{n+1} = h_1.$$

Problème discret : $\boxed{\min \left\{ \mathrm{f}(\mathrm{x},\mathrm{y}) \ ; \ (\mathrm{x},\mathrm{y}) \in K \right\}}$

avec

- $\mathrm{f}(\mathrm{x},\mathrm{y}) = \displaystyle\sum_{i=1}^{n} y_i$
- $K = \left\{ (\mathrm{x},\mathrm{y}) \in \mathbb{R}^{2n} \ ; \ (x_{i+1} - x_i)^2 + (y_{i+1} - y_i)^2 = \frac{L^2}{(n+1)^2}, \ \forall i = 0 \dots n \right\}.$

ÉCOLE CENTRALE LYON

Problème d'optimisation.

$$\boxed{\min_{u \in K} f(u)}$$

avec $K \subset \mathbb{R}^N$ et $f : K \to \mathbb{R}$.

Problème d'optimisation.

$$\boxed{\min_{\mathbf{u}\in K} \mathrm{f}(\mathbf{u})}$$

avec $K \subset \mathbb{R}^N$ et $\mathrm{f} : K \to \mathbb{R}$.

Vocabulaire :

- f : **fonction objectif** ou **énergie**.

Problème d'optimisation.

$$\boxed{\min_{u \in K} f(u)}$$

avec $K \subset \mathbb{R}^N$ et $f : K \to \mathbb{R}$.

Vocabulaire :

- f : **fonction objectif** ou **énergie**.
- $\mathbb{R}^N$: **espace des paramètres**.

Problème d'optimisation.

$$\boxed{\min_{u \in K} f(u)}$$

avec $K \subset \mathbb{R}^N$ et $f : K \to \mathbb{R}$.

Vocabulaire :

- f : **fonction objectif** ou **énergie**.
- $\mathbb{R}^N$: **espace des paramètres**.
- K : **ensemble admissible**.

Problème d'optimisation.

$$\boxed{\min_{u \in K} f(u)}$$

avec $K \subset \mathbb{R}^N$ et $f : K \to \mathbb{R}$.

Vocabulaire :

- f : **fonction objectif** ou **énergie**.
- $\mathbb{R}^N$: **espace des paramètres**.
- K : **ensemble admissible**.
- Solution : point $u^* \in K$ qui minimise f sur K ; on écrit

$$f(u^*) = \min_{u \in K} f(u) \quad \text{et aussi} \quad u^* = \underset{u \in K}{\operatorname{argmin}}\, f(u)$$

Problème d'optimisation.

$$\boxed{\min_{u \in K} f(u)}$$

avec $K \subset \mathbb{R}^N$ et $f : K \to \mathbb{R}$.

Vocabulaire :

- f : **fonction objectif** ou **énergie**.
- $\mathbb{R}^N$: **espace des paramètres**.
- K : **ensemble admissible**.
- Solution : point $u^* \in K$ qui minimise f sur K ; on écrit

$$f(u^*) = \min_{u \in K} f(u) \quad \text{et aussi} \quad u^* = \operatorname*{argmin}_{u \in K} f(u)$$

- $f(u^*)$: **minimum** de f sur K.

Problème d'optimisation.

$$\boxed{\min_{u \in K} f(u)}$$

avec $K \subset \mathbb{R}^N$ et $f : K \to \mathbb{R}$.

Vocabulaire :

- f : **fonction objectif** ou **énergie**.
- $\mathbb{R}^N$: **espace des paramètres**.
- K : **ensemble admissible**.
- Solution : point $u^* \in K$ qui minimise f sur K ; on écrit

$$f(u^*) = \min_{u \in K} f(u) \quad \text{et aussi} \quad u^* = \operatorname*{argmin}_{u \in K} f(u)$$

- $f(u^*)$: **minimum** de f sur K.

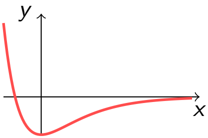
Attention : Il peut y avoir zéro, une ou plusieurs solutions au problème.

Optimisation libre dans $\mathbb{R}$

> **Définition.** Une fonction $f : \mathbb{R} \mapsto \mathbb{R}$ est **unimodale** s'il existe $x^* \in \mathbb{R}$ tel que f soit strictement décroissante sur $]-\infty, x^*[$ et strictement croissante sur $]x^*, +\infty[$.

Définition. Une fonction $f : \mathbb{R} \mapsto \mathbb{R}$ est **unimodale** s'il existe $x^* \in \mathbb{R}$ tel que f soit strictement décroissante sur $]-\infty, x^*[$ et strictement croissante sur $]x^*, +\infty[$.

Exemple : $x \mapsto e^{-2x} - 2e^{-x}$



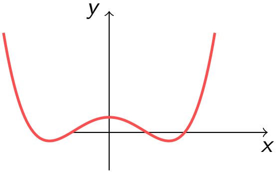
Définition. Une fonction $f : \mathbb{R} \mapsto \mathbb{R}$ est **unimodale** s'il existe $x^* \in \mathbb{R}$ tel que f soit strictement décroissante sur $]-\infty, x^*[$ et strictement croissante sur $]x^*, +\infty[$.

Définition. Une fonction $f : \mathbb{R} \mapsto \mathbb{R}$ est **coercive** si $\lim_{|x| \to +\infty} f(x) = +\infty$.

Définition. Une fonction $f : \mathbb{R} \mapsto \mathbb{R}$ est **unimodale** s'il existe $x^* \in \mathbb{R}$ tel que f soit strictement décroissante sur $]-\infty, x^*[$ et strictement croissante sur $]x^*, +\infty[$.

Définition. Une fonction $f : \mathbb{R} \mapsto \mathbb{R}$ est **coercive** si $\lim_{|x| \to +\infty} f(x) = +\infty$.

Exemple : $x \mapsto x^4 - 5x^3 + 4$



Définition. Une fonction $f : \mathbb{R} \mapsto \mathbb{R}$ est **unimodale** s'il existe $x^* \in \mathbb{R}$ tel que f soit strictement décroissante sur $]-\infty, x^*[$ et strictement croissante sur $]x^*, +\infty[$.

Définition. Une fonction $f : \mathbb{R} \mapsto \mathbb{R}$ est **coercive** si $\lim_{|x| \to +\infty} f(x) = +\infty$.

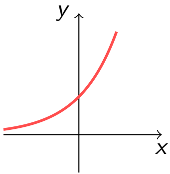
Définition. Une fonction $f : \mathbb{R} \mapsto \mathbb{R}$ de classe $\mathscr{C}^2$ est **fortement convexe** ssi $\exists \alpha > 0$ tel que pour tout $x \in \mathbb{R}$, $f''(x) \geqslant \alpha$. (On dit aussi **α-convexe**).

Définition. Une fonction $f : \mathbb{R} \mapsto \mathbb{R}$ est **unimodale** s'il existe $x^* \in \mathbb{R}$ tel que f soit strictement décroissante sur $]-\infty, x^*[$ et strictement croissante sur $]x^*, +\infty[$.

Définition. Une fonction $f : \mathbb{R} \mapsto \mathbb{R}$ est **coercive** si $\lim_{|x| \to +\infty} f(x) = +\infty$.

Définition. Une fonction $f : \mathbb{R} \mapsto \mathbb{R}$ de classe $\mathscr{C}^2$ est **fortement convexe** ssi $\exists \alpha > 0$ tel que pour tout $x \in \mathbb{R}$, $f''(x) \geqslant \alpha$. (On dit aussi **α-convexe**).

Exemple : $x \mapsto e^x$ strictement convexe, mais pas fortement convexe.



Définition. Une fonction $f : \mathbb{R} \mapsto \mathbb{R}$ est **unimodale** s'il existe $x^* \in \mathbb{R}$ tel que f soit strictement décroissante sur $]-\infty, x^*[$ et strictement croissante sur $]x^*, +\infty[$.

Définition. Une fonction $f : \mathbb{R} \mapsto \mathbb{R}$ est **coercive** si $\lim_{|x| \to +\infty} f(x) = +\infty$.

Définition. Une fonction $f : \mathbb{R} \mapsto \mathbb{R}$ de classe $\mathscr{C}^2$ est **fortement convexe** ssi $\exists \alpha > 0$ tel que pour tout $x \in \mathbb{R}$, $f''(x) \geqslant \alpha$. (On dit aussi α**-convexe**).

Proposition.

- fortement convexe $\Rightarrow$ coercive
- fortement convexe et coercive $\Rightarrow$ unimodale

But : résoudre $F(x) = 0$

- $F : \mathbb{R} \to \mathbb{R}$ **croissante** et **continue**, s'annule en $x^* \in [a_0, b_0]$.

But : résoudre $F(x) = 0$

- $F : \mathbb{R} \to \mathbb{R}$ **croissante** et **continue**, s'annule en $x^* \in [a_0, b_0]$.
- On construit (a_n), (b_n) par l'algorithme

On pose $x_n = \dfrac{a_n + b_n}{2}$.

- Si $F(x_n) \leq 0$,
$$a_{n+1} = x_n, \quad \text{et} \quad b_{n+1} = b_n.$$
- Si $F(x_n) > 0$,
$$a_{n+1} = a_n, \quad \text{et} \quad b_{n+1} = x_n.$$

But : résoudre $F(x) = 0$

- $F : \mathbb{R} \to \mathbb{R}$ **croissante** et **continue**, s'annule en $x^* \in [a_0, b_0]$.
- On construit (a_n), (b_n) par l'algorithme

On pose $x_n = \dfrac{a_n + b_n}{2}$.

- Si $F(x_n) \leq 0$,
$$a_{n+1} = x_n, \quad \text{et} \quad b_{n+1} = b_n.$$
- Si $F(x_n) > 0$,
$$a_{n+1} = a_n, \quad \text{et} \quad b_{n+1} = x_n.$$

Proposition. La méthode converge à vitesse géométrique :

$$\forall n \in \mathbb{N}, \quad |x_n - x^*| \leq \frac{b_0 - a_0}{2^{n+1}}.$$

But : résoudre $\min f(x)$

- $f : \mathbb{R} \to \mathbb{R}$ **unimodale** et **dérivable**, minimale en $x^* \in [a_0, b_0]$.

But : résoudre $\min f(x)$

- $f : \mathbb{R} \to \mathbb{R}$ **unimodale** et **dérivable**, minimale en $x^* \in [a_0, b_0]$.
- On construit (a_n), (b_n) par l'algorithme

On pose $x_n = \dfrac{a_n + b_n}{2}$.

- Si $f'(x_n) \leq 0$,
$$a_{n+1} = x_n, \quad \text{et} \quad b_{n+1} = b_n.$$
- Si $f'(x_n) > 0$,
$$a_{n+1} = a_n, \quad \text{et} \quad b_{n+1} = x_n.$$

But : résoudre $\min f(x)$

- $f : \mathbb{R} \to \mathbb{R}$ **unimodale** et **dérivable**, minimale en $x^* \in [a_0, b_0]$.
- On construit (a_n), (b_n) par l'algorithme

On pose $x_n = \dfrac{a_n + b_n}{2}$.

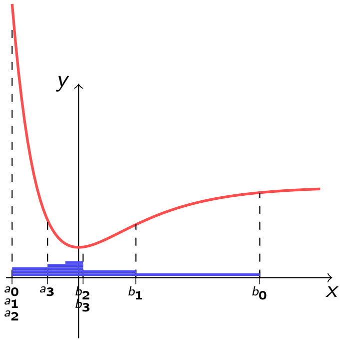
- Si $f'(x_n) \leq 0$,
$$a_{n+1} = x_n, \quad \text{et} \quad b_{n+1} = b_n.$$
- Si $f'(x_n) > 0$,
$$a_{n+1} = a_n, \quad \text{et} \quad b_{n+1} = x_n.$$

Proposition. La méthode converge à vitesse géométrique :

$$\forall n \in \mathbb{N}, \quad |x_n - x^*| \leq \frac{b_0 - a_0}{2^{n+1}}.$$

Optimisation numérique

Méthode de dichotomie 2



Avantages :

- Très simple à mettre en œuvre.
- Convergence rapide.

Inconvénient :

- Évaluation de f'.

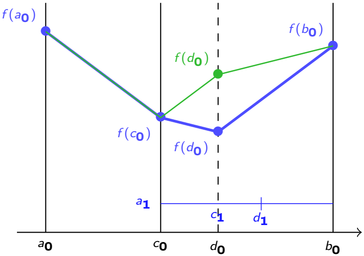
Coût : À chaque itération, on évalue seulement une fois la dérivée de f.

[*DichoGold.m*]

Optimisation numérique

Méthode du nombre d'or

Soit $f : \mathbb{R} \to \mathbb{R}$ **unimodale**, minimale en $x^* \in [a_0, b_0]$.



Principe : On évalue f en les 4 points

$$a_0 < c_0 < d_0 < b_0.$$

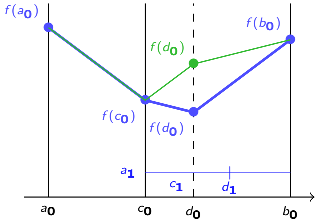
Deux cas :

- Si $f(d_0) < f(c_0)$ alors $x^* \in [c_0, b_0]$
- Si $f(d_0) > f(c_0)$ alors $x^* \in [a_0, d_0]$

- La taille de l'intervalle diminue à chaque étape.
- Coût : À chaque itération, une seule nouvelle évaluation de f.

[DichoGold.m]

Quel rapport avec le nombre d'or ?



- Par symétrie on impose
$$b_0 - c_0 = d_0 - a_0.$$
- Pour conserver le ratio des intervalles
$$\gamma = \frac{b_0 - a_0}{b_0 - c_0} = \frac{b_1 - a_1}{b_1 - c_1} = \frac{b_0 - c_0}{b_0 - d_0}$$

$$\frac{1}{\gamma} = \frac{b_0 - d_0}{b_0 - c_0} = \gamma - \frac{a_0 - d_0}{b_0 - c_0} = \gamma - 1$$

Donc $\gamma^2 - \gamma - 1 = 0$ et ainsi

$$\boxed{\gamma = \frac{1+\sqrt{5}}{2}}$$

- $f : \mathbb{R} \to \mathbb{R}$ **unimodale**, minimale en $x^* \in [a_0, b_0]$.
- On construit (a_n), (b_n) par l'algorithme

Si $f(c_n) \leq f(d_n)$, $$\left\lvert \begin{array}{lcl} a_{n+1} & = & a_n \\ b_{n+1} & = & d_n \\ c_{n+1} & = & b_{n+1} - \frac{b_{n+1} - a_{n+1}}{\gamma} \\ d_{n+1} & = & c_n \end{array} \right.$$	Si $f(c_n) > f(d_n)$, $$\left\lvert \begin{array}{lcl} a_{n+1} & = & c_n \\ b_{n+1} & = & b_n \\ c_{n+1} & = & d_n \\ d_{n+1} & = & a_{n+1} + \frac{b_{n+1} - a_{n+1}}{\gamma} \end{array} \right.$$

Proposition. La méthode converge à vitesse géométrique :

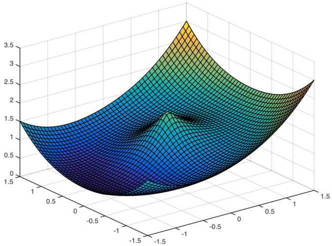
$$\forall n \in \mathbb{N}, \quad \left| \frac{a_n + b_n}{2} - x^* \right| \leq \frac{1}{2}(b_0 - a_0)\gamma^{-n}.$$

ÉCOLE CENTRALE LYON

Optimisation libre dans $\mathbb{R}^N$

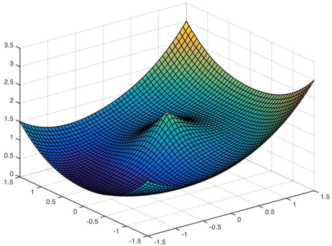
Exemple :

$$f(x) = \left(\sqrt{x_1^2 + x_2^2} - 1\right)^2 + \frac{x_1}{2} + 1.$$



Exemple :

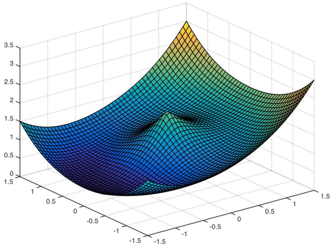
$$f(x) = \left(\sqrt{x_1^2 + x_2^2} - 1\right)^2 + \frac{x_1}{2} + 1.$$



À partir d'un point initial $x^0 \in \mathbb{R}^N$,

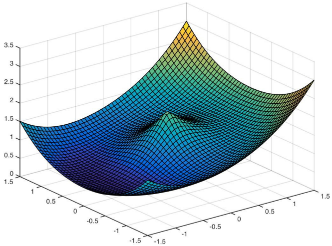
Exemple :

$$f(x) = \left(\sqrt{x_1^2 + x_2^2} - 1\right)^2 + \frac{x_1}{2} + 1.$$



À partir d'un point initial $x^0 \in \mathbb{R}^N$,

- on choisit une direction de descente $d^0 \in \mathbb{R}^N$ et un pas $\rho_0 > 0$,



Exemple :

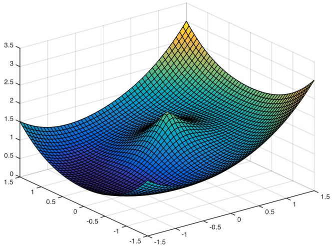
$$f(x) = \left(\sqrt{x_1^2 + x_2^2} - 1\right)^2 + \frac{x_1}{2} + 1.$$

À partir d'un point initial $x^0 \in \mathbb{R}^N$,

- on choisit une direction de descente $d^0 \in \mathbb{R}^N$ et un pas $\rho_0 > 0$,
- on obtient un nouveau point $x^1 = x^0 + \rho_0 d^0$,

Exemple :

$$f(x) = \left(\sqrt{x_1^2 + x_2^2} - 1\right)^2 + \frac{x_1}{2} + 1.$$



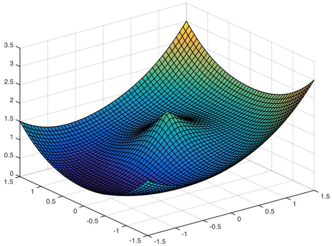
À partir d'un point initial $x^0 \in \mathbb{R}^N$,

- on choisit une direction de descente $d^0 \in \mathbb{R}^N$ et un pas $\rho_0 > 0$,
- on obtient un nouveau point $x^1 = x^0 + \rho_0 d^0$,
- on itère le processus : $\boxed{x^{n+1} = x^n + \rho_n d^n}$

Méthodes de descentes : principe

Exemple :

$$f(x) = \left(\sqrt{x_1^2 + x_2^2} - 1\right)^2 + \frac{x_1}{2} + 1.$$



À partir d'un point initial $x^0 \in \mathbb{R}^N$,

- on choisit une direction de descente $d^0 \in \mathbb{R}^N$ et un pas $\rho_0 > 0$,
- on obtient un nouveau point $x^1 = x^0 + \rho_0 d^0$,
- on itère le processus : $\boxed{x^{n+1} = x^n + \rho_n d^n}$

But. faire en sorte que $f(x^{n+1}) \leqslant f(x^n)$.

[*Grad.m*]

Comment choisir la direction d^n ?

- On veut $(f(x^n))_n$ décroissante.

Comment choisir la direction d^n ?

- On veut $(f(x^n))_n$ décroissante.
- On choisit d^n telle que $\varphi : t \mapsto f(x^n + td^n)$ soit décroissante au voisinage de 0.

Comment choisir la direction d^n ?

- On veut $(f(x^n))_n$ décroissante.
- On choisit d^n telle que $\varphi : t \mapsto f(x^n + td^n)$ soit décroissante au voisinage de 0.
- Or $\varphi'(0) = (\nabla f(x^n) \,|\, d^n)$.

Comment choisir la direction d^n ?

- On veut $(f(x^n))_n$ décroissante.
- On choisit d^n telle que $\varphi : t \mapsto f(x^n + td^n)$ soit décroissante au voisinage de 0.
- Or $\varphi'(0) = (\nabla f(x^n) \,|\, d^n)$.
- Un choix simple est
$$d^n = -\nabla f(x^n).$$

Comment choisir la direction d^n ?

- On veut $(f(x^n))_n$ décroissante.
- On choisit d^n telle que $\varphi : t \mapsto f(x^n + td^n)$ soit décroissante au voisinage de 0.
- Or $\varphi'(0) = (\nabla f(x^n) \,|\, d^n)$.
- Un choix simple est
$$d^n = -\nabla f(x^n).$$

Méthode du gradient à pas fixe. Elle est définie par

$$\begin{cases} x^0 \in \mathbb{R}^N, \\ x^{n+1} = x^n - \rho \nabla f(x^n). \end{cases}$$

avec $\rho > 0$.

Théorème. Soit $f \in \mathscr{C}^1(\mathbb{R}^N, \mathbb{R})$ strictement convexe et coercive telle que

$$(*) \qquad \exists M > 0, \quad \forall x, y \in \mathbb{R}^N, \quad \|\nabla f(x) - \nabla f(y)\|_2 \leq M \|x - y\|_2.$$

Si $0 < \rho < \frac{2}{M}$ alors la méthode du gradient à pas fixe converge.

La limite est l'unique point de minimum global de f.

Théorème. Soit $f \in \mathscr{C}^1(\mathbb{R}^N, \mathbb{R})$ strictement convexe et coercive telle que

$$(*) \qquad \exists M > 0, \quad \forall x, y \in \mathbb{R}^N, \quad \|\nabla f(x) - \nabla f(y)\|_2 \leq M \|x - y\|_2.$$

Si $0 < \rho < \frac{2}{M}$ alors la méthode du gradient à pas fixe converge.

La limite est l'unique point de minimum global de f.

Remarques :

- Condition **(??)** : $x \mapsto \nabla f(x)$ est globalement M-lipschitzienne sur $\mathbb{R}^N$.
- Si on ajoute l'hypothèse (f est fortement convexe)

$$\exists \alpha > 0, \quad \forall x, y \in \mathbb{R}^N, \quad (Hf(x)y|y) \geq \alpha \|y\|^2.$$

et $\rho < \dfrac{2\alpha}{M^2}$ alors la convergence est géométrique.

Pour des fonctions de classe $\mathscr{C}^2$

$f : \mathbb{R} \to \mathbb{R}$	$f : \mathbb{R}^N \to \mathbb{R}$	
$f'' \geq 0$	$\forall x, y \in \mathbb{R}^N, \ (Hf(x)y \mid y) \geq 0$	convexe
$f'' > 0$	$\forall x, y \in \mathbb{R}^N, \ y \neq 0, \ (Hf(x)y \mid y) > 0$	$\Longrightarrow$ strictement convexe
$f'' \geq \alpha$	$\forall x, y \in \mathbb{R}^N, \ (Hf(x)y \mid y) \geq \alpha \left\Vert y \right\Vert_2^2$	fortement convexe ($\alpha > 0$)

Peut-on faire mieux ?

- On peut chercher le meilleur pas possible à chaque étape :

$$\rho_n \quad \text{minimise} \quad \rho \mapsto f(x^n - \rho \nabla f(x^n)).$$

Vocabulaire : cette étape est dite *recherche linéaire*.

Méthode : type dichotomie.

Peut-on faire mieux ?

- On peut chercher le meilleur pas possible à chaque étape :

$$\rho_n \quad \text{minimise} \quad \rho \mapsto f(x^n - \rho \nabla f(x^n)).$$

Vocabulaire : cette étape est dite *recherche linéaire*.

Méthode : type dichotomie.

Méthode du gradient à pas optimal. Elle est définie par

$$\begin{cases} x^0 \in \mathbb{R}^N, \\ \rho_n = \arg\min_{\rho \in \mathbb{R}} f(x^n - \rho \nabla f(x^n)) \\ x^{n+1} = x^n - \rho_n \nabla f(x^n). \end{cases}$$

Peut-on faire mieux ?

- On peut chercher le meilleur pas possible à chaque étape :

$$\rho_n \quad \text{minimise} \quad \rho \mapsto f(x^n - \rho \nabla f(x^n)).$$

Vocabulaire : cette étape est dite *recherche linéaire*.

Méthode : type dichotomie.

Méthode du gradient à pas optimal. Elle est définie par

$$\begin{cases} x^0 \in \mathbb{R}^N, \\ \rho_n = \arg\min_{\rho \in \mathbb{R}} f(x^n - \rho \nabla f(x^n)) \\ x^{n+1} = x^n - \rho_n \nabla f(x^n). \end{cases}$$

Remarque. même type de résultats de convergence que pour le pas fixe.

Exemple : Minimisation de $f(x) = \|Ax - b\|_2^2$.

- On parle de résolution de $Ax = b$ *au sens des moindres carrés*.

Exemple : Minimisation de $f(x) = \|Ax - b\|_2^2$.

- On parle de résolution de $Ax = b$ *au sens des moindres carrés*.
- Éventuellement, A est rectangulaire (injective).

Exemple : Minimisation de $f(x) = \|Ax - b\|_2^2$.

- On parle de résolution de $Ax = b$ *au sens des moindres carrés*.
- Éventuellement, A est rectangulaire (injective).
- Lien avec l'approximation polynomiale.

Exemple : Minimisation de $f(x) = \|Ax - b\|_2^2$.

- On parle de résolution de $Ax = b$ *au sens des moindres carrés*.
- Éventuellement, A est rectangulaire (injective).
- Lien avec l'approximation polynomiale.
- Rappel : $\nabla f(x) = 2A^T(Ax - b)$

Exemple : Minimisation de $f(x) = \|Ax - b\|_2^2$.

- On parle de résolution de $Ax = b$ *au sens des moindres carrés*.
- Éventuellement, A est rectangulaire (injective).
- Lien avec l'approximation polynomiale.
- Rappel : $\nabla f(x) = 2A^T(Ax - b)$

Pour trouver le pas optimal, il faut résoudre

$$\min_{\rho > 0} f(x - \rho \nabla f(x)).$$

Exemple : Minimisation de $f(x) = \|Ax - b\|_2^2$.

- On parle de résolution de $Ax = b$ *au sens des moindres carrés*.
- Éventuellement, A est rectangulaire (injective).
- Lien avec l'approximation polynomiale.
- Rappel : $\nabla f(x) = 2A^T(Ax - b)$

Pour trouver le pas optimal, il faut résoudre

$$\min_{\rho > 0} f(x - \rho \nabla f(x)).$$

$$f(x - \rho \nabla f(x)) = \|Ax - b - \rho A \nabla f(x)\|_2^2$$

Exemple : Minimisation de $f(x) = \|Ax - b\|_2^2$.

- On parle de résolution de $Ax = b$ *au sens des moindres carrés*.
- Éventuellement, A est rectangulaire (injective).
- Lien avec l'approximation polynomiale.
- Rappel : $\nabla f(x) = 2A^T(Ax - b)$

Pour trouver le pas optimal, il faut résoudre

$$\min_{\rho > 0} f(x - \rho \nabla f(x)).$$

$$f(x - \rho \nabla f(x)) = \|Ax - b\|_2^2 - 2\rho(Ax - b|A\nabla f(x)) + \rho^2 \|A\nabla f(x)\|_2^2$$

Exemple : Minimisation de $f(x) = \|Ax - b\|_2^2$.

- On parle de résolution de $Ax = b$ *au sens des moindres carrés*.
- Éventuellement, A est rectangulaire (injective).
- Lien avec l'approximation polynomiale.
- Rappel : $\nabla f(x) = 2A^T(Ax - b)$

Pour trouver le pas optimal, il faut résoudre

$$\min_{\rho>0} f(x - \rho\nabla f(x)).$$

$$f(x - \rho\nabla f(x)) = \|Ax - b\|_2^2 - 2\rho(A^T Ax - A^T b|\nabla f(x)) + \rho^2 \|A\nabla f(x)\|_2^2$$

Exemple : Minimisation de $f(x) = \|Ax - b\|_2^2$.

- On parle de résolution de $Ax = b$ *au sens des moindres carrés*.
- Éventuellement, A est rectangulaire (injective).
- Lien avec l'approximation polynomiale.
- Rappel : $\nabla f(x) = 2A^T(Ax - b)$

Pour trouver le pas optimal, il faut résoudre

$$\min_{\rho>0} f(x - \rho\nabla f(x)).$$

$$f(x - \rho\nabla f(x)) = \rho^2 \|A\nabla f(x)\|_2^2 - \rho \|\nabla f(x)\|_2^2 + \|Ax - b\|_2^2$$

Exemple : Minimisation de $f(x) = \|Ax - b\|_2^2$.

- On parle de résolution de $Ax = b$ *au sens des moindres carrés*.
- Éventuellement, A est rectangulaire (injective).
- Lien avec l'approximation polynomiale.
- Rappel : $\nabla f(x) = 2A^T(Ax - b)$

Pour trouver le pas optimal, il faut résoudre

$$\min_{\rho>0} f(x - \rho\nabla f(x)).$$

$$f(x - \rho\nabla f(x)) = \rho^2 \|A\nabla f(x)\|_2^2 - \rho \|\nabla f(x)\|_2^2 + \|Ax - b\|_2^2$$

C'est un polynôme de degré 2 en ρ !

Exemple : Minimisation de $f(x) = \|Ax - b\|_2^2$.

- On parle de résolution de $Ax = b$ *au sens des moindres carrés*.
- Éventuellement, A est rectangulaire (injective).
- Lien avec l'approximation polynomiale.
- Rappel : $\nabla f(x) = 2A^T(Ax - b)$

Pour trouver le pas optimal, il faut résoudre

$$\min_{\rho>0} f(x - \rho\nabla f(x)).$$

$$f(x - \rho\nabla f(x)) = \rho^2 \|A\nabla f(x)\|_2^2 - \rho \|\nabla f(x)\|_2^2 + \|Ax - b\|_2^2$$

C'est un polynôme de degré 2 en ρ !

$$\boxed{\rho^* = \frac{\|\nabla f(x)\|_2^2}{2\|A\nabla f(x)\|_2^2}.}$$

- Résolution d'une équation $F(x) = 0$ avec $F \in \mathscr{C}^1(\mathbb{R}^N, \mathbb{R}^N)$.

Méthode de Newton pour les zeros de F

$$\begin{cases} x^0 \in \mathbb{R}^N, \\ x^{n+1} = x^n - JF(x^n)^{-1} F(x^n). \end{cases}$$

- Résolution d'une équation $F(x) = 0$ avec $F \in \mathscr{C}^1(\mathbb{R}^N, \mathbb{R}^N)$.

> **Méthode de Newton pour les zeros de F**
>
> $$\begin{cases} x^0 \in \mathbb{R}^N, \\ x^{n+1} = x^n - JF(x^n)^{-1}F(x^n). \end{cases}$$

$\boxed{\text{Faire un dessin 1D}}$

- Résolution d'une équation $F(x) = 0$ avec $F \in \mathscr{C}^1(\mathbb{R}^N, \mathbb{R}^N)$.

> **Méthode de Newton pour les zeros de F**
>
> $$\begin{cases} x^0 \in \mathbb{R}^N, \\ x^{n+1} = x^n - JF(x^n)^{-1}F(x^n). \end{cases}$$

- Minimisation de $f \in \mathscr{C}^2(\mathbb{R}^N, \mathbb{R})$. On résout $\nabla f(x) = 0$.

> **Méthode de Newton pour minimiser f**
>
> $$\begin{cases} x^0 \in \mathbb{R}^N, \\ x^{n+1} = x^n - Hf(x^n)^{-1}\nabla f(x^n). \end{cases}$$

- Résolution d'une équation $F(x) = 0$ avec $F \in \mathscr{C}^1(\mathbb{R}^N, \mathbb{R}^N)$.

> **Méthode de Newton pour les zeros de F**
>
> $$\begin{cases} x^0 \in \mathbb{R}^N, \\ x^{n+1} = x^n - JF(x^n)^{-1}F(x^n). \end{cases}$$

- Minimisation de $f \in \mathscr{C}^2(\mathbb{R}^N, \mathbb{R})$. On résout $\nabla f(x) = 0$.

> **Méthode de Newton pour minimiser f**
>
> $$\begin{cases} x^0 \in \mathbb{R}^N, \\ x^{n+1} = x^n - Hf(x^n)^{-1}\nabla f(x^n). \end{cases}$$

Remarques :

- La méthode est **locale**, mais rapide,
- Ne distingue pas min et max.

Optimisation sous contraintes

Exemple simple

Rappel : Un problème d'optimisation sous contraintes s'écrit

$$\min_{\mathbf{u}\in K} \mathrm{f}(\mathbf{u}), \qquad \text{avec} \quad K \subsetneq \mathbb{R}^N.$$

Rappel : Un problème d'optimisation sous contraintes s'écrit

$$\min_{\mathbf{u} \in K} \mathrm{f}(\mathbf{u}), \qquad \text{avec} \quad K \subsetneq \mathbb{R}^N.$$

En général, les algorithmes précédents vont converger **en dehors de K** !

Optimisation numérique

Exemple simple

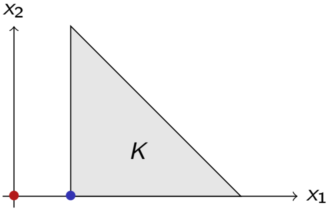
Rappel : Un problème d'optimisation sous contraintes s'écrit

$$\min_{u \in K} f(u), \qquad \text{avec} \quad K \subsetneq \mathbb{R}^N.$$

En général, les algorithmes précédents vont converger **en dehors de K** !

Exemple : $f(x) = x_1^2 + x_2^2$.

$$K = \left\{ x \in \mathbb{R}^2,\ x_1 \geq \tfrac{1}{2},\ x_2 \geq 0,\ x_1 + x_2 \leq 1 \right\}$$



[*GradProj.m*]

Soit $K \subset \mathbb{R}^N$ convexe fermé et non vide.

> **Définition.** Pour tout $x \in \mathbb{R}^N$, le projeté $\pi_K(x)$ est l'unique point de K qui minimise sa distance à x.
>
> $$\|\pi_K(x) - x\| = \min\left\{\|y - x\| \; ; \; y \in K\right\}.$$

Soit $K \subset \mathbb{R}^N$ convexe fermé et non vide.

Définition. Pour tout $x \in \mathbb{R}^N$, le projeté $\pi_K(x)$ est l'unique point de K qui minimise sa distance à x.

$$\|\pi_K(x) - x\| = \min \left\{ \|y - x\| \; ; \; y \in K \right\}.$$

Remarques :

- si $x \in K$ alors $\pi_K(x) = x$,
- si $x \notin K$ alors $\pi_K(x) \in \partial K$,
- π_K est contractante donc lipschitzienne donc continue.

Méthode du gradient projeté. Elle est définie par

$$\begin{cases} x^0 \in \mathbb{R}^N, \\ x^{n+1} = \pi_K\left(x^n - \rho \nabla f(x^n)\right). \end{cases}$$

avec $\rho > 0$.

Méthode du gradient projeté. Elle est définie par

$$\begin{cases} x^0 \in \mathbb{R}^N, \\ x^{n+1} = \pi_K\left(x^n - \rho \nabla f(x^n)\right). \end{cases}$$

avec $\rho > 0$.

Théorème. Soient $f \in \mathscr{C}^1(\mathbb{R}^N, \mathbb{R})$ strictement convexe et coercive, et K convexe fermé non vide, avec

$$\exists M > 0, \quad \forall x, y \in \mathbb{R}^N, \quad \|\nabla f(x) - \nabla f(y)\|_2 \leq M \|x - y\|_2 .$$

Si $0 < \rho < \frac{2}{M}$ alors la méthode du gradient projeté converge vers l'unique point de minimum global de f sur K.

Avantage :

- convergence vers la solution du problème contraint.

Avantage :

- convergence vers la solution du problème contraint.

Inconvénients :

- vitesse de convergence non garantie,
- le projecteur π_K peut être (très) difficile à calculer,

Avantage :

- convergence vers la solution du problème contraint.

Inconvénients :

- vitesse de convergence non garantie,
- le projecteur π_K peut être (très) difficile à calculer,

Alternative possible :

- Méthode de pénalisation.

Définition. On appelle fonction de pénalisation de K toute fonction $\beta : \mathbb{R}^N \to \mathbb{R}$ telle que

- β est continue
- $\beta \geq 0$ sur $\mathbb{R}^N$,
- $\beta(x) = 0 \Leftrightarrow x \in K$

Définition. On appelle fonction de pénalisation de K toute fonction $\beta : \mathbb{R}^N \to \mathbb{R}$ telle que

- β est continue
- $\beta \geq 0$ sur $\mathbb{R}^N$,
- $\beta(x) = 0 \Leftrightarrow x \in K$

Remarque : Si c'est possible on choisira β convexe. (Il faut K convexe).

Définition. On appelle fonction de pénalisation de K toute fonction $\beta : \mathbb{R}^N \to \mathbb{R}$ telle que

- β est continue
- $\beta \geq 0$ sur $\mathbb{R}^N$,
- $\beta(x) = 0 \Leftrightarrow x \in K$

Remarque : Si c'est possible on choisira β convexe. (Il faut K convexe).

Définition. Soit le problème d'optimisation sous contraintes

$$\min_{x \in K} f(x),$$

le problème pénalisé associé est

$$\min_{x \in \mathbb{R}^N} f(x) + \frac{1}{\varepsilon}\beta(x).$$

Définition. On appelle fonction de pénalisation de K toute fonction $\beta : \mathbb{R}^N \to \mathbb{R}$ telle que

- β est continue
- $\beta \geq 0$ sur $\mathbb{R}^N$,
- $\beta(x) = 0 \Leftrightarrow x \in K$

Remarque : Si c'est possible on choisira β convexe. (Il faut K convexe).

Définition. Soit le problème d'optimisation sous contraintes

$$\min_{x \in K} f(x),$$

le problème pénalisé associé est

$$\min_{x \in \mathbb{R}^N} f(x) + \frac{1}{\varepsilon}\beta(x).$$

Pourquoi ca marche ? On fait payer (*pénalise*) le fait de ne pas être dans K.

- Si $K_1 = \left\{ x \in \mathbb{R}^N,\ \varphi_i(x) = 0,\ i = 1 \ldots p \right\}$, on peut choisir

$$\beta(x) = \sum_{i=1}^{p} [\varphi_i(x)]^2 \,.$$

- Si $K_1 = \left\{ x \in \mathbb{R}^N, \ \varphi_i(x) = 0, \ i = 1 \dots p \right\}$, on peut choisir

$$\beta(x) = \sum_{i=1}^{p} [\varphi_i(x)]^2 .$$

- Si $K_2 = \left\{ x \in \mathbb{R}^N, \ \psi_j(x) \leq 0, \ i = 1 \dots q \right\}$, on peut choisir

$$\beta(x) = \sum_{j=1}^{q} \left[\psi_j^+(x)\right]^2 .$$

avec $\psi_j^+ = \max(0, \psi_j)$.

- Si $K_1 = \left\{x \in \mathbb{R}^N,\ \varphi_i(x) = 0,\ i = 1 \dots p\right\}$, on peut choisir

$$\beta(x) = \sum_{i=1}^{p} [\varphi_i(x)]^2 .$$

- Si $K_2 = \left\{x \in \mathbb{R}^N,\ \psi_j(x) \leq 0,\ i = 1 \dots q\right\}$, on peut choisir

$$\beta(x) = \sum_{j=1}^{q} \left[\psi_j^+(x)\right]^2 .$$

avec $\psi_j^+ = \max(0, \psi_j)$.

- Si $K = K_1 \cap K_2$, on somme :

$$\beta(x) = \sum_{i=1}^{p} [\varphi_i(x)]^2 + \sum_{j=1}^{q} \left[\psi_j^+(x)\right]^2 .$$

On résout le problème pénalisé (sans contrainte !)

$$\min_{x\in\mathbb{R}^N} f(x) + \frac{1}{\varepsilon}\beta(x).$$

par la méthode du gradient à pas fixe.

On résout le problème pénalisé (sans contrainte !)

$$\min_{x\in\mathbb{R}^N} f(x) + \frac{1}{\varepsilon}\beta(x).$$

par la méthode du gradient à pas fixe.

Méthode du gradient pénalisé. Elle est définie par

$$\begin{cases} x^0 \in \mathbb{R}^N, \\ x^{n+1} = x^n - \rho\nabla f(x^n) - \frac{\rho}{\varepsilon}\nabla\beta(x). \end{cases}$$

avec $\rho > 0$.

On résout le problème pénalisé (sans contrainte !)

$$\min_{x\in\mathbb{R}^N} f(x) + \frac{1}{\varepsilon}\beta(x).$$

par la méthode du gradient à pas fixe.

Méthode du gradient pénalisé. Elle est définie par

$$\begin{cases} x^0 \in \mathbb{R}^N, \\ x^{n+1} = x^n - \rho\nabla f(x^n) - \frac{\rho}{\varepsilon}\nabla\beta(x). \end{cases}$$

avec $\rho > 0$.

Remarque :

- Choix de ε difficile !
- Conditionne le choix de ρ...

Méthode du gradient pénalisé

Difficulté : La méthode du gradient pénalisé va converger vers x_ε solution de

$$\min_{\mathbf{u}\in\mathbb{R}^N} f(x) + \frac{1}{\varepsilon}\beta(x).$$

Si ε et petit, on s'attend a ce que x_ε soit proche de x^*, solution de

$$\min_{x\in K} f(x).$$

Difficulté : La méthode du gradient pénalisé va converger vers x_ε solution de

$$\min_{u \in \mathbb{R}^N} f(x) + \frac{1}{\varepsilon}\beta(x).$$

Si ε et petit, on s'attend a ce que x_ε soit proche de x^*, solution de

$$\min_{x \in K} f(x).$$

Théorème. Soient $f \in \mathscr{C}^1(\mathbb{R}^N, \mathbb{R})$ strictement convexe et coercive, et $K \subset \mathbb{R}^N$ convexe fermé non vide. Soit β une pénalisation de K. Alors pour tout $\varepsilon > 0$, le problème

$$\min_{x \in \mathbb{R}^N} f(x) + \frac{1}{\varepsilon}\beta(x)$$

admet une unique solution x_ε. Elle vérifie

$$\lim_{\varepsilon \to 0} x_\varepsilon = x^*$$

où x^* est l'unique solution du problème initial $\min_{x \in K} f(x)$.

COURS 4

Approximation numérique des équations différentielles

- Mécanique :

Approximation des EDO

Quelques usages

- Mécanique :
  - Prédiction de trajectoire (planètes, comètes, astéroïdes, débris,. . .)

Approximation des EDO

Quelques usages

- Mécanique :
  - Prédiction de trajectoire (planètes, comètes, astéroïdes, débris,. . .)
  - Orbite des satellites, manœuvre spatiales, rentrée atmosphérique,

Approximation des EDO

Quelques usages

- Mécanique :
  - Prédiction de trajectoire (planètes, comètes, astéroïdes, débris,. . .)
  - Orbite des satellites, manœuvre spatiales, rentrée atmosphérique,
  - Militaire (balistique),

Quelques usages

- Mécanique :
  - Prédiction de trajectoire (planètes, comètes, astéroïdes, débris,...)
  - Orbite des satellites, manœuvre spatiales, rentrée atmosphérique,
  - Militaire (balistique),
- Biologie, environnement, :

Approximation des EDO

Quelques usages

- Mécanique :
  - Prédiction de trajectoire (planètes, comètes, astéroïdes, débris,. . .)
  - Orbite des satellites, manœuvre spatiales, rentrée atmosphérique,
  - Militaire (balistique),
- Biologie, environnement, :
  - Evolution d'un écosystème, d'un population, croissance et extinction,

- Mécanique :
  - Prédiction de trajectoire (planètes, comètes, astéroïdes, débris,. . .)
  - Orbite des satellites, manœuvre spatiales, rentrée atmosphérique,
  - Militaire (balistique),
- Biologie, environnement, :
  - Evolution d'un écosystème, d'un population, croissance et extinction,
- Chimie, industrie :

- Mécanique :
  - Prédiction de trajectoire (planètes, comètes, astéroïdes, débris,. . .)
  - Orbite des satellites, manœuvre spatiales, rentrée atmosphérique,
  - Militaire (balistique),
- Biologie, environnement, :
  - Evolution d'un écosystème, d'un population, croissance et extinction,
- Chimie, industrie :
  - Dynamique des processus,

Approximation des EDO

Quelques usages

- Mécanique :
  - Prédiction de trajectoire (planètes, comètes, astéroïdes, débris,...)
  - Orbite des satellites, manœuvre spatiales, rentrée atmosphérique,
  - Militaire (balistique),
- Biologie, environnement, :
  - Evolution d'un écosystème, d'un population, croissance et extinction,
- Chimie, industrie :
  - Dynamique des processus,
- Economie, gestion :

- Mécanique :
  - Prédiction de trajectoire (planètes, comètes, astéroïdes, débris,. . .)
  - Orbite des satellites, manœuvre spatiales, rentrée atmosphérique,
  - Militaire (balistique),
- Biologie, environnement, :
  - Evolution d'un écosystème, d'un population, croissance et extinction,
- Chimie, industrie :
  - Dynamique des processus,
- Economie, gestion :
  - Modélisation macroéconomique,

Quelques usages

- Mécanique :
  - Prédiction de trajectoire (planètes, comètes, astéroïdes, débris,. . .)
  - Orbite des satellites, manœuvre spatiales, rentrée atmosphérique,
  - Militaire (balistique),
- Biologie, environnement, :
  - Evolution d'un écosystème, d'un population, croissance et extinction,
- Chimie, industrie :
  - Dynamique des processus,
- Economie, gestion :
  - Modélisation macroéconomique,
  - Modèles financiers,

Approximation des EDO

Quelques usages

- Mécanique :
  - Prédiction de trajectoire (planètes, comètes, astéroïdes, débris,. . .)
  - Orbite des satellites, manœuvre spatiales, rentrée atmosphérique,
  - Militaire (balistique),
- Biologie, environnement, :
  - Evolution d'un écosystème, d'un population, croissance et extinction,
- Chimie, industrie :
  - Dynamique des processus,
- Economie, gestion :
  - Modélisation macroéconomique,
  - Modèles financiers,
- Médecine :

Approximation des EDO

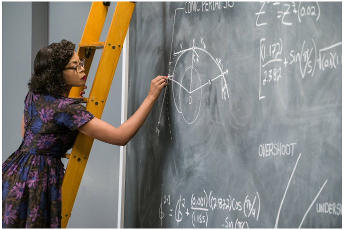
Quelques usages

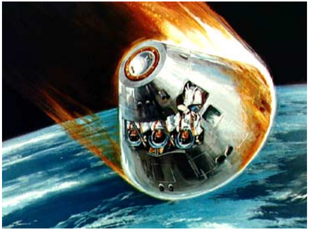
- Mécanique :
  - Prédiction de trajectoire (planètes, comètes, astéroïdes, débris,...)
  - Orbite des satellites, manœuvre spatiales, rentrée atmosphérique,
  - Militaire (balistique),
- Biologie, environnement, :
  - Evolution d'un écosystème, d'un population, croissance et extinction,
- Chimie, industrie :
  - Dynamique des processus,
- Economie, gestion :
  - Modélisation macroéconomique,
  - Modèles financiers,
- Médecine :
  - Posologie, absorption et efficacité des traitements, croissance tumorales,

- Mécanique :
  - Prédiction de trajectoire (planètes, comètes, astéroïdes, débris,. . .)
  - Orbite des satellites, manœuvre spatiales, rentrée atmosphérique,
  - Militaire (balistique),
- Biologie, environnement, :
  - Evolution d'un écosystème, d'un population, croissance et extinction,
- Chimie, industrie :
  - Dynamique des processus,
- Economie, gestion :
  - Modélisation macroéconomique,
  - Modèles financiers,
- Médecine :
  - Posologie, absorption et efficacité des traitements, croissance tumorales,
  - Epidémiologie.

Approximation des EDO

Une des premières utilisations célèbre





Mercury-Atlas 6 : comment calculer une trajectoire de rentrée atmosphérique ?

Comment viser le Kazakstan avec un Soyouz ?

Pourquoi seule l'analyse numérique peut résoudre le problème ?

Quelles méthodes ?

Quelle précision ?

Important : en général, **on ne sait pas résoudre explicitement une EDO !**

Approximation des EDO

Constat d'échec

Important : en général, **on ne sait pas résoudre explicitement une EDO !**

On s'en sort quand. . .

Approximation des EDO

Constat d'échec

Important : en général, **on ne sait pas résoudre explicitement une EDO !**

On s'en sort quand. . .

- l'équation différentielle est **linéaire** à coefficients explicitement intégrables,

Approximation des EDO

Constat d'échec

Important : en général, **on ne sait pas résoudre explicitement une EDO !**

On s'en sort quand...

- l'équation différentielle est **linéaire** à coefficients explicitement intégrables,
- l'exercice est fait pour qu'on y arrive,

Important : en général, **on ne sait pas résoudre explicitement une EDO !**

On s'en sort quand. . .

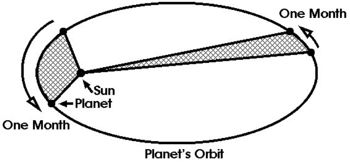
- l'équation différentielle est **linéaire** à coefficients explicitement intégrables,
- l'exercice est fait pour qu'on y arrive,
- on a beaucoup de **chance**.

Important : en général, **on ne sait pas résoudre explicitement une EDO !**

On s'en sort quand. . .

- l'équation différentielle est **linéaire** à coefficients explicitement intégrables,
- l'exercice est fait pour qu'on y arrive,
- on a beaucoup de **chance**.

Exemple de coup de chance : les orbites de planètes sont des ellipses.



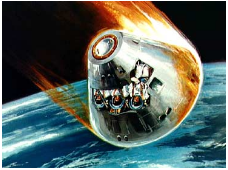
(Kepler)

Équation gravitationelle :

$$x'' = -Gm_S \frac{x}{\|x\|^3}.$$

Approximation des EDO

Modélisation rentrée atmosphérique



Modélisation du problème de rentrée atmosphérique :

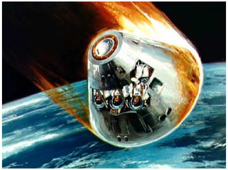
- Loi de Newton :

$$mx'' = \text{ force de gravité} + \text{force de frottement}$$

$$mx'' = -\frac{Gmm_T}{\|x\|^3}x - \alpha(x)\,\|x'\|\,x'$$

Approximation des EDO

Modélisation rentrée atmosphérique



Modélisation du problème de rentrée atmosphérique :

- Loi de Newton :

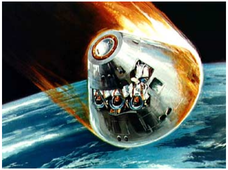
$$mx'' = \text{ force de gravité} + \text{force de frottement}$$

$$mx'' = -\frac{Gmm_T}{\|x\|^3}x - \alpha(x)\,\|x'\|\,x'$$

- Inconnue : x : $[0, T] \to \mathbb{R}^2$ ou $\mathbb{R}^3$

Approximation des EDO

Modélisation rentrée atmosphérique



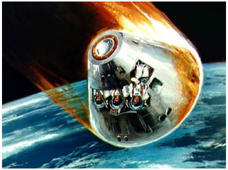
Modélisation du problème de rentrée atmosphérique :

- Loi de Newton :

$$mx'' = \text{ force de gravité} + \text{force de frottement}$$

$$mx'' = -\frac{Gmm_T}{\|x\|^3}x - \alpha(x)\,\|x'\|\,x'$$

- Inconnue : x : $[0, T] \to \mathbb{R}^2$ ou $\mathbb{R}^3$
- Domaine : pour tout $t \in [0, T]$, $x(t) \in \Omega := \{x \in \mathbb{R}^d, \|x\| > R_T\}$.



Modélisation du problème de rentrée atmosphérique :

- Loi de Newton :

$$mx'' = \text{ force de gravité} + \text{force de frottement}$$

$$mx'' = -\frac{Gmm_T}{\|x\|^3}x - \alpha(x)\,\|x'\|\,x'$$

- Inconnue : x : $[0, T] \to \mathbb{R}^2$ ou $\mathbb{R}^3$
- Domaine : pour tout $t \in [0, T]$, $x(t) \in \Omega := \{x \in \mathbb{R}^d, \|x\| > R_T\}$.
- Données : G cste de gravité, m masse du vaisseau, m_T masse de la Terre, $\alpha(x)$ coefficient de frottement aérodynamique, R_T rayon de la Terre.

Méthode : on met l'équation sous forme d'un problème de Cauchy.

Problème de Cauchy. C'est un problème différentiel de la forme

$$(\mathcal{C}) : \begin{cases} \mathrm{u}'(t) = \mathrm{f}(t, \mathrm{u}(t)) \quad t \in [0, T], \\ \mathrm{u}(0) = \mathrm{u}_0. \end{cases}$$

Méthode : on met l'équation sous forme d'un problème de Cauchy.

Problème de Cauchy. C'est un problème différentiel de la forme

$$(\mathcal{C}) : \begin{cases} \mathrm{u}'(t) = \mathrm{f}(t, \mathrm{u}(t)) & t \in [0, T], \\ \mathrm{u}(0) = \mathrm{u}_0. \end{cases}$$

- $\mathrm{u} : [0, T] \to \Omega \subset \mathbb{R}^d$ est la **fonction inconnue** ,

Méthode : on met l'équation sous forme d'un problème de Cauchy.

Problème de Cauchy. C'est un problème différentiel de la forme

$$(\mathcal{C}) : \begin{cases} \mathrm{u}'(t) = \mathrm{f}(t, \mathrm{u}(t)) \quad t \in [0, T], \\ \mathrm{u}(0) = \mathrm{u}_0. \end{cases}$$

- $\mathrm{u} : [0, T] \to \Omega \subset \mathbb{R}^d$ est la **fonction inconnue** ,
- le **domaine spatial** Ω est une partie ouverte connexe de l'espace $\mathbb{R}^d$,

Méthode : on met l'équation sous forme d'un problème de Cauchy.

Problème de Cauchy. C'est un problème différentiel de la forme

$$(\mathcal{C}) : \begin{cases} \mathrm{u}'(t) = \mathrm{f}(t, \mathrm{u}(t)) & t \in [0, T], \\ \mathrm{u}(0) = \mathrm{u}_0. \end{cases}$$

- $\mathrm{u} : [0, T] \to \Omega \subset \mathbb{R}^d$ est la **fonction inconnue** ,
- le **domaine spatial** Ω est une partie ouverte connexe de l'espace $\mathbb{R}^d$,
- $\mathrm{f} : \Omega \times [0, T] \mapsto \mathbb{R}^d$ est la **dynamique** du problème,

Méthode : on met l'équation sous forme d'un problème de Cauchy.

Problème de Cauchy. C'est un problème différentiel de la forme

$$(\mathcal{C}) : \begin{cases} \mathrm{u}'(t) = \mathrm{f}(t, \mathrm{u}(t)) \quad t \in [0, T], \\ \mathrm{u}(0) = \mathrm{u}_0. \end{cases}$$

- $\mathrm{u} : [0, T] \to \Omega \subset \mathbb{R}^d$ est la **fonction inconnue** ,
- le **domaine spatial** Ω est une partie ouverte connexe de l'espace $\mathbb{R}^d$,
- $\mathrm{f} : \Omega \times [0, T] \mapsto \mathbb{R}^d$ est la **dynamique** du problème,
- $\mathrm{u}_0 \in \Omega$ est l'**état initial** où condition initiale.

Exemple : rentrée atmosphérique,

$$\mathrm{x}'' = -\frac{Gm_T}{\|\mathrm{x}\|^3}x - \frac{\alpha(\mathrm{x})}{m}\,\|\mathrm{x}'\|\,\mathrm{x}'$$

Exemple : rentrée atmosphérique,

$$\mathrm{x}'' = -\frac{Gm_T}{\|\mathrm{x}\|^3} x - \frac{\alpha(\mathrm{x})}{m} \|\mathrm{x}'\| \, \mathrm{x}'$$

on pose $\mathrm{u} := \begin{bmatrix} u_1 \\ u_2 \\ u_3 \\ u_4 \end{bmatrix} := \begin{bmatrix} \mathrm{x} \\ \mathrm{x}' \end{bmatrix} = \begin{bmatrix} x_1 \\ x_2 \\ x_1' \\ x_2' \end{bmatrix} \in \mathbb{R}^4$ et on a

$$\mathrm{f}(\mathrm{u}) = \begin{bmatrix} u_3 \\ u_4 \\ -\frac{Gm_T}{\|(u_1, u_2)\|^3} \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} - \frac{\alpha(u_1, u_2)}{m} \|(u_3, u_4)\| \begin{bmatrix} u_3 \\ u_4 \end{bmatrix} \end{bmatrix}.$$

Exemple : rentrée atmosphérique,

$$\mathrm{x}'' = -\frac{Gm_T}{\|\mathrm{x}\|^3}x - \frac{\alpha(\mathrm{x})}{m}\|\mathrm{x}'\|\,\mathrm{x}'$$

on pose $\mathrm{u} := \begin{bmatrix} u_1 \\ u_2 \\ u_3 \\ u_4 \end{bmatrix} := \begin{bmatrix} \mathrm{x} \\ \mathrm{x}' \end{bmatrix} = \begin{bmatrix} x_1 \\ x_2 \\ x_1' \\ x_2' \end{bmatrix} \in \mathbb{R}^4$ et on a

$$\mathrm{f}(\mathrm{u}) = \begin{bmatrix} u_3 \\ u_4 \\ -\frac{Gm_T}{\|(u_1, u_2)\|^3}\begin{bmatrix} u_1 \\ u_2 \end{bmatrix} - \frac{\alpha(u_1, u_2)}{m}\|(u_3, u_4)\|\begin{bmatrix} u_3 \\ u_4 \end{bmatrix} \end{bmatrix}.$$

Le problème de Cauchy s'écrit

$$\begin{cases} \mathrm{u}'(t) = \mathrm{f}(\mathrm{u}(t)) & t \in [0, T], \\ \mathrm{u}(0) = \begin{bmatrix} \mathrm{x}_0 \\ \mathrm{v}_0 \end{bmatrix}. \end{cases}$$

Approximation des EDO

Comment approcher numériquement la solution du problème de Cauchy ?

Soit le problème de Cauchy

$$(\mathcal{C}) : \begin{cases} \mathrm{u}'(t) = \mathrm{f}(t, \mathrm{u}(t)) \quad t \in [0, T], \\ \mathrm{u}(0) = \mathrm{u}_0. \end{cases}$$

Comment approcher numériquement la solution du problème de Cauchy ?

Soit le problème de Cauchy

$$(\mathcal{C}) : \begin{cases} \mathrm{u}'(t) = \mathrm{f}(t, \mathrm{u}(t)) \quad t \in [0, T], \\ \mathrm{u}(0) = \mathrm{u}_0. \end{cases}$$

Existence et unicité de $u : [0, T^*[\to \Omega$ dès que $\mathrm{f} \in \mathscr{C}^1$ [Cauchy-Lipschitz].

Approximation des EDO

Comment approcher numériquement la solution du problème de Cauchy ?

Soit le problème de Cauchy

$$(\mathcal{C}) : \begin{cases} \mathrm{u}'(t) = \mathrm{f}(t, \mathrm{u}(t)) \quad t \in [0, T], \\ \mathrm{u}(0) = \mathrm{u}_0. \end{cases}$$

Existence et unicité de $u : [0, T^*[\to \Omega$ dès que $\mathrm{f} \in \mathscr{C}^1$ [Cauchy-Lipschitz].

Discrétisation du temps : soit $h > 0$, et par $t_n = nh$ avec $t_N = N_h h = T$.

ÉCOLE CENTRALE LYON

Approximation des EDO

Comment approcher numériquement la solution du problème de Cauchy ?

Soit le problème de Cauchy

$$(\mathcal{C}) : \begin{cases} \mathrm{u}'(t) = \mathrm{f}(t, \mathrm{u}(t)) \quad t \in [0, T], \\ \mathrm{u}(0) = \mathrm{u}_0. \end{cases}$$

Existence et unicité de $u : [0, T^*[\to \Omega$ dès que $\mathrm{f} \in \mathscr{C}^1$ [Cauchy-Lipschitz].

Discrétisation du temps : soit $h > 0$, et par $t_n = nh$ avec $t_N = N_h h = T$.

Cette suite dépend de h, on la note donc aussi $(t_n^h)_{n=0,1,\dots,N}$.

Approximation des EDO

Comment approcher numériquement la solution du problème de Cauchy ?

Soit le problème de Cauchy

$$(\mathcal{C}) : \begin{cases} \mathrm{u}'(t) = \mathrm{f}(t, \mathrm{u}(t)) & t \in [0, T], \\ \mathrm{u}(0) = \mathrm{u}_0. \end{cases}$$

Existence et unicité de $u : [0, T^*[\to \Omega$ dès que $\mathrm{f} \in \mathscr{C}^1$ [Cauchy-Lipschitz].

Discrétisation du temps : soit $h > 0$, et par $t_n = nh$ avec $t_N = N_h h = T$.
Cette suite dépend de h, on la note donc aussi $(t_n^h)_{n=0,1,\ldots,N}$.

Principe d'une méthode numérique : approcher les valeurs de la solution exacte $\mathrm{u}(t_n)$ par une suite $(\mathrm{U}_n)_{n=0,1,\ldots,N}$ de Ω.

Approximation des EDO

Comment approcher numériquement la solution du problème de Cauchy ?

Soit le problème de Cauchy

$$(\mathcal{C}) : \begin{cases} \mathrm{u}'(t) = \mathrm{f}(t, \mathrm{u}(t)) & t \in [0, T], \\ \mathrm{u}(0) = \mathrm{u}_0. \end{cases}$$

Existence et unicité de $u : [0, T^*[\to \Omega$ dès que $\mathrm{f} \in \mathscr{C}^1$ [Cauchy-Lipschitz].

Discrétisation du temps : soit $h > 0$, et par $t_n = nh$ avec $t_N = N_h h = T$.

Cette suite dépend de h, on la note donc aussi $(t_n^h)_{n=0,1,\dots,N}$.

Principe d'une méthode numérique : approcher les valeurs de la solution exacte $\mathrm{u}(t_n)$ par une suite $(\mathrm{U}_n)_{n=0,1,\dots,N}$ de Ω.

Vocabulaire : Un **schéma numérique** est une relation de récurrence qui permet de définir la suite (U_n).

Approximation des EDO

Comment approcher numériquement la solution du problème de Cauchy ?

Soit le problème de Cauchy

$$(\mathcal{C}) : \begin{cases} \mathrm{u}'(t) = \mathrm{f}(t, \mathrm{u}(t)) \quad t \in [0, T], \\ \mathrm{u}(0) = \mathrm{u}_0. \end{cases}$$

Existence et unicité de $u : [0, T^*[\to \Omega$ dès que $\mathrm{f} \in \mathscr{C}^1$ [Cauchy-Lipschitz].

Discrétisation du temps : soit $h > 0$, et par $t_n = nh$ avec $t_N = N_h h = T$.

Cette suite dépend de h, on la note donc aussi $(t_n^h)_{n=0,1,\ldots,N}$.

Principe d'une méthode numérique : approcher les valeurs de la solution exacte $\mathrm{u}(t_n)$ par une suite $(\mathrm{U}_n)_{n=0,1,\ldots,N}$ de Ω.

Vocabulaire : Un **schéma numérique** est une relation de récurrence qui permet de définir la suite (U_n).

Cette suite est dite solution numérique ou discrète.

Problème de Cauchy

$$(\mathcal{C}) : \begin{cases} \mathrm{u}'(t) = \mathrm{f}(t, \mathrm{u}(t)) & t \in [0, T], \\ \mathrm{u}(0) = \mathrm{u}_0. \end{cases}$$

- On écrit, pour $n = 0, 1, \ldots, N_h - 1$,

$$\mathrm{u}(t_{n+1}) = \mathrm{u}(t_n) + \int_{t_n}^{t_{n+1}} u'(s)\mathrm{d}s = \mathrm{u}(t_n) + \int_{t_n}^{t_{n+1}} \mathrm{f}(s, \mathrm{u}(s))\mathrm{d}s.$$

- Approximation de l'intégrale par la méthode des rectangles :

$$\int_{t_n}^{t_{n+1}} \mathrm{f}(s, \mathrm{u}(s))\mathrm{d}s \approx h\, \mathrm{f}(t_n, \mathrm{u}(t_n)).$$

- D'où la récurrence (explicite mais inexacte) :

$$\mathrm{u}(t_{n+1}) \approx \mathrm{u}(t_n) + h\, \mathrm{f}(t_n, \mathrm{u}(t_n)).$$

On a obtenu :

$$\mathrm{u}(t_{n+1}) \approx \mathrm{u}(t_n) + h\,\mathrm{f}(t_n, \mathrm{u}(t_n)).$$

Méthode d'Euler. C'est le système de récurrence/condition initiale suivant :

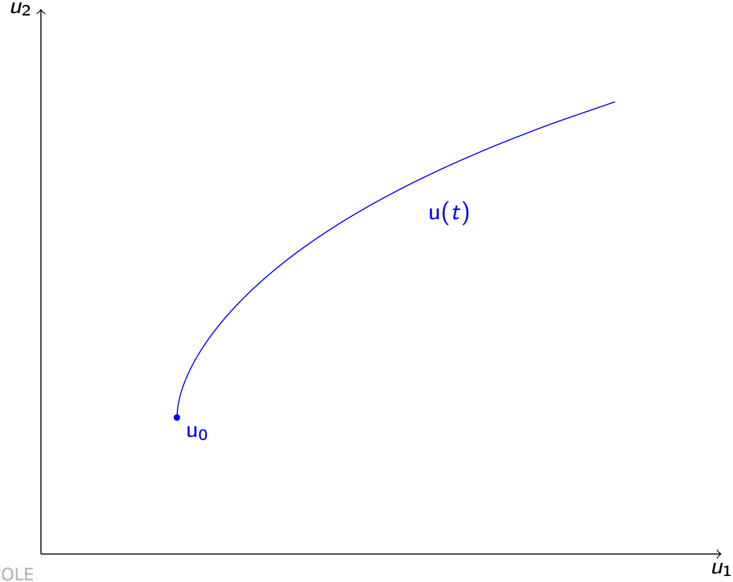
$$(E) : \begin{cases} \mathrm{U}_{n+1} = \mathrm{U}_n + h\,\mathrm{f}(t_n, \mathrm{U}_n), & n = 0, 1, \ldots, N_h - 1 \\ \mathrm{U}_0 = \mathrm{u}_0. \end{cases}$$

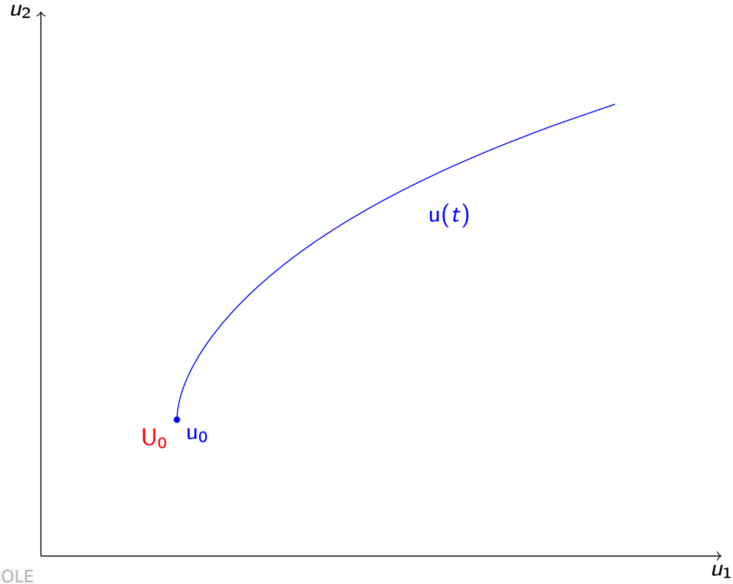
Remarques.

- a priori, on n'a pas $\mathrm{u}(t_n) = U_n$.
- il n'est pas évident non plus que $\mathrm{u}(t_n)$ approche U_n.

Approximation des EDO

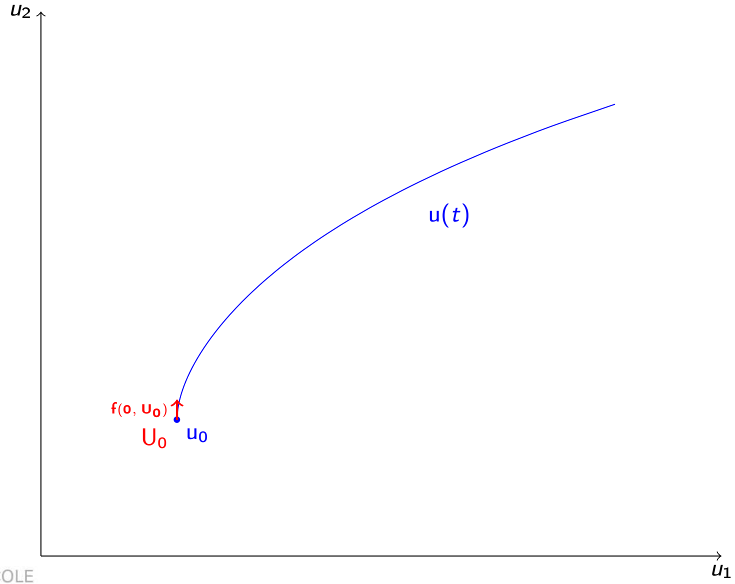
Méthode d'Euler

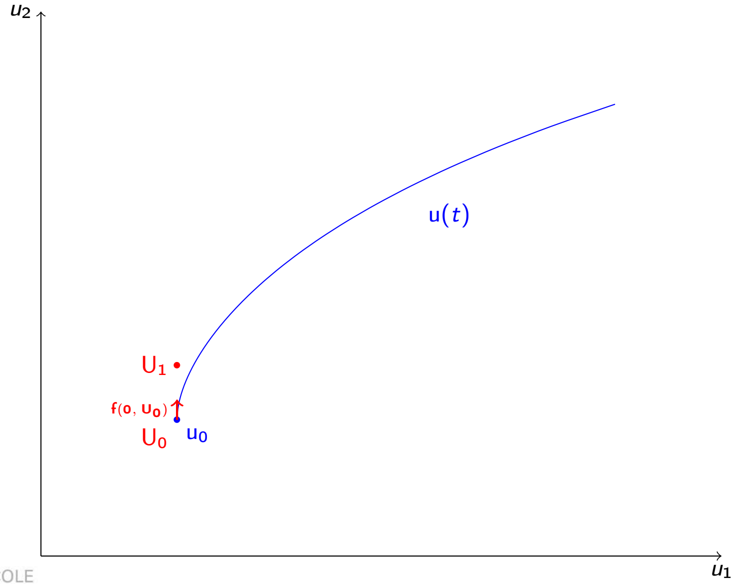


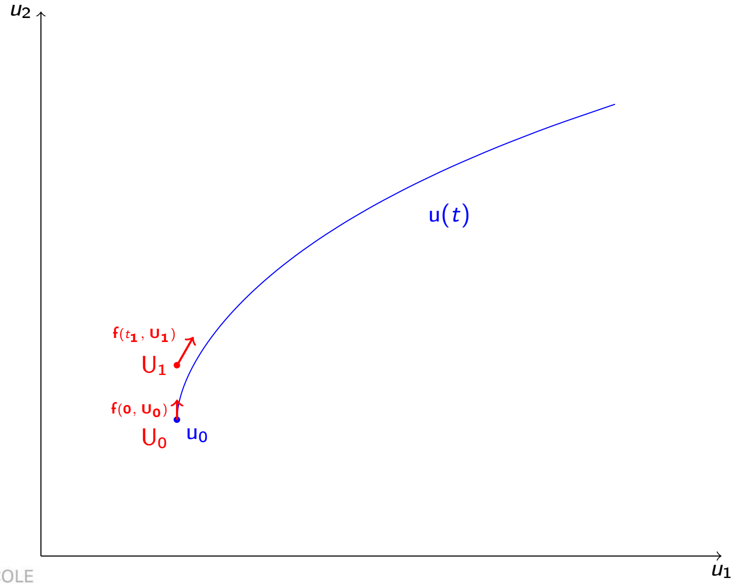

u_2
$u(t)$
U_0
u_0
u_1

Approximation des EDO

Méthode d'Euler

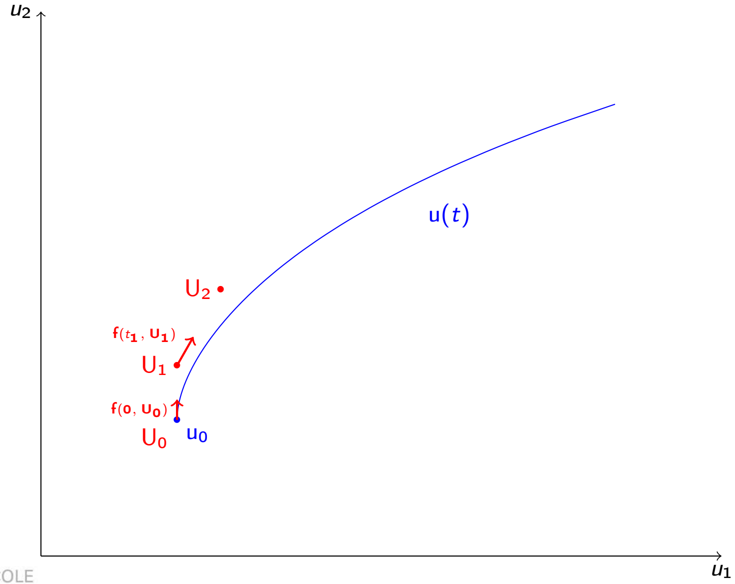



u_2
u_1
$\mathrm{u}(t)$
U_1
$\mathbf{f}(0, \mathbf{u}_0)$
U_0
u_0


u_2
u_1
$\mathbf{u}(t)$
$\mathbf{f}(t_1, \mathbf{U}_1)$
U_1
$\mathbf{f}(0, \mathbf{U}_0)$
U_0
$\mathbf{u}_0$

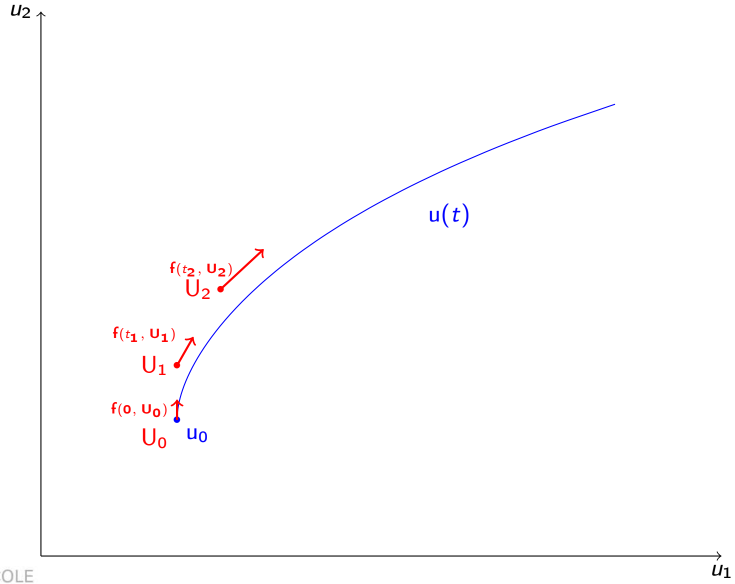
Approximation des EDO

Méthode d'Euler



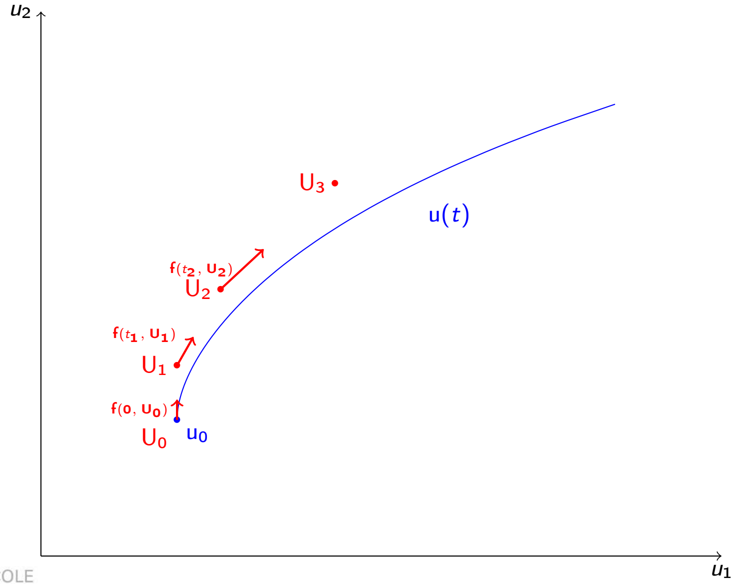
Approximation des EDO

Méthode d'Euler

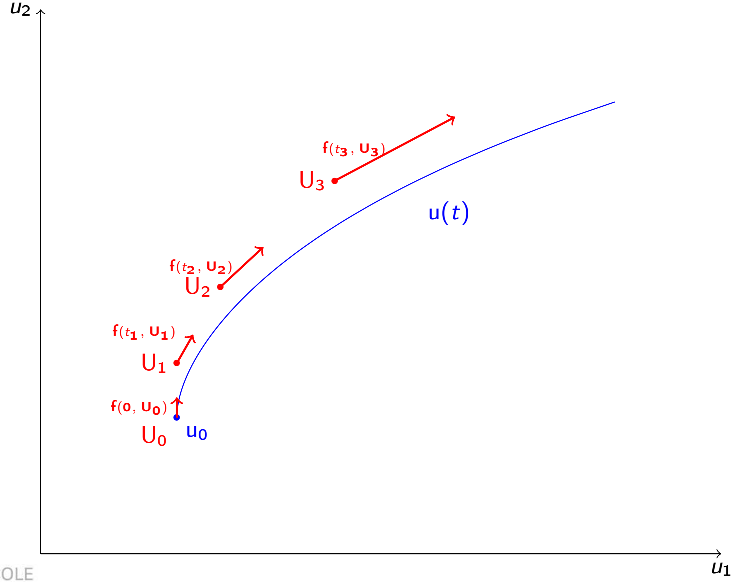


Approximation des EDO

Méthode d'Euler

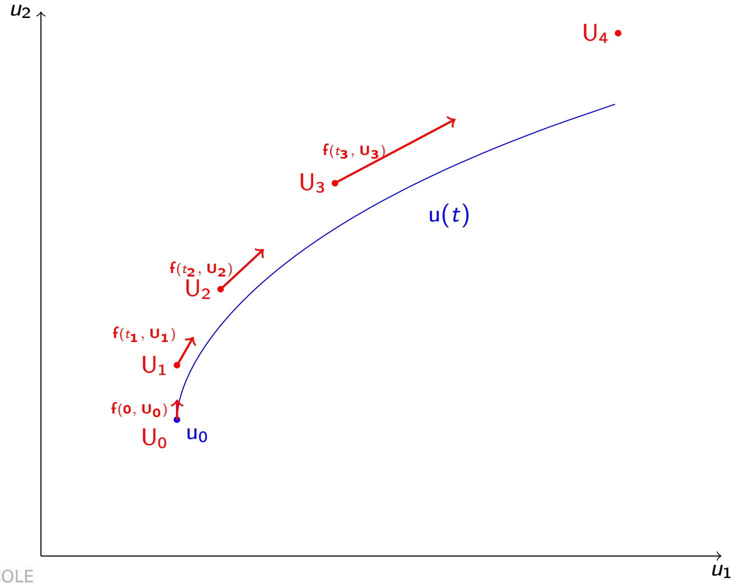


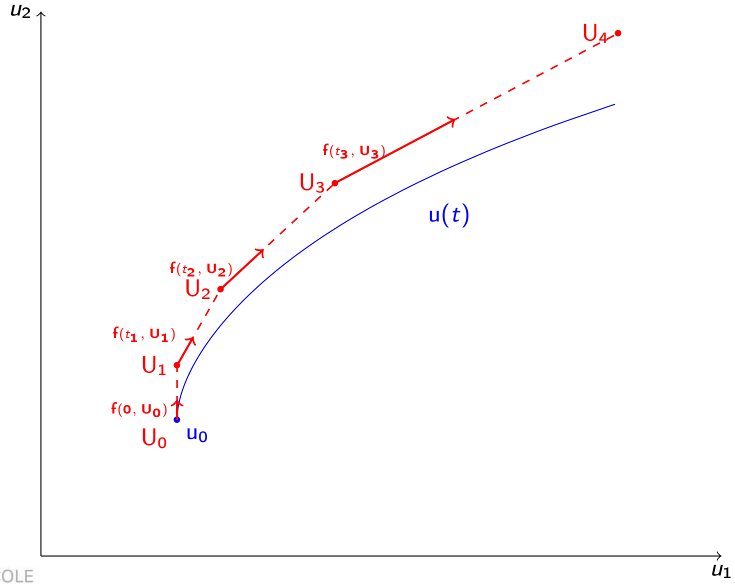
Méthode d'Euler



Approximation des EDO

Méthode d'Euler





u_2
U_4
$\mathbf{f}(t_3, \mathbf{U_3})$
U_3
$\mathrm{u}(t)$
$\mathbf{f}(t_2, \mathbf{U_2})$
U_2
$\mathbf{f}(t_1, \mathbf{U_1})$
U_1
$\mathbf{f}(0, \mathbf{U_0})$
U_0
u_0
u_1

Méthode d'Euler : Exemple

Considérons le problème de Cauchy

$$\begin{cases} u'(t) = 2tu(t) \quad t \in [0,1], \\ u(0) = 1, \end{cases}$$

dont la solution exacte est $u(t) = e^{t^2}$.

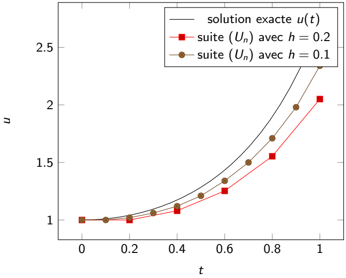
Considérons le problème de Cauchy

$$\begin{cases} u'(t) = 2tu(t) & t \in [0,1], \\ u(0) = 1, \end{cases}$$

dont la solution exacte est $u(t) = e^{t^2}$.

La méthode d'Euler s'écrit

$$\begin{cases} U_{n+1} = U_n + 2ht_nU_n & t \in [0,2], \\ U_0 = 1. \end{cases}$$



Considérons le problème de Cauchy

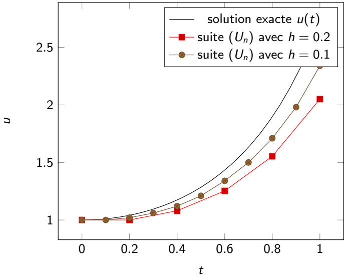
$$\begin{cases} u'(t) = 2tu(t) \quad t \in [0,1], \\ u(0) = 1, \end{cases}$$

dont la solution exacte est $u(t) = e^{t^2}$.

La méthode d'Euler s'écrit

$$\begin{cases} U_{n+1} = U_n + 2ht_nU_n \quad t \in [0,2], \\ U_0 = 1. \end{cases}$$

- La méthode d'Euler n'est pas exacte,



Méthode d'Euler : Exemple

Considérons le problème de Cauchy

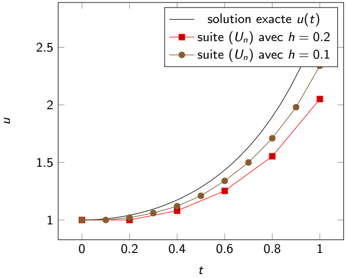
$$\begin{cases} u'(t) = 2tu(t) \quad t \in [0,1], \\ u(0) = 1, \end{cases}$$

dont la solution exacte est $u(t) = e^{t^2}$.

La méthode d'Euler s'écrit

$$\begin{cases} U_{n+1} = U_n + 2ht_nU_n \quad t \in [0,2], \\ U_0 = 1. \end{cases}$$

- La méthode d'Euler n'est pas exacte,
- l'erreur augmente avec le temps,



ÉCOLE CENTRALE LYON

Considérons le problème de Cauchy

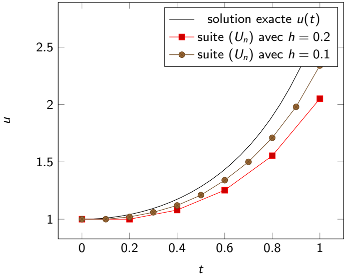
$$\begin{cases} u'(t) = 2tu(t) & t \in [0,1], \\ u(0) = 1, \end{cases}$$

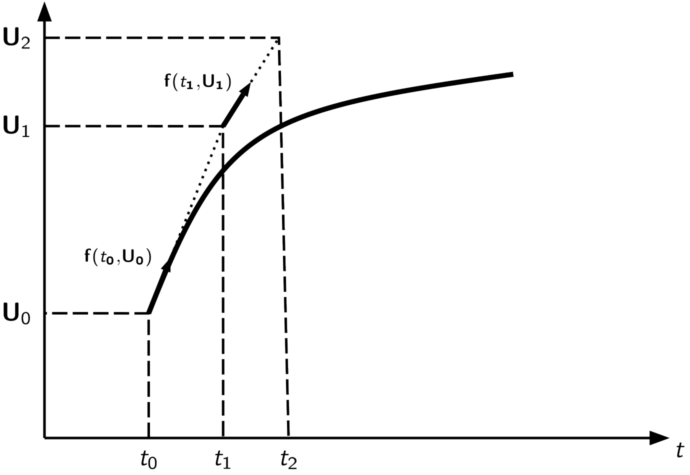
dont la solution exacte est $u(t) = e^{t^2}$.

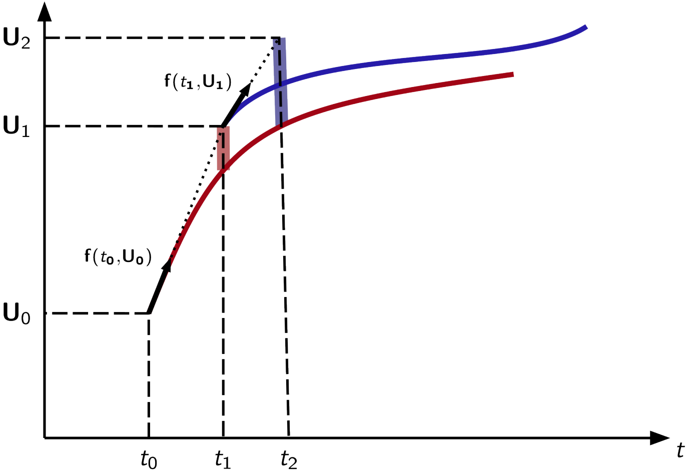
La méthode d'Euler s'écrit

$$\begin{cases} U_{n+1} = U_n + 2ht_nU_n & t \in [0,2], \\ U_0 = 1. \end{cases}$$

- La méthode d'Euler n'est pas exacte,
- l'erreur augmente avec le temps,
- l'erreur diminue si h diminue.




$\mathbf{U}_2$
$\mathbf{f}(t_1, \mathbf{U}_1)$
$\mathbf{U}_1$
$\mathbf{f}(t_0, \mathbf{U}_0)$
$\mathbf{U}_0$
t_0
t_1
t_2
t


$\mathbf{U}_2$
$\mathbf{U}_1$
$\mathbf{U}_0$
$\mathbf{f}(t_1, \mathbf{U}_1)$
$\mathbf{f}(t_0, \mathbf{U}_0)$
t_0
t_1
t_2
t

- **Convergence :** L'erreur globale du schéma tend vers 0

$$\max_{0 \leqslant n \leqslant N_h} \|U_n - u(t_n)\| \xrightarrow[h \to 0]{?} 0.$$

- Convergence : L'erreur globale du schéma tend vers 0

$$\max_{0 \leqslant n \leqslant N_h} \|\mathrm{U}_n - \mathrm{u}(t_n)\| \xrightarrow[h \to 0]{?} 0.$$

- Analyse en 2 temps :
  - Consistance : analyse de l'erreur locale (commise à chaque pas).

- Convergence : L'erreur globale du schéma tend vers 0

$$\max_{0 \leqslant n \leqslant N_h} \|\mathrm{U}_n - \mathrm{u}(t_n)\| \xrightarrow[h \to 0]{?} 0.$$

- Analyse en 2 temps :
  - Consistance : analyse de l'erreur locale (commise à chaque pas).
  - Stabilité : maîtrise de la propagation des erreurs locales.

- Convergence : L'erreur globale du schéma tend vers 0

$$\max_{0 \leqslant n \leqslant N_h} \|\mathrm{U}_n - \mathrm{u}(t_n)\| \xrightarrow[h \to 0]{?} 0.$$

- Analyse en 2 temps :
  - Consistance : analyse de l'erreur locale (commise à chaque pas).
  - Stabilité : maîtrise de la propagation des erreurs locales.

Théorème de Lax.

$$\text{Consistance et Stabilité} \Rightarrow \text{Convergence.}$$

On note u la solution exacte du problème de Cauchy.

Définition. L'erreur de consistance du schéma d'Euler est la suite

$$\varepsilon_n^h := \frac{\mathrm{u}(t_{n+1}) - \mathrm{u}(t_n)}{h} - \mathrm{f}(t_n, \mathrm{u}(t_n)), \quad 0 \leqslant n \leqslant N_h - 1.$$

On note u la solution exacte du problème de Cauchy.

Définition. L'erreur de consistance du schéma d'Euler est la suite

$$\varepsilon_n^h := \frac{\mathrm{u}(t_{n+1}) - \mathrm{u}(t_n)}{h} - \mathrm{f}(t_n, \mathrm{u}(t_n)), \quad 0 \leqslant n \leqslant N_h - 1.$$

Remarque : Lien avec l'approximation de l'intégration de u' sur $[t_n, t_n + h]$ par la méthode des rectangles à gauche :

$$\varepsilon_n^h := \int_0^1 \mathrm{u}'(t_n + hs)\mathrm{d}s - \mathrm{u}'(t_n).$$

Rappel : u désigne la solution exacte du problème de Cauchy.

Définition. L'erreur de consistance) du schéma d'Euler est la suite

$$\varepsilon_n^h := \frac{u(t_{n+1}) - u(t_n)}{h} - f(t_n, u(t_n)), \quad 0 \leqslant n \leqslant N_h - 1.$$

Définition. Le schéma est consistant si l'erreur de consistance converge vers 0 :

$$\max_{0 \leqslant n \leqslant N_h - 1} \left\| \varepsilon_n^h \right\| \to 0 \quad \text{quand} \quad h \to 0.$$

Si de plus il existe $k > 0$ tel que

$$\max_{0 \leqslant n \leqslant N_h - 1} \left\| \varepsilon_n^h \right\| = \mathcal{O}(h^k),$$

le schéma est dit consistant d'ordre k.

Consistance pour Euler

Proposition. Si la solution exacte u est de classe $\mathscr{C}^2$, alors la méthode d'Euler est consistante d'ordre 1.

> **Proposition.** Si la solution exacte u est de classe $\mathscr{C}^2$, alors la méthode d'Euler est consistante d'ordre 1.

Preuve : Par Taylor-Lagrange,

$$\mathrm{u}(t_{n+1}) = \mathrm{u}(t_n + h) = \mathrm{u}(t_n) + h\mathrm{u}'(t_n) + h^2\mathrm{R}_n(h),$$

avec $\mathrm{R}_n(h)$ bornée. Plus précisément, $\|\mathrm{R}_n(h)\| \leqslant C := \dfrac{1}{2} \sup\limits_{t\in[0,T]} \left\|\mathrm{u}''(t)\right\|$

$$\begin{aligned}
\varepsilon_n^h &:= \frac{\mathrm{u}(t_{n+1}) - \mathrm{u}(t_n)}{h} - \mathrm{f}(t_n, \mathrm{u}(t_n)),\\
\varepsilon_n^h &:= \mathrm{u}'(t_n) + h\mathrm{R}_n(h) - \mathrm{f}(t_n, \mathrm{u}(t_n)),\\
\varepsilon_n^h &:= h\mathrm{R}_n(h),
\end{aligned}$$

Ainsi

$$\max_{0\leqslant n\leqslant N_h-1} \left\|\varepsilon_n^h\right\| \leqslant Ch.$$

□

On considère le schéma d'Euler

$$(E) : \begin{cases} \mathrm{U}_{n+1} = \mathrm{U}_n + h\,\mathrm{f}(t_n, \mathrm{U}_n), & n \geqslant 0, \\ \mathrm{U}_0 = \mathrm{u}_0. \end{cases}$$

et on le perturbe en ajoutant une suite d'erreurs μ_n

On considère le schéma d'Euler

$$(E) : \begin{cases} \mathrm{U}_{n+1} = \mathrm{U}_n + h\ \mathrm{f}(t_n, \mathrm{U}_n), \quad n \geqslant 0, \\ \mathrm{U}_0 = \mathrm{u}_0. \end{cases}$$

et on le perturbe en ajoutant une suite d'erreurs μ_n

$$(E_{\text{pert}}) : \begin{cases} \mathrm{V}_{n+1} = \mathrm{V}_n + h\ \mathrm{f}(t_n, \mathrm{V}_n) + \mu_n, \quad n \geqslant 0, \\ \mathrm{V}_0 = \mathrm{u}_0. \end{cases}$$

On considère le schéma d'Euler

$$(E) : \begin{cases} \mathrm{U}_{n+1} = \mathrm{U}_n + h\, \mathrm{f}(t_n, \mathrm{U}_n), & n \geqslant 0, \\ \mathrm{U}_0 = \mathrm{u}_0. \end{cases}$$

et on le perturbe en ajoutant une suite d'erreurs μ_n

$$(E_{\text{pert}}) : \begin{cases} \mathrm{V}_{n+1} = \mathrm{V}_n + h\, \mathrm{f}(t_n, \mathrm{V}_n) + \mu_n, & n \geqslant 0, \\ \mathrm{V}_0 = \mathrm{u}_0. \end{cases}$$

Définition. Le schéma est stable s'il existe une constante $C > 0$ telle que

$$\max_{0 \leqslant n \leqslant N_h} \|\mathrm{V}_n - \mathrm{U}_n\| \leqslant C \sum_{0 \leqslant n \leqslant N_h - 1} \|\mu_n\| .$$

Proposition. Si f est L-lipschitzienne, la méthode d'Euler est stable et l'on a

$$\max_{0 \leqslant n \leqslant N_h} |V_n - U_n| \leqslant e^{LT} \sum_{0 \leqslant n \leqslant N_h - 1} \|\mu_n\| \,.$$

Proposition. Si f est L-lipschitzienne, la méthode d'Euler est stable et l'on a

$$\max_{0\leqslant n\leqslant N_h} |V_n - U_n| \leqslant e^{LT} \sum_{0\leqslant n\leqslant N_h-1} \|\mu_n\| .$$

Preuve : Par soustraction

$$\begin{aligned} V_{n+1} - U_{n+1} &= V_n - U_n + h\big(f(t_n, V_n) - f(t_n, U_n)\big) + \mu_n, \\ \|V_{n+1} - U_{n+1}\| &\leqslant \|V_n - U_n\| + h\,\|f(t_n, V_n) - f(t_n, U_n)\| + \|\mu_n\| , \\ \|V_{n+1} - U_{n+1}\| &\leqslant (1 + hL)\,\|V_n - U_n\| + \|\mu_n\| . \end{aligned}$$

Posons $e_n = \|V_n - U_n\|$, et divisons par $(1+hL)^{n+1}$;

$$\begin{aligned} \frac{e_{n+1}}{(1+hL)^{n+1}} &\leqslant \frac{e_n}{(1+hL)^n} + \frac{\|\mu_n\|}{(1+hL)^{n+1}}, \\ \frac{e_{n+1}}{(1+hL)^{n+1}} &\leqslant \frac{e_n}{(1+hL)^n} + \|\mu_n\|, \end{aligned}$$

Preuve (suite) : On a

$$\frac{e_{n+1}}{(1+hL)^{n+1}} - \frac{e_n}{(1+hL)^n} \leqslant \|\mu_n\| \,,$$

En sommant, on obtient

$$\frac{e_n}{(1+hL)^n} \leqslant \sum_{n=0}^{n-1} \|\mu_k\| \,, \quad \forall n \in \mathbb{N},$$

$$e_n \leqslant (1+hL)^n \sum_{n=0}^{n-1} \|\mu_k\| \,, \quad \forall n \in \mathbb{N}.$$

Or pour tout $n \leqslant N_h$,

$$(1+hL)^n \leqslant e^{Lt_N} \leqslant e^{N_h hL} = e^{LT}.$$

Ainsi,

$$\|\mathrm{V}_n - \mathrm{U}_n\| \leqslant e^{LT} \sum_{n=0}^{n-1} \|\mu_k\| \,, \quad \forall n = 0, 1, \ldots, N_h.$$

$\square$

> **Définition.** Le schéma <u>converge</u> si
>
> $$\max_{0 \leqslant n \leqslant N_h} |\mathrm{U}_n - \mathrm{u}(t_n)| \longrightarrow 0 \quad \text{quand} \quad h \to 0.$$

Définition. Le schéma converge si

$$\max_{0 \leqslant n \leqslant N_h} |\mathrm{U}_n - \mathrm{u}(t_n)| \longrightarrow 0 \quad \text{quand} \quad h \to 0.$$

Théorème de Lax. Un schéma consistant et stable est convergent.

Définition. Le schéma converge si

$$\max_{0\leqslant n\leqslant N_h} |\mathrm{U}_n - \mathrm{u}(t_n)| \longrightarrow 0 \quad \text{quand} \quad h \to 0.$$

Théorème de Lax. Un schéma consistant d'ordre k et stable est convergent d'ordre k.

Preuve (Pour Euler) : Par définition,

$$\varepsilon_n^h := \frac{\mathrm{u}(t_{n+1}) - \mathrm{u}(t_n)}{h} - \mathrm{f}(t_n, \mathrm{u}(t_n)).$$

Donc la suite de valeurs exactes est solution du schéma perturbé :

$$\begin{cases} \mathrm{u}(t_{n+1}) = \mathrm{u}(t_n) + h\mathrm{f}(t_n, \mathrm{u}(t_n)) + h\varepsilon_n^h, \\ \mathrm{u}(t_0) = u_0. \end{cases}$$

Preuve (suite) : On utilise la stabilité du schéma :

$$\max_{0\leqslant n\leqslant N_h} |\mathrm{U}_n - \mathrm{u}(t_n)| \leqslant e^{LT} h \sum_{0\leqslant n\leqslant N_h-1} \left\| \varepsilon_n^h \right\| .$$

Or,

$$\sum_{0\leqslant n\leqslant N_h-1} \left\| \varepsilon_n^h \right\| \leqslant N_h \max_{0\leqslant n\leqslant N_h-1} \left\| \varepsilon_n^h \right\| ,$$

Ainsi,

$$\max_{0\leqslant n\leqslant N_h} |\mathrm{U}_n - \mathrm{u}(t_n)| \leqslant Te^{LT} \max_{0\leqslant n\leqslant N_h-1} \left\| \varepsilon_n^h \right\| .$$

Or, par consistance d'ordre 1,

$$\max_{0\leqslant n\leqslant N_h-1} \left\| \varepsilon_n^h \right\| \leqslant Ch.$$

Finalement,

$$\max_{0\leqslant n\leqslant N_h} |\mathrm{U}_n - \mathrm{u}(t_n)| \leqslant CTe^{LT} h.$$

$\square$

Proposition. Le schéma d'Euler est convergent d'ordre 1 :

$$\max_{0 \leqslant n \leqslant N_h} |\mathrm{U}_n - \mathrm{u}(t_n)| \leqslant C T e^{LT} h,$$

avec $C = \dfrac{1}{2} \sup_{t \in [0,T]} \left\| \mathrm{u}''(t) \right\|$.

Proposition. Le schéma d'Euler est convergent d'ordre 1 :

$$\max_{0 \leqslant n \leqslant N_h} |\mathrm{U}_n - \mathrm{u}(t_n)| \leqslant CTe^{LT} h,$$

avec $C = \frac{1}{2} \sup_{t \in [0,T]} \left\| \mathrm{u}''(t) \right\|$.

Remarques :

- On a besoin que $\mathrm{u} \in \mathscr{C}^2([0, T])$.
- La constante Te^{LT} augmente exponentiellement
  - avec T (problèmes en temps long) ,
  - avec L (problèmes « raides »).

Définition. Un schéma à un pas (constant) pour approcher la solution du problème de Cauchy

$$(\mathcal{C}) : \begin{cases} \mathrm{u}'(t) = \mathrm{f}(t, \mathrm{u}(t)) & t \in [0, T], \\ \mathrm{u}(0) = \mathrm{u}_0. \end{cases}$$

est une relation de récurrence de la forme

$$\begin{cases} \mathrm{U}_{n+1} = \mathrm{U}_n + h\mathrm{G}_h(t_n, \mathrm{U}_n, \mathrm{U}_{n+1}), & n = 0, 1, \ldots, N_h - 1, \\ \mathrm{U}_0 = \mathrm{u}_0. \end{cases}$$

où $h > 0$ est le pas et G_h est une fonction qui définit le schéma.

Définition. Un schéma à un pas (constant) pour approcher la solution du problème de Cauchy

$$(\mathcal{C}) : \begin{cases} u'(t) = f(t, u(t)) & t \in [0, T], \\ u(0) = u_0. \end{cases}$$

est une relation de récurrence de la forme

$$\begin{cases} U_{n+1} = U_n + hG_h(t_n, U_n, U_{n+1}), & n = 0, 1, \ldots, N_h - 1, \\ U_0 = u_0. \end{cases}$$

où $h > 0$ est le pas et G_h est une fonction qui définit le schéma.

Vocabulaire : Si G_h dépend effectivement de U_{n+1} le schéma est dit implicite, sinon il est explicite.

Définition. Un schéma à un pas (constant) pour approcher la solution du problème de Cauchy

$$(\mathcal{C}) : \begin{cases} u'(t) = f(t, u(t)) & t \in [0, T], \\ u(0) = u_0. \end{cases}$$

est une relation de récurrence de la forme

$$\begin{cases} U_{n+1} = U_n + hG_h(t_n, U_n, U_{n+1}), & n = 0, 1, \dots, N_h - 1, \\ U_0 = u_0. \end{cases}$$

où $h > 0$ est le pas et G_h est une fonction qui définit le schéma.

Vocabulaire : Si G_h dépend effectivement de U_{n+1} le schéma est dit implicite, sinon il est explicite.

Exemple : Le schéma d'Euler $U_{n+1} = U_n + hf(t_n, U_n)$ correspond à

$$G_h(t_n, U_n, U_{n+1}) = f(t_n, U_n).$$

Rappel : u désigne la solution exacte du problème de Cauchy.

Définition. L'erreur de consistance d'un schéma général à un pas est

$$\varepsilon_n^h := \frac{\mathrm{u}(t_{n+1}) - \mathrm{u}(t_n)}{h} - G_h\big(t_n, \mathrm{u}(t_n), \mathrm{u}(t_{n+1})\big), \quad 0 \leqslant n \leqslant N_h.$$

Remarque : Lien avec l'erreur de l'approximation de l'intégration de u′ sur $[t_n, t_n + h]$:

$$\varepsilon_n^h := \int_0^1 \mathrm{u}'(t_n + hs)\mathrm{d}s - G_h\big(t_n, \mathrm{u}(t_n), \mathrm{u}(t_{n+1})\big).$$

On considère un schéma général

$$(E) : \begin{cases} \mathrm{U}_{n+1} = \mathrm{U}_n + hG_h(t_n, \mathrm{U}_n, \mathrm{U}_{n+1}), \quad n = 0, 1, \ldots, N_h - 1, \\ \mathrm{U}_0 = \mathrm{u}_0. \end{cases}$$

et on le perturbe en ajoutant une suite d'erreurs μ_n :

$$(E_{\text{pert}}) : \begin{cases} \mathrm{V}_{n+1} = \mathrm{V}_n + hG_h(t_n, \mathrm{V}_n, \mathrm{V}_{n+1}) + \mu_n, \quad n = 0, 1, \ldots, N_h - 1, \\ \mathrm{V}_0 = \mathrm{u}_0. \end{cases}$$

Définition. Le schéma est stable s'il existe $C > 0$ t.q.

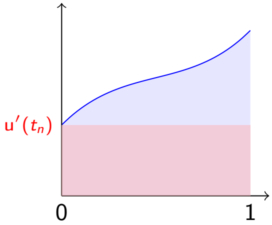
$$\max_{0 \leqslant n \leqslant N_h} \|\mathrm{V}_n - \mathrm{U}_n\| \leqslant C \sum_{0 \leqslant n \leqslant N_h - 1} \|\mu_n\| \,.$$

$$\int_{t_n}^{t_{n+1}} u'(t)\,\mathrm{d}t = h\int_0^1 u'(t_n+hs)\,\mathrm{d}s$$

Rectangle à gauche :

$$I(h) = \int_0^1 \mathrm{u}'(t_n+hs)\mathrm{d}s \approx \mathrm{u}'(t_n) = \mathrm{f}(t_n,\mathrm{u}(t_n))$$

conduit à poser
Euler : $\mathrm{U}_{n+1} = \mathrm{U}_n + h\mathrm{f}(t_n,\mathrm{U}_n)$.

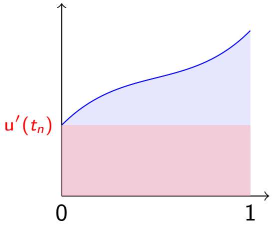


Rectangle à gauche :

$$I(h) = \int_0^1 u'(t_n + hs)ds$$
$$\approx u'(t_n) = f(t_n, u(t_n))$$

conduit à poser

Euler : $U_{n+1} = U_n + hf(t_n, U_n)$**.**

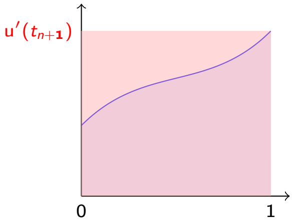


Rectangle à droite :

$$I(h) = \int_0^1 u'(t_n + hs)ds$$
$$\approx u'(t_{n+1}) = f(t_{n+1}, u(t_{n+1}))$$

conduit à poser

Euler (implicite) : $U_{n+1} = U_n + hf(t_{n+1}, U_{n+1})$

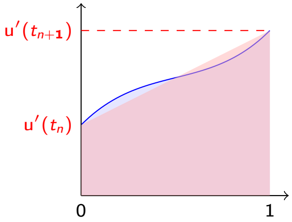


Trapèze :

$$I(h) \approx \frac{u'(t_n) + u'(t_{n+1})}{2}$$

Crank-Nicolson (implicite)

$$U_{n+1} = U_n + \frac{h}{2}\left(f(t_n, U_n) + f(t_{n+1}, U_{n+1})\right)$$

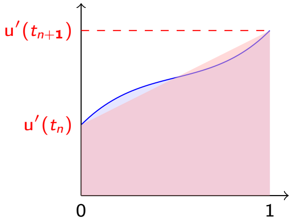


Trapèze :

$$I(h) \approx \frac{u'(t_n) + u'(t_{n+1})}{2}$$

Crank-Nicolson (implicite)

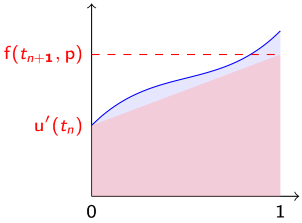
$$U_{n+1} = U_n + \frac{h}{2}\left(f(t_n, U_n) + f(t_{n+1}, U_{n+1})\right)$$



Trapèze approché : Le vecteur p approche $u(t_{n+1})$ par Euler :

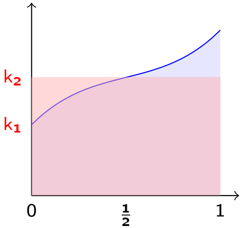
Heun (explicite)

$$\left|\begin{array}{l} p = U_n + hf(t_n, U_n). \\ U_{n+1} = U_n + \frac{h}{2}\left(f(t_n, U_n) + f(t_{n+1}, p)\right) \end{array}\right.$$



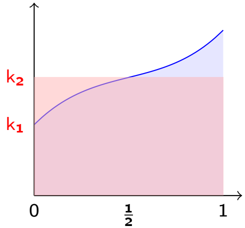
Point milieu approché : **RK-2** (explicite)

$$
\left|\begin{aligned}
\mathrm{k}_1 &= \mathrm{f}\left(t_n, \mathrm{U}_n\right).\\
\mathrm{k}_2 &= \mathrm{f}\left(t_n + \frac{h}{2}, \mathrm{U}_n + \frac{h}{2}k_1\right).\\
\mathrm{U}_{n+1} &= \mathrm{U}_n + hk_2.
\end{aligned}\right.
$$



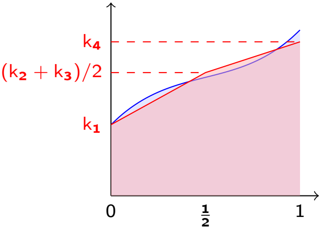
Point milieu approché : **RK-2** (explicite)

$$\left|\begin{aligned} \mathrm{k}_1 &= \mathrm{f}\left(t_n, \mathrm{U}_n\right). \\ \mathrm{k}_2 &= \mathrm{f}\left(t_n + \frac{h}{2}, \mathrm{U}_n + \frac{h}{2} k_1\right). \\ \mathrm{U}_{n+1} &= \mathrm{U}_n + h k_2. \end{aligned}\right.$$



Méthode de Simpson approchée : **RK-4** (explicite)

$$\left|\begin{aligned} \mathrm{k}_1 &= \mathrm{f}\left(t_n, \mathrm{U}_n\right). \\ \mathrm{k}_2 &= \mathrm{f}\left(t_n + \frac{h}{2}, \mathrm{U}_n + \frac{h}{2} k_1\right). \\ \mathrm{k}_3 &= \mathrm{f}\left(t_n + \frac{h}{2}, \mathrm{U}_n + \frac{h}{2} k_2\right). \\ \mathrm{k}_4 &= \mathrm{f}\left(t_n + h, \mathrm{U}_n + h\mathrm{k}_3\right). \\ \mathrm{U}_{n+1} &= \mathrm{U}_n + \frac{h}{6}\left(k_1 + 2k_2 + 2k_3 + k_4\right). \end{aligned}\right.$$



Pourquoi les méthodes implicites ?

On considère le problème de Cauchy :

$$\begin{cases} y'(t) = -1000y(t), \\ y(0) = 1. \end{cases}$$

On considère le problème de Cauchy :

$$\begin{cases} y'(t) = -1000y(t), \\ y(0) = 1. \end{cases}$$

- Problème **raide** : $f(y) = -1000y$ est 1000-lipschitzienne.
- Solution exacte : $y(t) = e^{-1000t}$ (positive, converge vers 0 en $+\infty$).
- Constante d'erreur pour Euler proportionnelle à Te^{1000T}.

On considère le problème de Cauchy :

$$\begin{cases} y'(t) = -1000y(t), \\ y(0) = 1. \end{cases}$$

- Problème **raide** : $f(y) = -1000y$ est 1000-lipschitzienne.
- Solution exacte : $y(t) = e^{-1000t}$ (positive, converge vers 0 en $+\infty$).
- Constante d'erreur pour Euler proportionnelle à Te^{1000T}.

> **Méthode d'Euler :**
>
> $$\begin{cases} Y_{n+1} = Y_n - 1000hY_n, \\ Y_0 = 1, \end{cases}$$
>
> soit $Y_n = (1 - 1000h)^n$.

On considère le problème de Cauchy :

$$\begin{cases} y'(t) = -1000y(t), \\ y(0) = 1. \end{cases}$$

- Problème **raide** : $f(y) = -1000y$ est 1000-lipschitzienne.
- Solution exacte : $y(t) = e^{-1000t}$ (positive, converge vers 0 en $+\infty$).
- Constante d'erreur pour Euler proportionnelle à Te^{1000T}.

Méthode d'Euler :

$$\begin{cases} Y_{n+1} = Y_n - 1000hY_n, \\ Y_0 = 1, \end{cases}$$

soit $Y_n = (1 - 1000h)^n$.

La solution numérique reste positive ssi $h \leqslant 10^{-3}$

On considère le problème de Cauchy :

$$\begin{cases} y'(t) = -1000y(t), \\ y(0) = 1. \end{cases}$$

- Problème **raide** : $f(y) = -1000y$ est 1000-lipschitzienne.
- Solution exacte : $y(t) = e^{-1000t}$ (positive, converge vers 0 en $+\infty$).
- Constante d'erreur pour Euler proportionnelle à Te^{1000T}.

Méthode d'Euler :

$$\begin{cases} Y_{n+1} = Y_n - 1000hY_n, \\ Y_0 = 1, \end{cases}$$

soit $Y_n = (1 - 1000h)^n$.

$\boxed{\text{La solution numérique tend vers 0 ssi } h < 2 \cdot 10^{-3}}$

Euler implicite :

$$\begin{cases} Y_{n+1} = Y_n - 1000hY_{n+1}, \\ Y_0 = 1, \end{cases}$$

soit $Y_n = (1 + 1000h)^{-n}$.

Euler implicite :

$$\begin{cases} Y_{n+1} = Y_n - 1000hY_{n+1}, \\ Y_0 = 1, \end{cases}$$

soit $Y_n = (1 + 1000h)^{-n}$.

La solution numérique **reste positive** et **converge vers 0** pour tout $h > 0$.

$\boxed{\text{Pas de contrainte sur } h.}$

Euler implicite :

$$\begin{cases} Y_{n+1} = Y_n - 1000hY_{n+1}, \\ Y_0 = 1, \end{cases}$$

soit $Y_n = (1 + 1000h)^{-n}$.

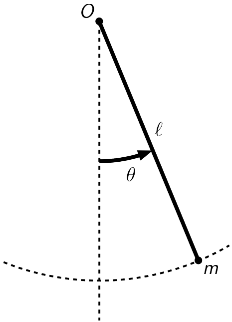
La solution numérique **reste positive** et **converge vers 0** pour tout $h > 0$.

Pas de contrainte sur h.

- Méthode implicite plus coûteuse (pour les problèmes non-linéaires).
- Mais utile pour les problèmes raides en temps long.

Modélisation du pendule sans frottement :

$$\begin{cases} \theta''(t) = -\dfrac{g}{\ell}\sin\left(\theta(t)\right) \\ \theta(0), \theta'(0) \text{ donnés.} \end{cases}$$



Énergie du système :

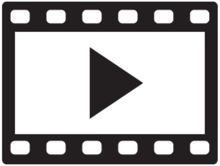
$$E(t) = \frac{1}{2}\theta'(t)^2 - \frac{g}{\ell}\cos\left(\theta(t)\right).$$

Résultat : $E'(t) = 0$.

Question : est-ce respecté par les schémas ?

Approximation des EDO

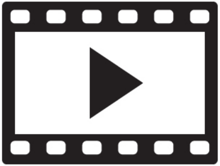
Conservation d'énergie



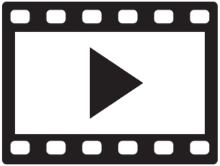
Méthode d'Euler explicite (ordre 1) pour $h = 0.1$.

Approximation des EDO

Conservation d'énergie



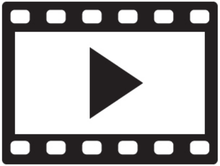
Méthode d'Euler implicite (ordre 1) pour $h = 0.1$.



Méthode RK4 (ordre 4) pour $h = 0.1$.

Approximation des EDO

Conservation d'énergie



Méthode d'Euler symplectique (ordre 1) pour $h = 0.1$.

COURS 5

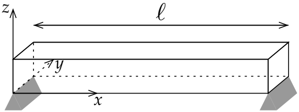
Discrétisation des EDP

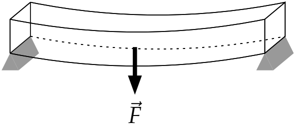
l'équation de Laplace

ÉCOLE CENTRALE LYON

Discrétisation de l'équation de Laplace

Un modèle de poutre en flexion



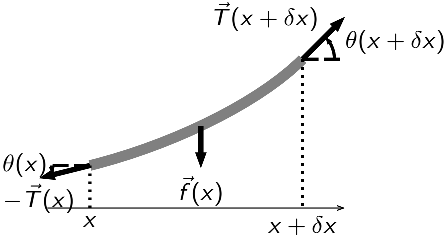


- Modèle 1D :
  - Déplacements verticaux seulement.
  - Ligne moyenne donnée par $z = u(x)$.
  - Densité linéique de force appliquée $x \mapsto f(x)$.

- Principe de Newton

$$-\vec{T}(x) + \vec{T}(x+\delta x) = \vec{f}(x)\delta x.$$

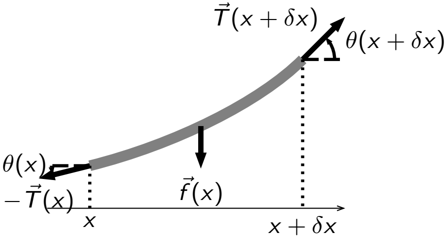
(élément de corde situé entre x et $x+\delta x$).



- Principe de Newton

$$-\vec{T}(x) + \vec{T}(x+\delta x) = \vec{f}(x)\delta x.$$

(élément de corde situé entre x et $x+\delta x$).

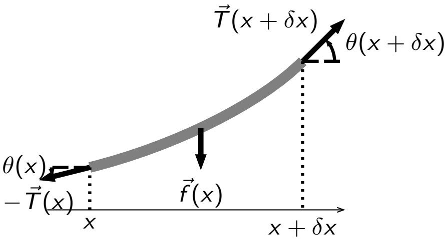


- Proj. horizontale : $-T(x)\cos(\theta(x)) + T(x+\delta x)\cos(\theta(x+\delta x)) = 0.$

- Principe de Newton

$$-\vec{T}(x) + \vec{T}(x+\delta x) = \vec{f}(x)\delta x.$$

(élément de corde situé entre x et $x+\delta x$).



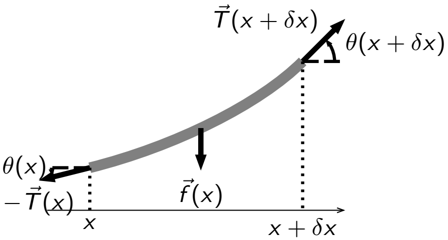
- Proj. horizontale : $-T(x)\cos(\theta(x)) + T(x+\delta x)\cos(\theta(x+\delta x)) = 0.$

$$\Longrightarrow T(x)\cos(\theta(x)) = k.$$

- Principe de Newton

$$-\vec{T}(x) + \vec{T}(x+\delta x) = \vec{f}(x)\delta x.$$

(élément de corde situé entre x et $x+\delta x$).



- Proj. horizontale : $-T(x)\cos(\theta(x)) + T(x+\delta x)\cos(\theta(x+\delta x)) = 0.$

$$\Longrightarrow T(x)\cos(\theta(x)) = k.$$

- Proj. verticale : $-T(x)\sin(\theta(x)) + T(x+\delta x)\sin(\theta(x+\delta x)) = f(x)\delta x.$

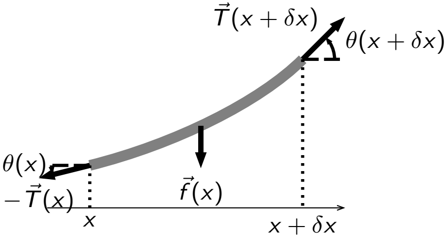
Discrétisation de l'équation de Laplace

Un modèle de poutre en flexion

- Principe de Newton

  $$-\vec{T}(x) + \vec{T}(x+\delta x) = \vec{f}(x)\delta x.$$

  (élément de corde situé entre x et $x+\delta x$).



- Proj. horizontale : $-T(x)\cos(\theta(x)) + T(x+\delta x)\cos(\theta(x+\delta x)) = 0.$

  $$\Longrightarrow T(x)\cos(\theta(x)) = k.$$

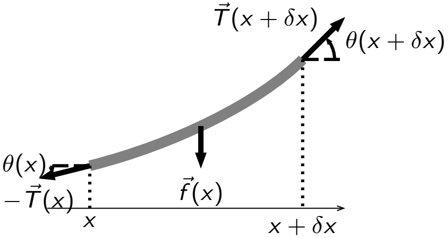
- Proj. verticale : $-T(x)\sin(\theta(x)) + T(x+\delta x)\sin(\theta(x+\delta x)) = f(x)\delta x.$

  $$\Longrightarrow -k\tan(\theta(x)) + k\tan(\theta(x+\delta x)) = f(x)\delta x.$$

- Principe de Newton

$$-\vec{T}(x) + \vec{T}(x+\delta x) = \vec{f}(x)\delta x.$$

(élément de corde situé entre x et $x+\delta x$).



- Proj. horizontale : $-T(x)\cos(\theta(x)) + T(x+\delta x)\cos(\theta(x+\delta x)) = 0.$

$$\Longrightarrow T(x)\cos(\theta(x)) = k.$$

- Proj. verticale : $-T(x)\sin(\theta(x)) + T(x+\delta x)\sin(\theta(x+\delta x)) = f(x)\delta x.$

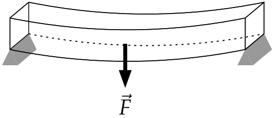
$$\Longrightarrow -k\tan(\theta(x)) + k\tan(\theta(x+\delta x)) = f(x)\delta x.$$

- $\delta x \to 0$: $-k\frac{d}{dx}(\tan\theta(x)) = f(x)$

$$\Longrightarrow \boxed{-ku''(x) = f(x).}$$

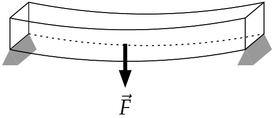
- Problème 1D

$$\begin{cases} -ku''(x) = f(x), \quad x \in]0,1[, \\ u(0) = u(1) = 0. \end{cases}$$



- Problème 1D

$$\begin{cases} -ku''(x) = f(x), & x \in]0,1[, \\ u(0) = u(1) = 0. \end{cases}$$

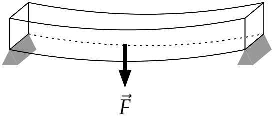


- Résolution explicite dans ce cas particulier

$$u(x) = \frac{1}{k}\int_0^x \left(\int_0^y f(s)\,\mathrm{d}s\right)\mathrm{d}y - \frac{x}{k}\int_0^1 \left(\int_0^y f(s)\,\mathrm{d}s\right)\mathrm{d}y.$$

- Problème 1D

$$\begin{cases} -ku''(x) = f(x), & x \in]0,1[, \\ u(0) = u(1) = 0. \end{cases}$$



- Résolution explicite dans ce cas particulier

$$u(x) = \frac{1}{k}\int_0^x \left(\int_0^y f(s)\,\mathrm{d}s\right)\mathrm{d}y - \frac{x}{k}\int_0^1 \left(\int_0^y f(s)\,\mathrm{d}s\right)\mathrm{d}y.$$

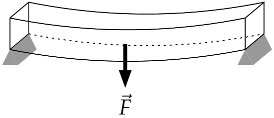
- Problème plus général :

$$\begin{cases} -ku''(x) + b(x)u(x) = f(x), & x \in]0,1[, \\ u(0) = u(1) = 0. \end{cases}$$

Pas de formule explicite...

- Problème 1D

$$\begin{cases} -ku''(x) = f(x), & x \in]0,1[, \\ u(0) = u(1) = 0. \end{cases}$$

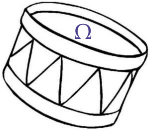


- Résolution explicite dans ce cas particulier

$$u(x) = \frac{1}{k}\int_0^x \left(\int_0^y f(s)\,\mathrm{d}s\right)\mathrm{d}y - \frac{x}{k}\int_0^1 \left(\int_0^y f(s)\,\mathrm{d}s\right)\mathrm{d}y.$$

- En dimension supérieure : $(\Delta = \partial_x^2 + \partial_y^2)$

$$\begin{cases} -k\Delta u(x,y) = f(x,y), & (x,y) \in \Omega \\ u(x,y) = 0, & (x,y) \in \partial\Omega. \end{cases}$$



Pas de formule explicite...

- Problème modèle

$$\begin{cases} -ku''(x) = f(x), \quad x \in]0,1[, \\ u(0) = \alpha, \;\; u(1) = \beta. \end{cases}$$

- Problème modèle aux limites / Problème de Cauchy

$$\begin{cases} -ku''(x) = f(x), & x \in]0,1[, \\ u(0) = \alpha, \ u(1) = \beta. \end{cases} \qquad \begin{cases} -ku''(x) = f(x), & x \in]0,1[, \\ u(0) = \alpha, \ u'(0) = q. \end{cases}$$

- Problème modèle aux limites / Problème de Cauchy

$$\begin{cases} -ku''(x) = f(x), \quad x \in]0,1[, \\ u(0) = \alpha, \ u(1) = \beta. \end{cases} \qquad \begin{cases} -ku_q''(x) = f(x), \quad x \in]0,1[, \\ u_q(0) = \alpha, \ u_q'(0) = q. \end{cases}$$

- Pour $q \in \mathbb{R}$, il existe une unique solution u_q.

- Problème modèle aux limites / Problème de Cauchy

$$\begin{cases} -ku''(x) = f(x), & x \in]0,1[, \\ u(0) = \alpha, \ u(1) = \beta. \end{cases} \qquad \begin{cases} -ku_q''(x) = f(x), & x \in]0,1[, \\ u_q(0) = \alpha, \ u_q'(0) = q. \end{cases}$$

- Pour $q \in \mathbb{R}$, il existe une unique solution u_q.
- **Question** : peut-on trouver $q \in \mathbb{R}$ pour que $u_q(1) = \beta$?

- Problème modèle aux limites / Problème de Cauchy

$$\begin{cases} -ku''(x) = f(x), & x \in]0,1[, \\ u(0) = \alpha, \ u(1) = \beta. \end{cases} \qquad \begin{cases} -ku_q''(x) = f(x), & x \in]0,1[, \\ u_q(0) = \alpha, \ u_q'(0) = q. \end{cases}$$

- Pour $q \in \mathbb{R}$, il existe une unique solution u_q.

- **Question** : peut-on trouver $q \in \mathbb{R}$ pour que $u_q(1) = \beta$?

- *[Tir1.m]* *[Tir2.m]* *[Tir3.m]*

Discrétisation de l'équation de Laplace

Problème 1D : méthode de tirs

- Problème modèle aux limites / Problème de Cauchy

$$\begin{cases} -ku''(x) = f(x), & x \in]0,1[, \\ u(0) = \alpha, \ u(1) = \beta. \end{cases} \qquad \begin{cases} -ku_q''(x) = f(x), & x \in]0,1[, \\ u_q(0) = \alpha, \ u_q'(0) = q. \end{cases}$$

- Pour $q \in \mathbb{R}$, il existe une unique solution u_q.
- **Question** : peut-on trouver $q \in \mathbb{R}$ pour que $u_q(1) = \beta$?
- [*Tir1.m*] [*Tir2.m*] [*Tir3.m*]

Théorème : l'application $q \longmapsto u_q(1)$ est affine.

Discrétisation de l'équation de Laplace

Problème 1D : méthode de tirs

- Problème modèle aux limites / Problème de Cauchy

$$\left\{\begin{array}{l} -ku''(x) = f(x), \quad x \in]0,1[, \\ u(0) = \alpha, \;\; u(1) = \beta. \end{array}\right. \qquad \left\{\begin{array}{l} -ku_q''(x) = f(x), \quad x \in]0,1[, \\ u_q(0) = \alpha, \;\; u_q'(0) = q. \end{array}\right.$$

- Pour $q \in \mathbb{R}$, il existe une unique solution u_q.
- **Question** : peut-on trouver $q \in \mathbb{R}$ pour que $u_q(1) = \beta$?
- [*Tir1.m*] [*Tir2.m*] [*Tir3.m*]

> **Théorème :** l'application $q \longmapsto u_q(1)$ est affine.

- Pas généralisable en dimension supérieure !

- Problème modèle

$$\left\{\begin{array}{ll} -ku''(x) = f(x), & x \in]0,1[, \\ u(0) = \alpha, \ u(1) = \beta. & \end{array}\right.$$

- Discrétisation de pas $h = \frac{1}{N+1}$: $x_i = ih$.

- Approximation des dérivées pour $g \in \mathscr{C}^2$

$$\left| \begin{array}{l} g(x+h) = g(x) + hg'(x) + \frac{h^2}{2} g''(x) + h^2 \varepsilon_+(h) \\ g(x-h) = g(x) - hg'(x) + \frac{h^2}{2} g''(x) + h^2 \varepsilon_-(h) \end{array}\right.$$

D'où

$$g''(x) = \frac{g(x+h) + g(x-h) - 2g(x)}{h^2} + \varepsilon(h).$$

ÉCOLE CENTRALE LYON

- Problème exact (en x_i)

$$k\frac{-u(x_{i-1}) + 2u(x_i) - u(x_{i+1})}{h^2} + \varepsilon_i(h) = f(x_i).$$

- Problème discret

$$k\frac{-u_{i-1}+2u_i-u_{i+1}}{h^2}=f(x_i).$$

- Problème discret

$$k\frac{-u_{i-1}+2u_i-u_{i+1}}{h^2}=f(x_i).$$

- Aux extrémités

$$k\frac{-u_0+2u_1-u_2}{h^2}=f(x_1),\qquad k\frac{-u_{N-1}+2u_N-u_{N+1}}{h^2}=f(x_N)$$

- **Problème discret**

$$k\frac{-u_{i-1}+2u_i-u_{i+1}}{h^2}=f(x_i).$$

- **Aux extrémités**

$$k\frac{-\alpha+2u_1-u_2}{h^2}=f(x_1),\qquad k\frac{-u_{N-1}+2u_N-u_{N+1}}{h^2}=f(x_N)$$

- Problème discret

$$k\frac{-u_{i-1}+2u_i-u_{i+1}}{h^2}=f(x_i).$$

- Aux extrémités

$$k\frac{-\alpha+2u_1-u_2}{h^2}=f(x_1),\qquad k\frac{-u_{N-1}+2u_N-\beta}{h^2}=f(x_N)$$

- Problème discret

$$k\frac{-u_{i-1}+2u_i-u_{i+1}}{h^2}=f(x_i).$$

- Aux extrémités

$$k\frac{-\alpha+2u_1-u_2}{h^2}=f(x_1),\qquad k\frac{-u_{N-1}+2u_N-\beta}{h^2}=f(x_N)$$

- Écriture matricielle $AU=F$ avec

- Problème discret
$$k\frac{-u_{i-1}+2u_i-u_{i+1}}{h^2}=f(x_i).$$

- Aux extrémités
$$k\frac{-\alpha+2u_1-u_2}{h^2}=f(x_1),\qquad k\frac{-u_{N-1}+2u_N-\beta}{h^2}=f(x_N)$$

- Écriture matricielle $AU=F$ avec
$$A=\frac{k}{h^2}\begin{pmatrix}2 & -1 & & \\ -1 & \ddots & \ddots & \\ & \ddots & \ddots & -1 \\ & & -1 & 2\end{pmatrix},\quad U=\begin{pmatrix}u_1\\ u_2\\ \vdots\\ u_N\end{pmatrix},\quad F=\begin{pmatrix}f(x_1)+k\alpha/h^2\\ f(x_2)\\ \vdots\\ f(x_{N-1})\\ f(x_N)+k\beta/h^2\end{pmatrix}.$$

- Problème-test

$$\begin{cases} -u''(x) = 10\sin(3x), \quad x \in]0,1[, \\ u(0) = 1, \ u(1) = 2. \end{cases}$$

- Solution exacte

$$u(x) = \frac{10}{9}\sin(3x) + \left(1 - \frac{10}{9}\sin(3)\right)x + 1.$$

- [*Lap1D.m*]

- [*Erreur.m*]

Théorème. A est symétrique définie-positive. Ses valeurs propres sont

$$\lambda_\ell = \frac{4}{h^2}\sin^2\left(\frac{\ell\pi h}{2}\right), \quad \ell = 1, 2, \dots N.$$

Théorème. A est symétrique définie-positive. Ses valeurs propres sont

$$\lambda_\ell = \frac{4}{h^2}\sin^2\left(\frac{\ell\pi h}{2}\right), \quad \ell = 1, 2, \dots N.$$

Preuve. Soit $x \in \mathbb{R}^N$.

$$h^2(Ax, x) = 2\sum_{i=1}^{N} x_i^2 - 2\sum_{i=1}^{N-1} x_i x_{i+1} = x_1^2 + \sum_{i=1}^{N-1}(x_i - x_{i+1})^2 + x_N^2 \geqslant 0$$

Si $(Ax, x) = 0$, alors $x_1 = x_N = 0$ et $x_i = x_{i+1}$ pour $i = 1, 2, \dots, N-1$. D'où $x = 0$.

Pour les *vp*, on vérifie que le vecteur suivant est $\overrightarrow{vp}$:

$$v_\ell = \big(\sin(\ell\pi h), \sin(2\ell\pi h), \dots, \sin(N\ell\pi h)\big).$$

$\square$

On pose $U_{\mathrm{ex}} = \big(u(x_1), \ldots, u(x_N)\big)^{\mathsf{T}}$.

Lemme.

$$\|U_{\mathrm{ex}} - U\| \leqslant |||A^{-1}||| \times \|AU_{\mathrm{ex}} - F\|,$$

où

$$|||A^{-1}||| = \sup_{x \neq 0} \frac{\|A^{-1}x\|}{\|x\|}.$$

La quantité $AU_{\mathrm{ex}} - F$ est appelée **résidu**.

On pose $U_{\text{ex}} = \big(u(x_1), \dots, u(x_N)\big)^{\mathsf{T}}$.

Lemme.

$$\|U_{\text{ex}} - U\| \leqslant |||A^{-1}||| \times \|AU_{\text{ex}} - F\|,$$

où

$$|||A^{-1}||| = \sup_{x \neq 0} \frac{\|A^{-1}x\|}{\|x\|}.$$

La quantité $AU_{\text{ex}} - F$ est appelée **résidu**.

Lemme. On a les estimations suivantes

$$|||A^{-1}|||_2 \leqslant \frac{1}{4} \quad \text{et} \quad |||A^{-1}|||_\infty \leqslant \frac{1}{8}.$$

Preuve. On a, pour la norme 2,

$$|||A^{-1}|||_2 = \frac{1}{\lambda_{\min}(A)} = \frac{h^2}{4\sin^2\left(\frac{\pi h}{2}\right)} \leqslant \frac{1}{4}.$$

(en fait $\sim \frac{1}{\pi^2}$). Plus difficile en norme ∞. $\square$

On pose $U_{\mathrm{ex}} = \big(u(x_1), \dots, u(x_N)\big)^{\mathsf{T}}$.

> **Lemme.** Si $u \in \mathscr{C}^4([0,1])$, alors
>
> $$\|AU_{\mathrm{ex}} - F\|_\infty \leqslant \frac{h^2}{12} \sup_{x\in[0,1]} \left|u^{(4)}(x)\right|.$$

On pose $U_{\text{ex}} = \big(u(x_1), \dots, u(x_N)\big)^{\mathsf{T}}$.

Lemme. Si $u \in \mathscr{C}^4([0,1])$, alors

$$\|AU_{\text{ex}} - F\|_\infty \leqslant \frac{h^2}{12} \sup_{x\in[0,1]} \left|u^{(4)}(x)\right|.$$

Preuve. On a, pour $1 \leqslant i \leqslant N$,

$$(AU_{\text{ex}} - F)_i = \frac{-u(x_{i-1}) + 2u(x_i) - u(x_{i+1})}{h^2} - f(x_i).$$

À l'aide d'un développement de Taylor en x_i,

$$(AU_{\text{ex}} - F)_i = \frac{h^2}{24}\left[u^{(4)}(\xi_-) + u^{(4)}(\xi_+)\right].$$

On pose $U_{\text{ex}} = \big(u(x_1), \ldots, u(x_N)\big)^{\mathsf{T}}$.

Lemme. Si $u \in \mathscr{C}^4([0,1])$, alors

$$\|AU_{\text{ex}} - F\|_\infty \leqslant \frac{h^2}{12} \sup_{x\in[0,1]} \left|u^{(4)}(x)\right|.$$

Corollaire. $\|AU_{\text{ex}} - F\|_2 \leqslant \dfrac{h^{\frac{3}{2}}}{12} \sup\limits_{x\in[0,1]} \left|u^{(4)}(x)\right|.$

On pose $U_{\text{ex}} = \big(u(x_1), \dots, u(x_N)\big)^{\mathsf{T}}$.

Lemme. Si $u \in \mathscr{C}^4([0,1])$, alors

$$\|AU_{\text{ex}} - F\|_\infty \leqslant \frac{h^2}{12} \sup_{x\in[0,1]} \left|u^{(4)}(x)\right|.$$

Corollaire. $\|AU_{\text{ex}} - F\|_2 \leqslant \dfrac{h^{\frac{3}{2}}}{12} \sup\limits_{x\in[0,1]} \left|u^{(4)}(x)\right|.$

Preuve. Pour $V \in \mathbb{R}^N$,

$$\|V\|_2^2 = \sum_{i=1}^{N} V_i^2 \leqslant N\|V\|_\infty^2 \leqslant \frac{1}{h}\|V\|_\infty^2.$$

Théorème. Si $u \in \mathscr{C}^4([0,1])$, alors

$$\|U_{\text{ex}} - U\|_\infty \leqslant \frac{h^2}{96} \sup_{x\in[0,1]} \left|u^{(4)}(x)\right|.$$

et

$$\|U_{\text{ex}} - U\|_2 \leqslant \frac{h^{\frac{3}{2}}}{48} \sup_{x\in[0,1]} \left|u^{(4)}(x)\right|.$$

Remarques.

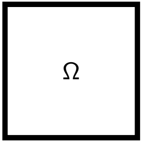
- En norme 2, l'estimation correspondant à $\mathrm{L}^2([0,1])$ est

$$\sqrt{h}\|U_{\text{ex}} - U\|_2 \leqslant \frac{h^2}{48} \sup_{x\in[0,1]} \left|u^{(4)}(x)\right|.$$

- Pour avoir $u \in \mathscr{C}^4$, il suffit d'avoir $f \in \mathscr{C}^2$.

- Problème modèle

$$\begin{cases} -\Delta u = f, & \text{dans } \Omega \\ u = 0, & \text{sur } \partial\Omega. \end{cases}$$

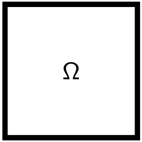


Discrétisation de l'équation de Laplace

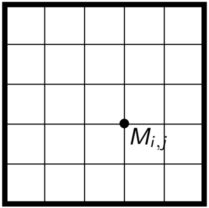
En dimension 2

- **Problème modèle**

$$\begin{cases} -\Delta u = f, & \text{dans } \Omega \\ u = 0, & \text{sur } \partial\Omega. \end{cases}$$



- **Discrétisation**



Pour $N \in \mathbb{N}$,

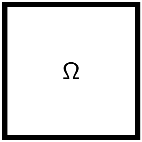
- $h = \frac{1}{N+1}$,
- $M_{i,j} = \left(\dfrac{i}{N+1}, \dfrac{j}{N+1} \right)$.

Discrétisation de l'équation de Laplace

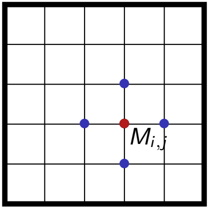
En dimension 2

- Problème modèle

$$\begin{cases} -\Delta u = f, & \text{dans } \Omega \\ u = 0, & \text{sur } \partial\Omega. \end{cases}$$



- Discrétisation



Pour $N \in \mathbb{N}$,

- $h = \frac{1}{N+1}$,
- $M_{i,j} = \left(\dfrac{i}{N+1}, \dfrac{j}{N+1}\right)$.

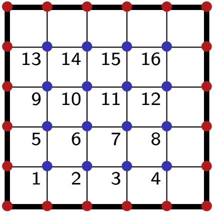
$$-\Delta u\left(M_{i,j}\right) \simeq \frac{4u\left(M_{i,j}\right) - u\left(M_{i-1,j}\right) - u\left(M_{i+1,j}\right) - u\left(M_{i,j-1}\right) - u\left(M_{i,j+1}\right)}{h^2}$$

Discrétisation de l'équation de Laplace

En dimension 2

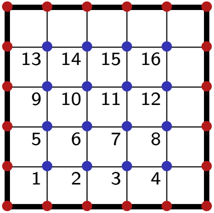
$$-\Delta u\left(M_{i,j}\right) \simeq \frac{4u\left(M_{i,j}\right) - u\left(M_{i-1,j}\right) - u\left(M_{i+1,j}\right) - u\left(M_{i,j-1}\right) - u\left(M_{i,j+1}\right)}{h^2}$$

$$-\Delta u\left(M_{i,j}\right) \simeq \frac{4u\left(M_{i,j}\right) - u\left(M_{i-1,j}\right) - u\left(M_{i+1,j}\right) - u\left(M_{i,j-1}\right) - u\left(M_{i,j+1}\right)}{h^2}$$

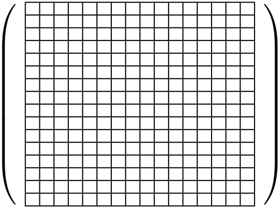


- Inconnues en bleu.
- Données en rouge.
- $\{6, 7, 10, 11\}$: aucun voisin au bord.
- Autres nœuds : interaction avec bord.

$$-\Delta u\left(M_{i,j}\right) \simeq \frac{4u\left(M_{i,j}\right) - u\left(M_{i-1,j}\right) - u\left(M_{i+1,j}\right) - u\left(M_{i,j-1}\right) - u\left(M_{i,j+1}\right)}{h^2}$$



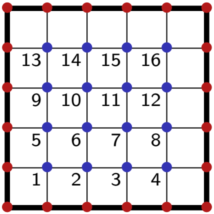
- Inconnues en bleu.
- Données en rouge.
- $\{6, 7, 10, 11\}$: aucun voisin au bord.
- Autres nœuds : interaction avec bord.

$A = \dfrac{1}{h^2}$ 

Discrétisation de l'équation de Laplace

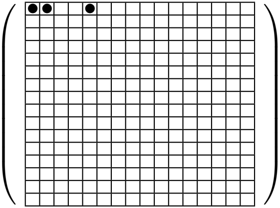
En dimension 2

$$-\Delta u\left(M_{i,j}\right) \simeq \frac{4u\left(M_{i,j}\right) - u\left(M_{i-1,j}\right) - u\left(M_{i+1,j}\right) - u\left(M_{i,j-1}\right) - u\left(M_{i,j+1}\right)}{h^2}$$



- Inconnues en bleu.
- Données en rouge.
- $\{6, 7, 10, 11\}$: aucun voisin au bord.
- Autres nœuds : interaction avec bord.

$$A = \frac{1}{h^2}$$

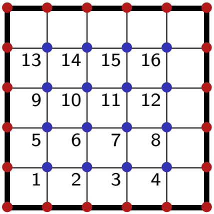


ÉCOLE CENTRALE LYON

Discrétisation de l'équation de Laplace

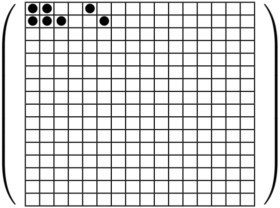
En dimension 2

$$-\Delta u\left(M_{i,j}\right) \simeq \frac{4u\left(M_{i,j}\right) - u\left(M_{i-1,j}\right) - u\left(M_{i+1,j}\right) - u\left(M_{i,j-1}\right) - u\left(M_{i,j+1}\right)}{h^2}$$



- Inconnues en bleu.
- Données en rouge.
- $\{6, 7, 10, 11\}$: aucun voisin au bord.
- Autres nœuds : interaction avec bord.

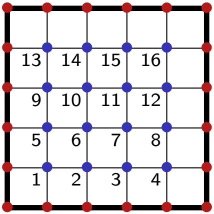
$A = \dfrac{1}{h^2}$



Discrétisation de l'équation de Laplace

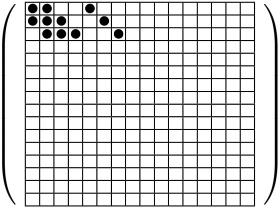
En dimension 2

$$-\Delta u\left(M_{i,j}\right) \simeq \frac{4u\left(M_{i,j}\right) - u\left(M_{i-1,j}\right) - u\left(M_{i+1,j}\right) - u\left(M_{i,j-1}\right) - u\left(M_{i,j+1}\right)}{h^2}$$



- Inconnues en bleu.
- Données en rouge.
- $\{6, 7, 10, 11\}$: aucun voisin au bord.
- Autres nœuds : interaction avec bord.

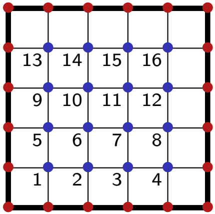
$A = \dfrac{1}{h^2}$



Discrétisation de l'équation de Laplace

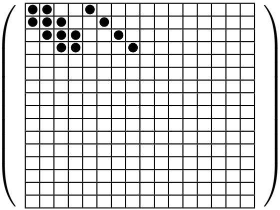
En dimension 2

$$-\Delta u\left(M_{i,j}\right) \simeq \frac{4u\left(M_{i,j}\right) - u\left(M_{i-1,j}\right) - u\left(M_{i+1,j}\right) - u\left(M_{i,j-1}\right) - u\left(M_{i,j+1}\right)}{h^2}$$



- Inconnues en bleu.
- Données en rouge.
- $\{6, 7, 10, 11\}$: aucun voisin au bord.
- Autres nœuds : interaction avec bord.

$$A = \frac{1}{h^2}$$

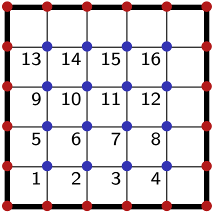


ÉCOLE CENTRALE LYON

Discrétisation de l'équation de Laplace

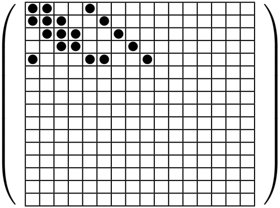
En dimension 2

$$-\Delta u\left(M_{i,j}\right) \simeq \frac{4u\left(M_{i,j}\right) - u\left(M_{i-1,j}\right) - u\left(M_{i+1,j}\right) - u\left(M_{i,j-1}\right) - u\left(M_{i,j+1}\right)}{h^2}$$



- Inconnues en bleu.
- Données en rouge.
- $\{6, 7, 10, 11\}$: aucun voisin au bord.
- Autres nœuds : interaction avec bord.

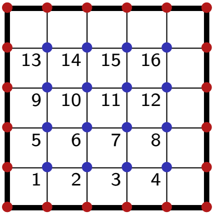
$$A = \frac{1}{h^2}$$



Discrétisation de l'équation de Laplace

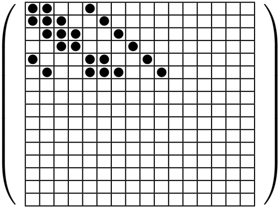
En dimension 2

$$-\Delta u\left(M_{i,j}\right) \simeq \frac{4u\left(M_{i,j}\right) - u\left(M_{i-1,j}\right) - u\left(M_{i+1,j}\right) - u\left(M_{i,j-1}\right) - u\left(M_{i,j+1}\right)}{h^2}$$



- Inconnues en bleu.
- Données en rouge.
- $\{6, 7, 10, 11\}$: aucun voisin au bord.
- Autres nœuds : interaction avec bord.

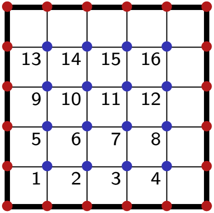
$A = \dfrac{1}{h^2}$



Discrétisation de l'équation de Laplace

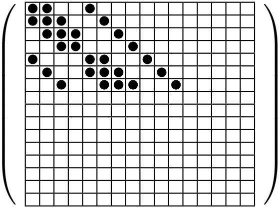
En dimension 2

$$-\Delta u\left(M_{i,j}\right) \simeq \frac{4u\left(M_{i,j}\right) - u\left(M_{i-1,j}\right) - u\left(M_{i+1,j}\right) - u\left(M_{i,j-1}\right) - u\left(M_{i,j+1}\right)}{h^2}$$



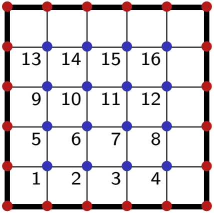
- Inconnues en bleu.
- Données en rouge.
- $\{6, 7, 10, 11\}$: aucun voisin au bord.
- Autres nœuds : interaction avec bord.

$A = \dfrac{1}{h^2}$



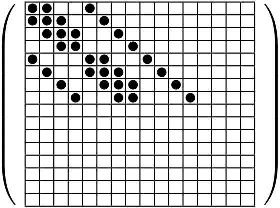
ÉCOLE CENTRALE LYON

$$-\Delta u\left(M_{i,j}\right) \simeq \frac{4u\left(M_{i,j}\right) - u\left(M_{i-1,j}\right) - u\left(M_{i+1,j}\right) - u\left(M_{i,j-1}\right) - u\left(M_{i,j+1}\right)}{h^2}$$



- Inconnues en bleu.
- Données en rouge.
- $\{6, 7, 10, 11\}$: aucun voisin au bord.
- Autres nœuds : interaction avec bord.

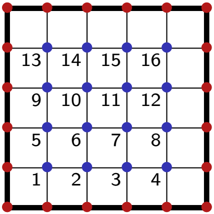
$A = \dfrac{1}{h^2}$



Discrétisation de l'équation de Laplace

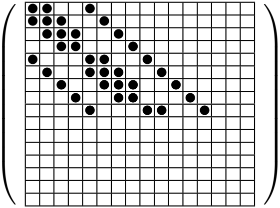
En dimension 2

$$-\Delta u\left(M_{i,j}\right) \simeq \frac{4u\left(M_{i,j}\right) - u\left(M_{i-1,j}\right) - u\left(M_{i+1,j}\right) - u\left(M_{i,j-1}\right) - u\left(M_{i,j+1}\right)}{h^2}$$

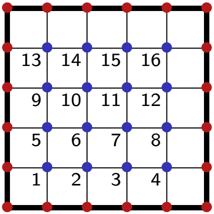


- Inconnues en bleu.
- Données en rouge.
- $\{6, 7, 10, 11\}$: aucun voisin au bord.
- Autres nœuds : interaction avec bord.

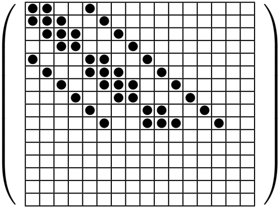
$A = \dfrac{1}{h^2}$



$$-\Delta u\left(M_{i,j}\right) \simeq \frac{4u\left(M_{i,j}\right) - u\left(M_{i-1,j}\right) - u\left(M_{i+1,j}\right) - u\left(M_{i,j-1}\right) - u\left(M_{i,j+1}\right)}{h^2}$$



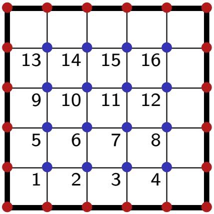
- Inconnues en bleu.
- Données en rouge.
- $\{6, 7, 10, 11\}$: aucun voisin au bord.
- Autres nœuds : interaction avec bord.

$A = \dfrac{1}{h^2}$ 

Discrétisation de l'équation de Laplace

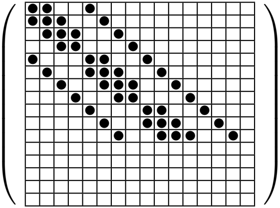
En dimension 2

$$-\Delta u\left(M_{i,j}\right) \simeq \frac{4u\left(M_{i,j}\right) - u\left(M_{i-1,j}\right) - u\left(M_{i+1,j}\right) - u\left(M_{i,j-1}\right) - u\left(M_{i,j+1}\right)}{h^2}$$



- Inconnues en bleu.
- Données en rouge.
- $\{6, 7, 10, 11\}$: aucun voisin au bord.
- Autres nœuds : interaction avec bord.

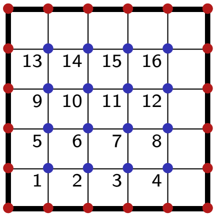
$A = \frac{1}{h^2}$



Discrétisation de l'équation de Laplace

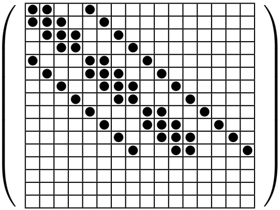
En dimension 2

$$-\Delta u\left(M_{i,j}\right) \simeq \frac{4u\left(M_{i,j}\right) - u\left(M_{i-1,j}\right) - u\left(M_{i+1,j}\right) - u\left(M_{i,j-1}\right) - u\left(M_{i,j+1}\right)}{h^2}$$



- Inconnues en bleu.
- Données en rouge.
- $\{6, 7, 10, 11\}$: aucun voisin au bord.
- Autres nœuds : interaction avec bord.

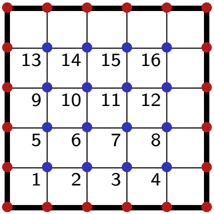
$$A = \frac{1}{h^2}$$



Discrétisation de l'équation de Laplace

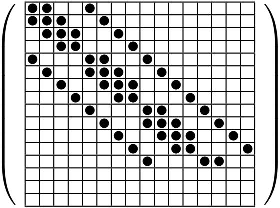
En dimension 2

$$-\Delta u\left(M_{i,j}\right) \simeq \frac{4u\left(M_{i,j}\right) - u\left(M_{i-1,j}\right) - u\left(M_{i+1,j}\right) - u\left(M_{i,j-1}\right) - u\left(M_{i,j+1}\right)}{h^2}$$



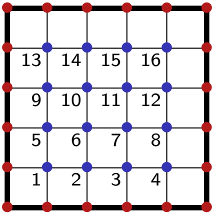
- Inconnues en bleu.
- Données en rouge.
- $\{6, 7, 10, 11\}$: aucun voisin au bord.
- Autres nœuds : interaction avec bord.

$A = \dfrac{1}{h^2}$



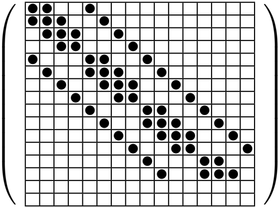
Discrétisation de l'équation de Laplace
En dimension 2

$$-\Delta u\left(M_{i,j}\right) \simeq \frac{4u\left(M_{i,j}\right) - u\left(M_{i-1,j}\right) - u\left(M_{i+1,j}\right) - u\left(M_{i,j-1}\right) - u\left(M_{i,j+1}\right)}{h^2}$$



- Inconnues en bleu.
- Données en rouge.
- $\{6, 7, 10, 11\}$: aucun voisin au bord.
- Autres nœuds : interaction avec bord.

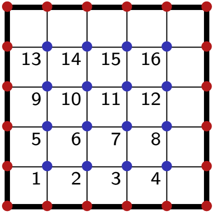
$A = \frac{1}{h^2}$



Discrétisation de l'équation de Laplace

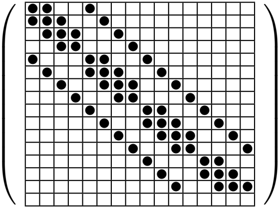
En dimension 2

$$-\Delta u\left(M_{i,j}\right) \simeq \frac{4u\left(M_{i,j}\right) - u\left(M_{i-1,j}\right) - u\left(M_{i+1,j}\right) - u\left(M_{i,j-1}\right) - u\left(M_{i,j+1}\right)}{h^2}$$



- Inconnues en bleu.
- Données en rouge.
- $\{6, 7, 10, 11\}$: aucun voisin au bord.
- Autres nœuds : interaction avec bord.

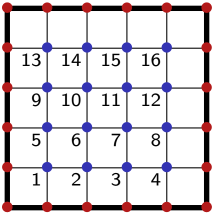
$A = \dfrac{1}{h^2}$



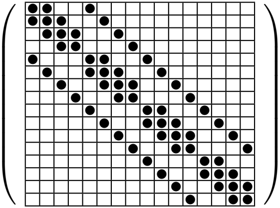
Discrétisation de l'équation de Laplace

En dimension 2

$$-\Delta u\left(M_{i,j}\right) \simeq \frac{4u\left(M_{i,j}\right) - u\left(M_{i-1,j}\right) - u\left(M_{i+1,j}\right) - u\left(M_{i,j-1}\right) - u\left(M_{i,j+1}\right)}{h^2}$$



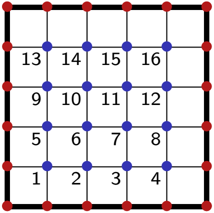
- Inconnues en bleu.
- Données en rouge.
- $\{6, 7, 10, 11\}$: aucun voisin au bord.
- Autres nœuds : interaction avec bord.

$A = \frac{1}{h^2}$ 

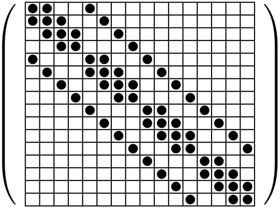
Discrétisation de l'équation de Laplace

En dimension 2

$$-\Delta u\left(M_{i,j}\right) \simeq \frac{4u\left(M_{i,j}\right) - u\left(M_{i-1,j}\right) - u\left(M_{i+1,j}\right) - u\left(M_{i,j-1}\right) - u\left(M_{i,j+1}\right)}{h^2}$$



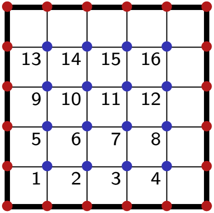
- Inconnues en bleu.
- Données en rouge.
- $\{6, 7, 10, 11\}$: aucun voisin au bord.
- Autres nœuds : interaction avec bord.

$A = \dfrac{1}{h^2}$ 

- Élements diagonaux $= 4$.

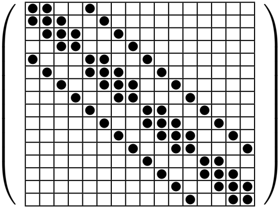
Discrétisation de l'équation de Laplace
En dimension 2

$$-\Delta u\left(M_{i,j}\right) \simeq \frac{4u\left(M_{i,j}\right) - u\left(M_{i-1,j}\right) - u\left(M_{i+1,j}\right) - u\left(M_{i,j-1}\right) - u\left(M_{i,j+1}\right)}{h^2}$$



- Inconnues en bleu.
- Données en rouge.
- $\{6, 7, 10, 11\}$: aucun voisin au bord.
- Autres nœuds : interaction avec bord.

$A = \dfrac{1}{h^2}$



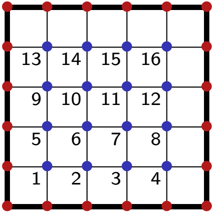
- Élements diagonaux $= 4$.
- Élements hors-diagonaux $= -1$.

ÉCOLE CENTRALE LYON

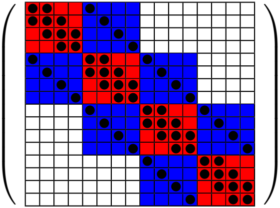
Discrétisation de l'équation de Laplace

En dimension 2

$$-\Delta u\left(M_{i,j}\right) \simeq \frac{4u\left(M_{i,j}\right) - u\left(M_{i-1,j}\right) - u\left(M_{i+1,j}\right) - u\left(M_{i,j-1}\right) - u\left(M_{i,j+1}\right)}{h^2}$$



- Inconnues en bleu.
- Données en rouge.
- $\{6, 7, 10, 11\}$: aucun voisin au bord.
- Autres nœuds : interaction avec bord.

$A = \dfrac{1}{h^2}$ 

- Élements diagonaux $= 4$.
- Élements hors-diagonaux $= -1$.
- Matrice tridiagonale par blocs.

Cas général.

$$A = \frac{1}{h^2}\left(\begin{array}{c|c|c|c} H & -I & & \\ \hline -I & \ddots & \ddots & \\ \hline & \ddots & \ddots & -I \\ \hline & & -I & H \end{array}\right) \in \mathbb{R}^{N^2 \times N^2},$$

avec

$$H = \begin{pmatrix} 4 & -1 & & \\ -1 & \ddots & \ddots & \\ & \ddots & \ddots & -1 \\ & & -1 & 4 \end{pmatrix} \in \mathbb{R}^{N \times N}, \quad I = I_N \in \mathbb{R}^{N \times N}.$$

Discrétisation de l'équation de Laplace

En dimension 2

- Comme en dimension 1...

> **Théorème.** Si $u \in \mathscr{C}^4(\overline{\Omega})$, alors
>
> $$\|U_{\text{ex}} - U\|_\infty \leqslant \frac{h^2}{48} \sup_{(x,y)\in\overline{\Omega}} \left(\left| \frac{\partial^{(4)} u}{\partial x^4}(x,y) \right| + \left| \frac{\partial^{(4)} u}{\partial y^4}(x,y) \right| \right).$$

- Comme en dimension 1...

> **Théorème.** Si $u \in \mathscr{C}^4(\overline{\Omega})$, alors
> $$\|U_{\mathrm{ex}} - U\|_\infty \leqslant \frac{h^2}{48} \sup_{(x,y)\in\overline{\Omega}} \left(\left| \frac{\partial^{(4)} u}{\partial x^4}(x,y) \right| + \left| \frac{\partial^{(4)} u}{\partial y^4}(x,y) \right| \right).$$

- ... mais des difficultés supplémentaires

- Comme en dimension 1…

> **Théorème.** Si $u \in \mathscr{C}^4(\overline{\Omega})$, alors
> $$\|U_{ex} - U\|_\infty \leqslant \frac{h^2}{48} \sup_{(x,y)\in\overline{\Omega}} \left(\left| \frac{\partial^{(4)} u}{\partial x^4}(x,y) \right| + \left| \frac{\partial^{(4)} u}{\partial y^4}(x,y) \right| \right).$$

- … mais des difficultés supplémentaires
  - L'hypothèse $f \in \mathscr{C}^2(\overline{\Omega})$ n'implique $u \in \mathscr{C}^4(\overline{\Omega})$ que si Ω est régulier.

- Comme en dimension 1...

> **Théorème.** Si $u \in \mathscr{C}^4(\overline{\Omega})$, alors
> $$\|U_{ex} - U\|_\infty \leqslant \frac{h^2}{48} \sup_{(x,y)\in\overline{\Omega}} \left(\left| \frac{\partial^{(4)} u}{\partial x^4}(x,y) \right| + \left| \frac{\partial^{(4)} u}{\partial y^4}(x,y) \right| \right).$$

- ... mais des difficultés supplémentaires
  - L'hypothèse $f \in \mathscr{C}^2(\overline{\Omega})$ n'implique $u \in \mathscr{C}^4(\overline{\Omega})$ que si Ω est régulier.
  - La matrice est creuse, mais la numérotation influe sur sa structure.

- Comme en dimension 1...

> **Théorème.** Si $u \in \mathscr{C}^4(\overline{\Omega})$, alors
> $$\|U_{ex} - U\|_\infty \leqslant \frac{h^2}{48} \sup_{(x,y)\in\overline{\Omega}} \left(\left| \frac{\partial^{(4)} u}{\partial x^4}(x,y) \right| + \left| \frac{\partial^{(4)} u}{\partial y^4}(x,y) \right| \right).$$

- ... mais des difficultés supplémentaires
  - L'hypothèse $f \in \mathscr{C}^2(\overline{\Omega})$ n'implique $u \in \mathscr{C}^4(\overline{\Omega})$ que si Ω est régulier.
  - La matrice est creuse, mais la numérotation influe sur sa structure.
  - La prise en compte de géométries complexes n'est pas aisée.

- Comme en dimension 1...

> **Théorème.** Si $u \in \mathscr{C}^4(\overline{\Omega})$, alors
>
> $$\|U_{ex} - U\|_\infty \leqslant \frac{h^2}{48} \sup_{(x,y)\in\overline{\Omega}} \left(\left| \frac{\partial^{(4)} u}{\partial x^4}(x,y) \right| + \left| \frac{\partial^{(4)} u}{\partial y^4}(x,y) \right| \right).$$

- ... mais des difficultés supplémentaires
  - L'hypothèse $f \in \mathscr{C}^2(\overline{\Omega})$ n'implique $u \in \mathscr{C}^4(\overline{\Omega})$ que si Ω est régulier.
  - La matrice est creuse, mais la numérotation influe sur sa structure.
  - La prise en compte de géométries complexes n'est pas aisée.
- Problème aux valeurs propres associé à A
  - vibrations de la membrane.

COURS 6

Discrétisation des EDP

l'équation de transport

Discrétisation de l'équation de transport

Pollution d'une rivière

Comment un polluant est-il transporté par le courant d'une rivière ?

Quel est l'impact du courant sur sa concentration ?

Comment anticiper les zones où la concentration sera la plus forte ?

Discrétisation de l'équation de transport

Modélisation

- L'espace (la rivière) est modélisé par $\Omega = \mathbb{R}$.

- L'espace (la rivière) est modélisé par $\Omega = \mathbb{R}$.
- Le temps est modélisé par l'intervalle $[0, T]$.

- L'espace (la rivière) est modélisé par $\Omega = \mathbb{R}$.
- Le temps est modélisé par l'intervalle $[0, T]$.
- L'inconnue est la concentration de polluant $u(x, t)$ avec

$$u : \Omega \times [0, T] \to \mathbb{R}^+$$

- L'espace (la rivière) est modélisé par $\Omega = \mathbb{R}$.
- Le temps est modélisé par l'intervalle $[0, T]$.
- L'inconnue est la concentration de polluant $u(x, t)$ avec
$$u : \Omega \times [0, T] \to \mathbb{R}^+$$
- Cette quantité évolue car soumise a un champ de vitesse $v(x, t)$ supposé connu.

- L'espace (la rivière) est modélisé par $\Omega = \mathbb{R}$.
- Le temps est modélisé par l'intervalle $[0, T]$.
- L'inconnue est la concentration de polluant $u(x, t)$ avec
$$u : \Omega \times [0, T] \to \mathbb{R}^+$$
- Cette quantité évolue car soumise a un champ de vitesse $v(x, t)$ supposé connu.
- Loi : Conservation de la masse.

Discrétisation de l'équation de transport

Conservation de la masse

La quantité de polluant dans un intervalle quelconque $[a, b] \subset \mathbb{R}$, au temps t est

$$Q(t) = \int_a^b u(x,t)\,\mathrm{d}x$$

La quantité de polluant dans un intervalle quelconque $[a, b] \subset \mathbb{R}$, au temps t est

$$Q(t) = \int_a^b u(x, t)\, \mathrm{d}x$$

Bilan entrées-sorties :

$$Q'(t) = u(a, t)v(a, t) - u(b, t)v(b, t).$$

Discrétisation de l'équation de transport

Conservation de la masse

La quantité de polluant dans un intervalle quelconque $[a, b] \subset \mathbb{R}$, au temps t est

$$Q(t) = \int_a^b u(x,t)\,\mathrm{d}x$$

Bilan entrées-sorties :

$$Q'(t) = u(a,t)v(a,t) - u(b,t)v(b,t).$$

D'autre part, $Q'(t) = \displaystyle\int_a^b \partial_t u(x,t)\,\mathrm{d}x$ et ainsi

Discrétisation de l'équation de transport

Conservation de la masse

La quantité de polluant dans un intervalle quelconque $[a, b] \subset \mathbb{R}$, au temps t est

$$Q(t) = \int_a^b u(x,t)\,\mathrm{d}x$$

Bilan entrées-sorties :

$$Q'(t) = u(a,t)v(a,t) - u(b,t)v(b,t).$$

D'autre part, $Q'(t) = \displaystyle\int_a^b \partial_t u(x,t)\,\mathrm{d}x$ et ainsi

$$\int_a^b \partial_t u(x,t)\,\mathrm{d}x + u(b,t)v(b,t) - u(a,t)v(a,t) = 0.$$

Conservation de la masse

$$\int_a^b \partial_t u(x,t)\,\mathrm{d}x + u(b,t)v(b,t) - u(a,t)v(a,t) = 0,$$

$$\int_a^b \left[\partial_t u(x,t) + \partial_x(uv)(x,t)\right]\,\mathrm{d}x = 0, \quad \forall t \in [0,T]$$

Ce calcul est valable pour tout intervalle $[a,b]$, donc on peut écrire directement

$$\boxed{\partial_t u(x,t) + \partial_x(uv)(x,t) = 0, \quad \forall (x,t) \in \mathbb{R} \times [0,T]}$$

C'est une EDP linéaire d'ordre 1 aussi appelée **équation de transport linéaire**

Remarque. l'EDP peut se réécrire

$$\partial_t u + v\partial_x u + u\partial_x v = 0, \quad \text{sur } \mathbb{R} \times [0,T].$$

Définition. On appelle équation de transport linéaire en dimension 1 l'EDP suivante

$$\begin{cases} \partial_t u(x,t) + v(x,t)\partial_x u(x,t) + a(x,t)u(x,t) = f(x,t), \\ u(x,0) = u_0(x), \end{cases}$$

où v, a et f sont des fonctions données continues sur $[0,T] \times \mathbb{R}$ et l'inconnue u est une fonction de classe $\mathscr{C}^1([0,T] \times \mathbb{R}, \mathbb{R})$.

Remarques :

Définition. On appelle équation de transport linéaire en dimension 1 l'EDP suivante

$$\begin{cases} \partial_t u(x,t) + v(x,t)\partial_x u(x,t) + a(x,t)u(x,t) = f(x,t), \\ u(x,0) = u_0(x), \end{cases}$$

où v, a et f sont des fonctions données continues sur $[0,T] \times \mathbb{R}$ et l'inconnue u est une fonction de classe $\mathscr{C}^1([0,T] \times \mathbb{R}, \mathbb{R})$.

Remarques :

- v est homogène à une vitesse et s'appelle le champ de vitesse de l'équation.

Définition. On appelle équation de transport linéaire en dimension 1 l'EDP suivante

$$\left\{\begin{array}{l} \partial_t u(x,t) + v(x,t)\partial_x u(x,t) + a(x,t)u(x,t) = f(x,t), \\ \\ u(x,0) = u_0(x), \end{array}\right.$$

où v, a et f sont des fonctions données continues sur $[0,T] \times \mathbb{R}$ et l'inconnue u est une fonction de classe $\mathscr{C}^1([0,T] \times \mathbb{R}, \mathbb{R})$.

Remarques :

- v est homogène à une vitesse et s'appelle le champ de vitesse de l'équation.
- Si $f = 0$, l'équation est dite homogène.

Définition. On appelle <u>équation de transport linéaire en dimension 1</u> l'EDP suivante

$$\begin{cases} \partial_t u(x,t) + v(x,t)\partial_x u(x,t) + a(x,t)u(x,t) = f(x,t), \\ u(x,0) = u_0(x), \end{cases}$$

où v, a et f sont des fonctions données continues sur $[0,T] \times \mathbb{R}$ et l'inconnue u est une fonction de classe $\mathscr{C}^1([0,T] \times \mathbb{R}, \mathbb{R})$.

Remarques :

- v est homogène à une vitesse et s'appelle le champ de vitesse de l'équation.
- Si $f = 0$, l'équation est dite homogène.
- Dans le modèle vu plus haut, les coefficients a et v sont liés.

Méthode des caractéristiques

Discrétisation de l'équation de transport

Cas vitesse constante, $a = 0$ et $f = 0$

Supposons le champ de vitesse constant $v(x, t) = c$, $a = 0$ et $f = 0$.

$$(\mathcal{T}) \qquad \begin{cases} \partial_t u + c\partial_x u = 0, & \text{sur } \mathbb{R} \times [0, T], \\ u(x, 0) = u_0(x), & \text{pour } x \in \mathbb{R}. \end{cases}$$

Supposons le champ de vitesse constant $v(x,t) = c$, $a = 0$ et $f = 0$.

$$(\mathcal{T}) \qquad \begin{cases} \partial_t u + c\partial_x u = 0, & \text{sur } \mathbb{R} \times [0,T], \\ u(x,0) = u_0(x), & \text{pour } x \in \mathbb{R}. \end{cases}$$

- La solution satisfaisant $u(x,0) = u_0(x)$ est donnée par

$$u(x,t) = u_0(x - ct).$$

Discrétisation de l'équation de transport

Cas vitesse constante, $a = 0$ et $f = 0$

Supposons le champ de vitesse constant $v(x,t) = c$, $a = 0$ et $f = 0$.

$$(\mathcal{T}) \qquad \begin{cases} \partial_t u + c\partial_x u = 0, & \text{sur } \mathbb{R} \times [0,T], \\ u(x,0) = u_0(x), & \text{pour } x \in \mathbb{R}. \end{cases}$$

- La solution satisfaisant $u(x,0) = u_0(x)$ est donnée par

$$u(x,t) = u_0(x - ct).$$

- Autrement dit, si on note $X(t) = k + ct$, alors

$$t \mapsto u(X(t),t) \text{ est constante.}$$

Définition. On appelle caractéristiques associées à l'équation $(\mathcal{T})$ toutes les solutions de l'équation différentielle

$$X'(t) = c.$$

Leurs graphes $t \to (X(t), t)$ sont appelés courbes caractéristiques.

Discrétisation de l'équation de transport

Cas vitesse constante, $a = 0$ et $f = 0$

Définition. On appelle caractéristiques associées à l'équation $(\mathcal{T})$ toutes les solutions de l'équation différentielle

$$X'(t) = c.$$

Leurs graphes $t \to (X(t), t)$ sont appelés courbes caractéristiques.

Remarque : Ce sont les fonctions affines $X(t) = ct + k$, $k \in \mathbb{R}$.

Définition. On appelle caractéristiques associées à l'équation $(\mathcal{T})$ toutes les solutions de l'équation différentielle

$$X'(t) = c.$$

Leurs graphes $t \to (X(t), t)$ sont appelés courbes caractéristiques.

Remarque : Ce sont les fonctions affines $X(t) = ct + k$, $k \in \mathbb{R}$.

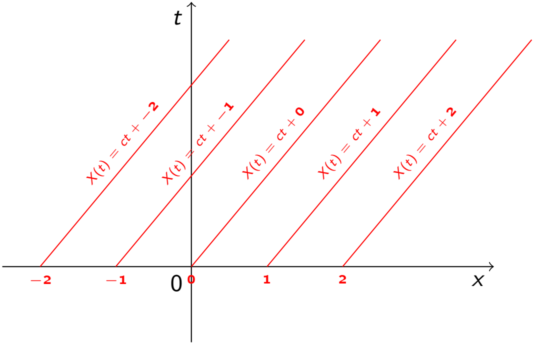
Proposition. La solution $u \in \mathscr{C}^1([0, T] \times \mathbb{R})$ de $(\mathcal{T})$, est constante sur les courbes caractéristiques : si X est une caractéristique associée à $(\mathcal{T})$, alors la fonction

$$t \mapsto u(X(t), t)$$

est constante.

Discrétisation de l'équation de transport

Cas vitesse constante, $a = 0$ et $f = 0$



- Les courbes caractéristiques sont des courbes de niveau de la solution u sur $\mathbb{R} \times [0, T]$.
- Remarque : Si on connaît la solution u pour $t = 0$ alors on la connaît en tout temps.

Retour à notre équation de transport :

$$(\mathcal{T}) \qquad \begin{cases} \partial_t u + c\partial_x u = 0, & \text{sur } \mathbb{R} \times [0, T], \\ u(x, 0) = u_0(x), & \forall x \in \mathbb{R}. \end{cases}$$

Retour à notre équation de transport :

$$(\mathcal{T}) \qquad \begin{cases} \partial_t u + c\partial_x u = 0, & \text{sur } \mathbb{R} \times [0, T], \\ u(x, 0) = u_0(x), & \forall x \in \mathbb{R}. \end{cases}$$

Méthodologie : pour trouver la valeur de la solution u en un point quelconque $(x_0, t_0) \in \mathbb{R} \times [0, T]$,

- on identifie l'unique caractéristique X_{x_0, t_0} qui passe par x_0 en $t = t_0$,

Retour à notre équation de transport :

$$(\mathcal{T}) \qquad \begin{cases} \partial_t u + c\partial_x u = 0, & \text{sur } \mathbb{R} \times [0, T], \\ u(x,0) = u_0(x), & \forall x \in \mathbb{R}. \end{cases}$$

Méthodologie : pour trouver la valeur de la solution u en un point quelconque $(x_0, t_0) \in \mathbb{R} \times [0, T]$,

- on identifie l'unique caractéristique X_{x_0,t_0} qui passe par x_0 en $t = t_0$,
- on évalue u à l'origine de cette caractéristique : $u_0(X_{x_0,t_0}(0))$,

Retour à notre équation de transport :

$$(\mathcal{T}) \qquad \begin{cases} \partial_t u + c\partial_x u = 0, & \text{sur } \mathbb{R} \times [0, T], \\ u(x, 0) = u_0(x), & \forall x \in \mathbb{R}. \end{cases}$$

Méthodologie : pour trouver la valeur de la solution u en un point quelconque $(x_0, t_0) \in \mathbb{R} \times [0, T]$,

- on identifie l'unique caractéristique X_{x_0, t_0} qui passe par x_0 en $t = t_0$,
- on évalue u à l'origine de cette caractéristique : $u_0(X_{x_0, t_0}(0))$,
- on reporte cette valeur au point (x_0, t_0) :

$$u(x_0, t_0) = u_0(X_{x_0, t_0}(0)).$$

Retour à notre équation de transport :

$$(\mathcal{T}) \qquad \begin{cases} \partial_t u + c\partial_x u = 0, & \text{sur } \mathbb{R}\times[0,T], \\ u(x,0) = u_0(x), & \forall x \in \mathbb{R}. \end{cases}$$

Méthodologie : pour trouver la valeur de la solution u en un point quelconque $(x_0, t_0) \in \mathbb{R}\times[0,T]$,

- on identifie l'unique caractéristique X_{x_0,t_0} qui passe par x_0 en $t = t_0$,
- on évalue u à l'origine de cette caractéristique : $u_0(X_{x_0,t_0}(0))$,
- on reporte cette valeur au point (x_0, t_0) :

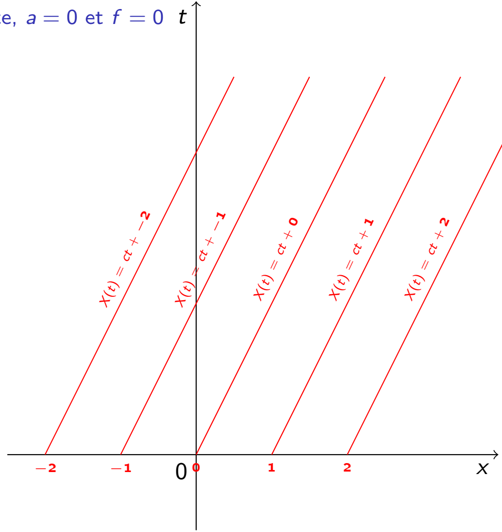
$$u(x_0, t_0) = u_0(X_{x_0,t_0}(0)).$$

Ici : $X_{x_0,t_0}(t) = ct + x_0 - ct_0$ passe par (x_0, t_0). À l'origine elle vaut $X_{x_0,t_0}(0) = x_0 - ct_0$. Ainsi,

$$u(x_0, t_0) = u_0(x_0 - ct_0).$$

Discrétisation de l'équation de transport

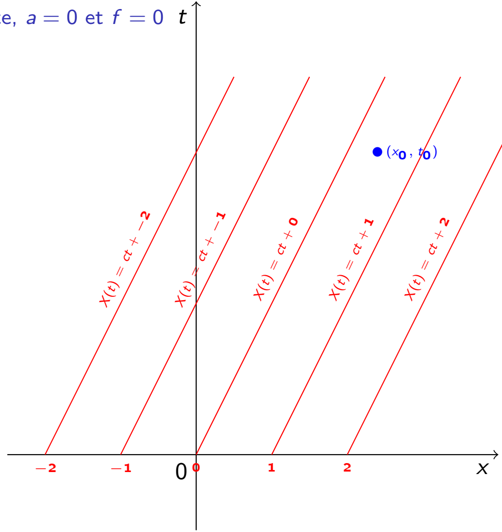
Cas vitesse constante, $a = 0$ et $f = 0$



$$u(x_0, t_0) = u_0(x_0 - ct_0).$$

Discrétisation de l'équation de transport

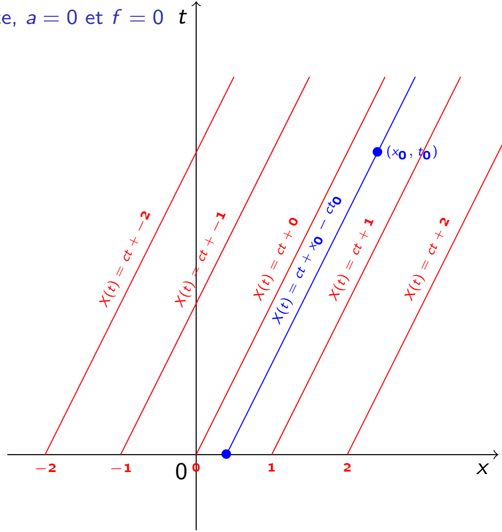
Cas vitesse constante, $a = 0$ et $f = 0$



$$u(x_0, t_0) = u_0(x_0 - ct_0).$$

Discrétisation de l'équation de transport

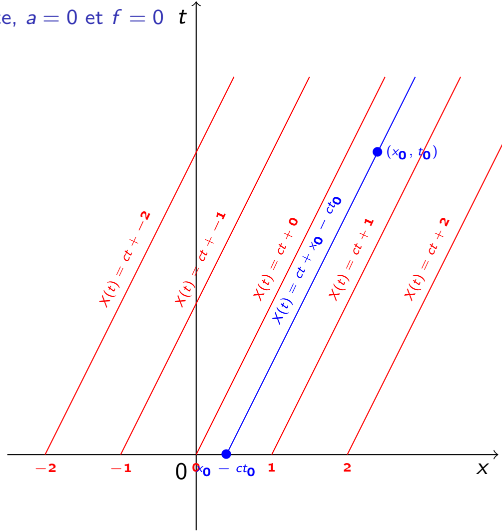
Cas vitesse constante, $a = 0$ et $f = 0$



$$u(x_0, t_0) = u_0(x_0 - ct_0).$$

Discrétisation de l'équation de transport

Cas vitesse constante, $a = 0$ et $f = 0$



$$u(x_0, t_0) = u_0(x_0 - ct_0).$$

Proposition. Soit $c \in \mathbb{R}$, et $u_0 \in \mathscr{C}^1(\mathbb{R})$, le problème de transport

$$(\mathcal{T}) \qquad \begin{cases} \partial_t u + c\partial_x u = 0, & \text{sur } \mathbb{R} \times [0, T], \\ u(x,0) = u_0(x), & \forall x \in \mathbb{R}. \end{cases}$$

admet une unique solution $u \in \mathscr{C}^1(\mathbb{R} \times [0, T])$ donnée par

$$u(x,t) = u_0(x - ct).$$

Proposition. Soit $c \in \mathbb{R}$, et $u_0 \in \mathscr{C}^1(\mathbb{R})$, le problème de transport

$$(\mathcal{T}) \qquad \begin{cases} \partial_t u + c\partial_x u = 0, & \text{sur } \mathbb{R} \times [0, T], \\ u(x,0) = u_0(x), & \forall x \in \mathbb{R}. \end{cases}$$

admet une unique solution $u \in \mathscr{C}^1(\mathbb{R} \times [0, T])$ donnée par

$$u(x, t) = u_0(x - ct).$$

Exemple : La solution du problème

$$\begin{cases} \partial_t u + 2\partial_x u = 0, & \text{sur } \mathbb{R} \times [0, T], \\ u(x,0) = e^{-x^2}, & \forall x \in \mathbb{R}. \end{cases}$$

est

$$u(x, t) = e^{-(x-2t)^2}.$$

On choisit toujours que $a = 0$ et $f = 0$, et on suppose :

$$\forall x, y \in \mathbb{R},\ \forall t \in [0, T], \quad |v(x,t) - v(y,t)| \leq k|x - y|.$$

On considère

$$(\mathcal{T}) \qquad \begin{cases} \partial_t u + v\partial_x u = 0, & \text{sur } \mathbb{R} \times [0, T], \\ u(x, 0) = u_0(x), & \forall x \in \mathbb{R}. \end{cases}$$

On choisit toujours que $a = 0$ et $f = 0$, et on suppose :

$$\forall x, y \in \mathbb{R},\ \forall t \in [0, T], \quad |v(x,t) - v(y,t)| \leq k|x - y|.$$

On considère

$$(\mathcal{T}) \qquad \begin{cases} \partial_t u + v \partial_x u = 0, & \text{sur } \mathbb{R} \times [0, T], \\ u(x, 0) = u_0(x), & \forall x \in \mathbb{R}. \end{cases}$$

Définition. On appelle caractéristiques associées à l'équation $(\mathcal{T})$ toutes les solutions de l'équation différentielle

$$X'(t) = v(X(t), t).$$

Leurs graphes $t \mapsto (X(t), t)$ sont appelés courbes caractéristiques.

Discrétisation de l'équation de transport

Cas vitesse non constante, $a = 0$ et $f = 0$

On choisit toujours que $a = 0$ et $f = 0$, et on suppose :

$$\forall x, y \in \mathbb{R},\ \forall t \in [0, T], \quad |v(x,t) - v(y,t)| \leq k|x - y|.$$

On considère

$$(\mathcal{T}) \qquad \begin{cases} \partial_t u + v\partial_x u = 0, & \text{sur } \mathbb{R} \times [0, T], \\ u(x, 0) = u_0(x), & \forall x \in \mathbb{R}. \end{cases}$$

Définition. On appelle caractéristiques associées à l'équation $(\mathcal{T})$ toutes les solutions de l'équation différentielle

$$X'(t) = v(X(t), t).$$

Leurs graphes $t \mapsto (X(t), t)$ sont appelés courbes caractéristiques.

Remarque : Cauchy-Lipschitz $\implies X \in \mathscr{C}^1([0, T])$

Discrétisation de l'équation de transport

Cas vitesse non constante, $a = 0$ et $f = 0$

Proposition. Toute solution $u \in \mathscr{C}^1([0, T] \times \mathbb{R})$ de $(\mathcal{T})$, est constante sur les courbes caractéristiques : si X est une caractéristique associée à $(\mathcal{T})$, alors la fonction

$$t \mapsto u(X(t), t)$$

est constante.

Cas vitesse non constante, $a = 0$ et $f = 0$

Proposition. Toute solution $u \in \mathscr{C}^1([0, T] \times \mathbb{R})$ de $(\mathcal{T})$, est constante sur les courbes caractéristiques : si X est une caractéristique associée à $(\mathcal{T})$, alors la fonction

$$t \mapsto u(X(t), t)$$

est constante.

Preuve : On pose $\varphi(t) = u(X(t), t)$. Par dérivation composée,

$$\varphi'(t) = \partial_x u(X(t), t) X'(t) + \partial_t u(X(t), t)$$

Cas vitesse non constante, $a = 0$ et $f = 0$

Proposition. Toute solution $u \in \mathscr{C}^1([0, T] \times \mathbb{R})$ de $(\mathcal{T})$, est constante sur les courbes caractéristiques : si X est une caractéristique associée à $(\mathcal{T})$, alors la fonction

$$t \mapsto u(X(t), t)$$

est constante.

Preuve : On pose $\varphi(t) = u(X(t), t)$. Par dérivation composée,

$$\begin{aligned}\varphi'(t) &= \partial_x u(X(t), t) X'(t) + \partial_t u(X(t), t) \\ &= \partial_x u(X(t), t) v(X(t), t) + \partial_t u(X(t), t)\end{aligned}$$

Cas vitesse non constante, $a = 0$ et $f = 0$

Proposition. Toute solution $u \in \mathscr{C}^1([0, T] \times \mathbb{R})$ de $(\mathcal{T})$, est constante sur les courbes caractéristiques : si X est une caractéristique associée à $(\mathcal{T})$, alors la fonction

$$t \mapsto u(X(t), t)$$

est constante.

Preuve : On pose $\varphi(t) = u(X(t), t)$. Par dérivation composée,

$$\begin{aligned}
\varphi'(t) &= \partial_x u(X(t), t) X'(t) + \partial_t u(X(t), t) \\
&= \partial_x u(X(t), t) v(X(t), t) + \partial_t u(X(t), t) \\
&= (v \partial_x u + \partial_t u)(X(t), t)
\end{aligned}$$

Discrétisation de l'équation de transport

Cas vitesse non constante, $a = 0$ et $f = 0$

Proposition. Toute solution $u \in \mathscr{C}^1([0, T] \times \mathbb{R})$ de $(\mathcal{T})$, est constante sur les courbes caractéristiques : si X est une caractéristique associée à $(\mathcal{T})$, alors la fonction

$$t \mapsto u(X(t), t)$$

est constante.

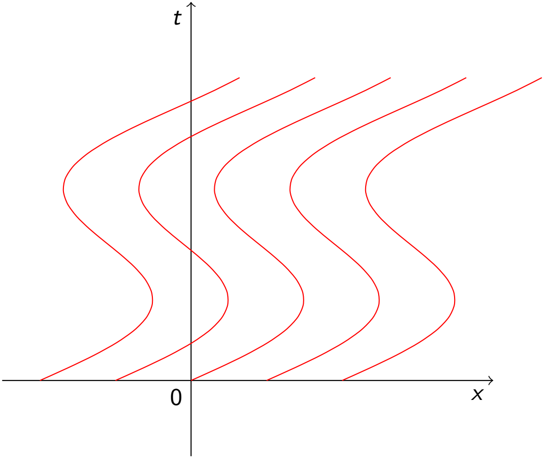
Preuve : On pose $\varphi(t) = u(X(t), t)$. Par dérivation composée,

$$\begin{aligned}
\varphi'(t) &= \partial_x u(X(t), t)X'(t) + \partial_t u(X(t), t) \\
&= \partial_x u(X(t), t)v(X(t), t) + \partial_t u(X(t), t) \\
&= (v\partial_x u + \partial_t u)(X(t), t) \\
&= 0.
\end{aligned}$$

□

Discrétisation de l'équation de transport

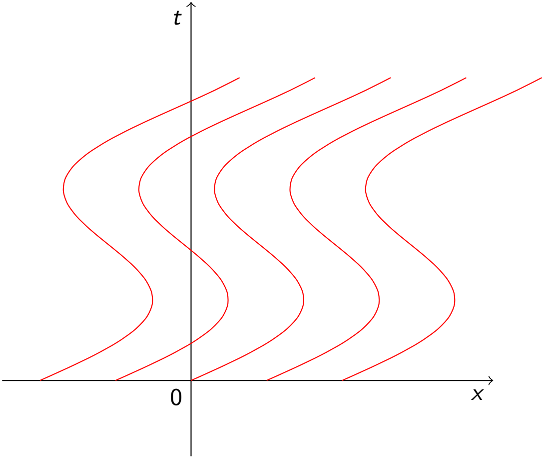
Cas vitesse non constante, $a = 0$ et $f = 0$



Remarques :

Discrétisation de l'équation de transport

Cas vitesse non constante, $a = 0$ et $f = 0$

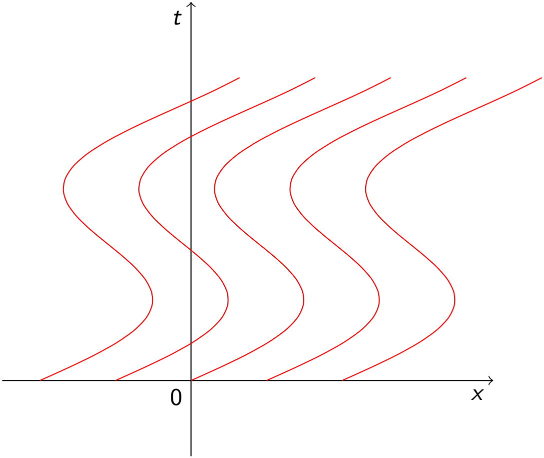


Remarques :

- Les courbes caractéristiques sont les lignes de champ associées au champ de vitesse $v(x, t)$

Discrétisation de l'équation de transport

Cas vitesse non constante, $a = 0$ et $f = 0$

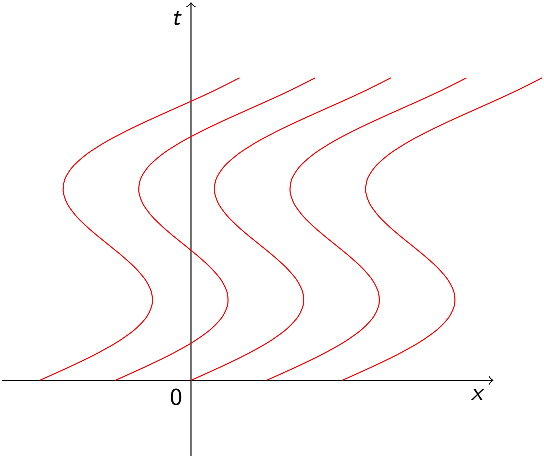


Remarques :

- Les courbes caractéristiques sont les lignes de champ associées au champ de vitesse $v(x, t)$
- Les courbes caractéristiques sont les courbes de niveau de la solution u sur $\mathbb{R} \times [0, T]$.

Discrétisation de l'équation de transport

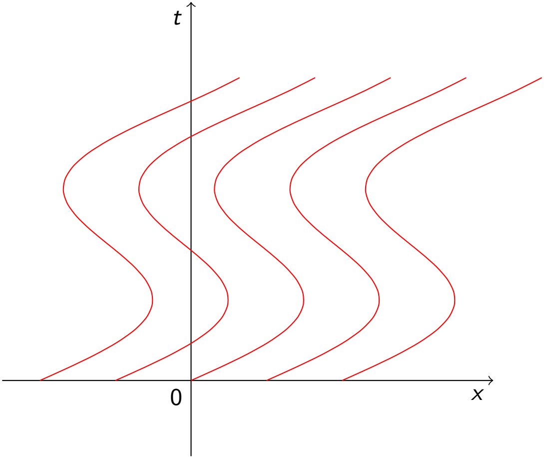
Cas vitesse non constante, $a = 0$ et $f = 0$



Remarques :

- Les courbes caractéristiques sont les lignes de champ associées au champ de vitesse $v(x, t)$
- Les courbes caractéristiques sont les courbes de niveau de la solution u sur $\mathbb{R} \times [0, T]$.
- Les courbes caractéristiques recouvrent le domaine d'espace/temps $\mathbb{R} \times [0, T]$

Cas vitesse non constante, $a = 0$ et $f = 0$



Remarques :

- Les courbes caractéristiques sont les lignes de champ associées au champ de vitesse $v(x, t)$
- Les courbes caractéristiques sont les courbes de niveau de la solution u sur $\mathbb{R} \times [0, T]$.
- Les courbes caractéristiques recouvrent le domaine d'espace/temps $\mathbb{R} \times [0, T]$
- Si on connaît la solution u pour $t = 0$ alors on la connait en tout temps.

Pour résoudre le problème

$$(\mathcal{T}) \qquad \begin{cases} \partial_t u + v\partial_x u = 0, & \text{sur } \mathbb{R} \times [0, T], \\ u(x, 0) = u_0(x), & \forall x \in \mathbb{R}. \end{cases}$$

par la méthode des caractéristiques, on procède comme suit.

Pour résoudre le problème

$$(\mathcal{T}) \qquad \begin{cases} \partial_t u + v\partial_x u = 0, & \text{sur } \mathbb{R} \times [0, T], \\ u(x, 0) = u_0(x), & \forall x \in \mathbb{R}. \end{cases}$$

par la méthode des caractéristiques, on procède comme suit.

- On choisit un point $(x_0, t_0) \in \mathbb{R} \times [0, T]$,

Pour résoudre le problème

$$(\mathcal{T}) \qquad \begin{cases} \partial_t u + v\partial_x u = 0, & \text{sur } \mathbb{R}\times[0,T], \\ u(x,0) = u_0(x), & \forall x \in \mathbb{R}. \end{cases}$$

par la méthode des caractéristiques, on procède comme suit.

- On choisit un point $(x_0, t_0) \in \mathbb{R}\times[0,T]$,
- on identifie l'unique caractéristique X_{x_0,t_0} qui passe par (x_0,t_0), unique solution du problème de Cauchy

$$(\mathcal{C}_{x_0,t_0}) : \begin{cases} X'(t) = v(X(t),t), \\ X(t_0) = x_0. \end{cases}$$

Pour résoudre le problème

$$(\mathcal{T}) \qquad \begin{cases} \partial_t u + v\partial_x u = 0, & \text{sur } \mathbb{R}\times[0,T], \\ u(x,0) = u_0(x), & \forall x \in \mathbb{R}. \end{cases}$$

par la méthode des caractéristiques, on procède comme suit.

- On choisit un point $(x_0, t_0) \in \mathbb{R}\times[0,T]$,
- on identifie l'unique caractéristique X_{x_0,t_0} qui passe par (x_0, t_0), unique solution du problème de Cauchy

$$(\mathcal{C}_{x_0,t_0}) : \begin{cases} X'(t) = v(X(t),t), \\ X(t_0) = x_0. \end{cases}$$

- on évalue u à l'origine de cette caractéristique : $u_0(X_{x_0,t_0}(0))$,

Pour résoudre le problème

$$(\mathcal{T}) \qquad \begin{cases} \partial_t u + v\partial_x u = 0, & \text{sur } \mathbb{R}\times[0,T], \\ u(x,0) = u_0(x), & \forall x \in \mathbb{R}. \end{cases}$$

par la méthode des caractéristiques, on procède comme suit.

- On choisit un point $(x_0, t_0) \in \mathbb{R}\times[0,T]$,
- on identifie l'unique caractéristique X_{x_0,t_0} qui passe par (x_0, t_0), unique solution du problème de Cauchy

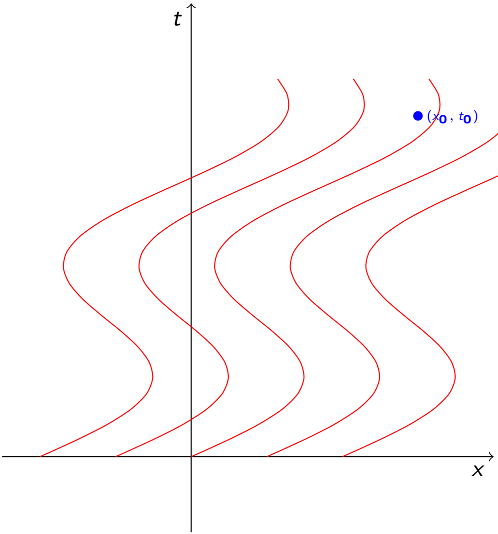
$$(\mathcal{C}_{x_0,t_0}) : \begin{cases} X'(t) = v(X(t), t), \\ X(t_0) = x_0. \end{cases}$$

- on évalue u à l'origine de cette caractéristique : $u_0(X_{x_0,t_0}(0))$,
- on reporte cette valeur au point (x_0, t_0) :

$$u(x_0, t_0) = u_0(X_{x_0,t_0}(0)).$$

Discrétisation de l'équation de transport

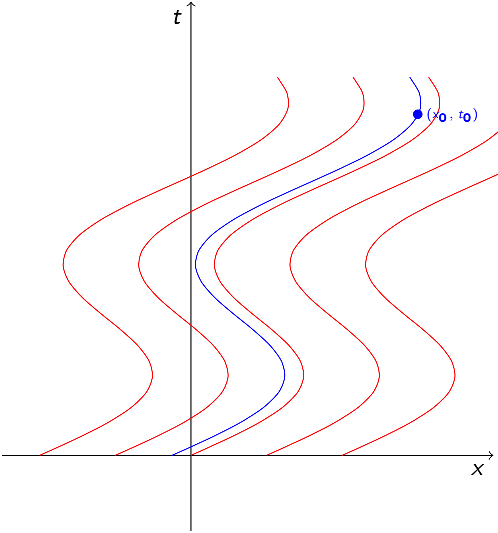
Cas vitesse non constante, $a = 0$ et $f = 0$



$$\boxed{u(x_0, t_0) = u_0(X_{x_0,t_0}(0))}$$

Discrétisation de l'équation de transport

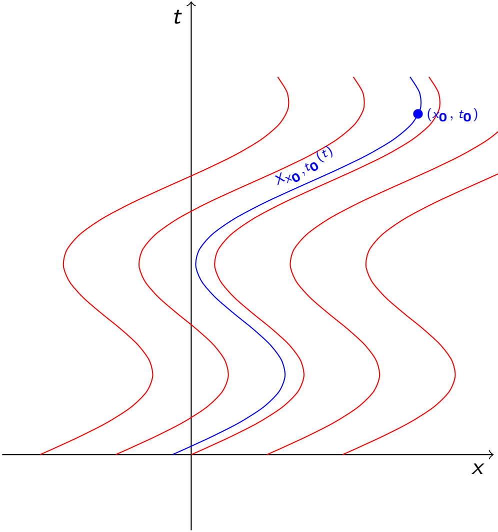
Cas vitesse non constante, $a = 0$ et $f = 0$



$$\boxed{u(x_0, t_0) = u_0(X_{x_0,t_0}(0))}$$

Discrétisation de l'équation de transport

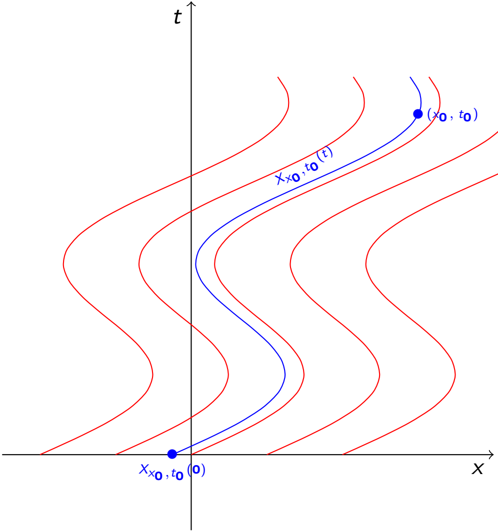
Cas vitesse non constante, $a = 0$ et $f = 0$



$$\boxed{u(x_0, t_0) = u_0(X_{x_0,t_0}(0))}$$

Discrétisation de l'équation de transport

Cas vitesse non constante, $a = 0$ et $f = 0$



$$\boxed{u(x_0, t_0) = u_0(X_{x_0,t_0}(0))}$$

Discrétisation de l'équation de transport

Cas général

$$(\mathcal{T}) \qquad \begin{cases} \partial_t u + v\partial_x u + au = f, & \text{sur } \mathbb{R} \times [0, T], \\ u(x, 0) = u_0(x), & \forall x \in \mathbb{R}. \end{cases}$$

avec v, a, f continue sur $\mathbb{R} \times [0, T]$ et v k-lipschitzienne en espace.

Les caractéristiques sont toujours définies comme les solutions de

$$X'(t) = v(X(t), t).$$

Discrétisation de l'équation de transport

Cas général

$$(\mathcal{T}) \qquad \begin{cases} \partial_t u + v\partial_x u + au = f, & \text{sur } \mathbb{R} \times [0, T], \\ u(x,0) = u_0(x), & \forall x \in \mathbb{R}. \end{cases}$$

avec v, a, f continue sur $\mathbb{R} \times [0, T]$ et v k-lipschitzienne en espace.

Les caractéristiques sont toujours définies comme les solutions de

$$X'(t) = v(X(t), t).$$

Difficulté : Cette fois, la solution u n'est plus constante sur les caractéristiques.

Posons $\varphi(t) = u(X(t), t)$, on a

$$\begin{aligned} \varphi'(t) &= \partial_x u(X(t), t)X'(t) + \partial_t u(X(t), t) \\ &= (v\partial_x u + \partial_t u)(X(t), t) \\ &= f(X(t), t) - a(X(t), t)u(X(t), t) \\ &= f(X(t), t) - a(X(t), t)\varphi(t) \end{aligned}$$

Ainsi φ est solution d'un EDO linéaire et n'est pas constante en général.

Proposition. Soient v continue et k-lipschitzienne en espace, a, f continues et u_0 de classe $\mathscr{C}^1$. Le problème de transport

$$(\mathcal{T}) \qquad \begin{cases} \partial_t u + v\partial_x u + au = f, & \text{sur } \mathbb{R} \times [0, T], \\ u(x, 0) = u_0(x), & \forall x \in \mathbb{R}. \end{cases}$$

admet une unique solution $u \in \mathscr{C}^1(\mathbb{R} \times [0, T])$.

Proposition. Soient v continue et k-lipschitzienne en espace, a, f continues et u_0 de classe $\mathscr{C}^1$. Le problème de transport

$$(\mathcal{T}) \qquad \begin{cases} \partial_t u + v\partial_x u + au = f, & \text{sur } \mathbb{R}\times[0,T], \\ u(x,0) = u_0(x), & \forall x \in \mathbb{R}. \end{cases}$$

admet une unique solution $u \in \mathscr{C}^1(\mathbb{R}\times[0,T])$.

Preuve d'unicité (constructive) : Supposons qu'il existe une solution. Soit $(x_0,t_0) \in \mathbb{R}\times[0,T]$, on va calculer $u(x_0,t_0)$. Considérons la caractéristique X_{x_0,t_0}, elle vérifie $X_{x_0,t_0}(t_0) = x_0$. Posons maintenant $\varphi_{x_0,t_0}(t) = u\big(X_{x_0,t_0}(t),t\big)$. D'après le calcul précédent, c'est l'unique solution du problème de Cauchy

$$\begin{cases} \varphi_{x_0,t_0}(t) = f\big(X_{x_0,t_0}(t),t\big) - a\big(X_{x_0,t_0}(t),t\big)\varphi_{x_0,t_0}(t) \\ \varphi_{x_0,t_0}(0) = u_0\big(X_{x_0,t_0}(0)\big). \end{cases}$$

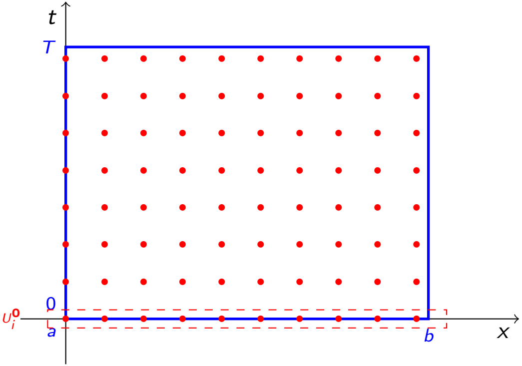
Cette EDO est linéaire donc φ_{x_0,t_0} existe et est unique. Elle est entièrement déterminée par f, a, u_0 et X_{x_0,t_0} et on a $u(x_0,t_0) = \varphi_{x_0,t_0}(t_0)$.

Méthode des différences finies

Problème : La méthode des caractéristiques est très coûteuse s'il on veut connaitre la solution u sur un domaine large. Il faut résoudre un très grand nombre d'EDO.

Autre méthode : La méthode des différences finies consiste a calculer directement des approximations notés U_i^n de u sur les points de la grille (x_i, t_n) par récurrence sur n.

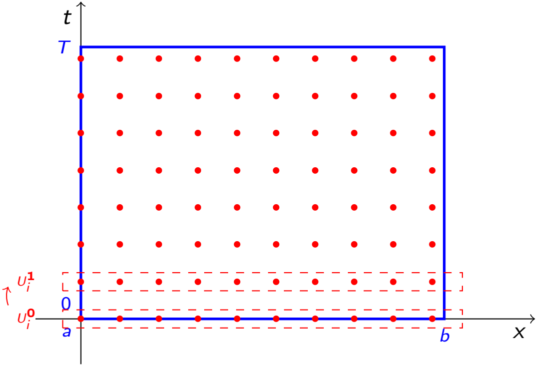
- Les valeurs U_i^0 sont définies par la condition initiale : $U_i^0 = u_0(x_i)$
- Les valeurs U_i^1 sont calculées à partir des valeurs U_i^0
- Les valeurs U_i^2 sont calculées à partir des valeurs U_i^1
- ...



Problème : La méthode des caractéristiques est très coûteuse s'il on veut connaitre la solution u sur un domaine large. Il faut résoudre un très grand nombre d'EDO.

Autre méthode : La méthode des différences finies consiste a calculer directement des approximations notés U_i^n de u sur les points de la grille (x_i, t_n) par récurrence sur n.

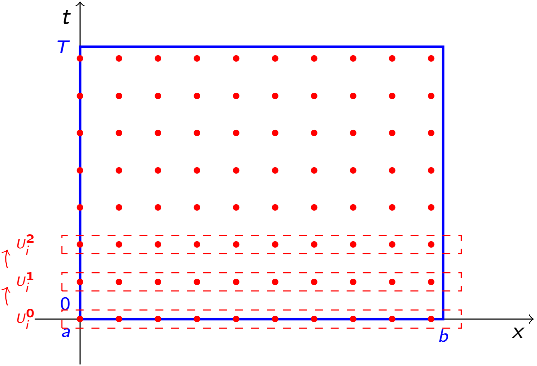
- Les valeurs U_i^0 sont définies par la condition initiale : $U_i^0 = u_0(x_i)$
- Les valeurs U_i^1 sont calculées à partir des valeurs U_i^0
- Les valeurs U_i^2 sont calculées à partir des valeurs U_i^1
- ...



Problème : La méthode des caractéristiques est très coûteuse s'il on veut connaitre la solution u sur un domaine large. Il faut résoudre un très grand nombre d'EDO.

Autre méthode : La méthode des différences finies consiste a calculer directement des approximations notés U_i^n de u sur les points de la grille (x_i, t_n) par récurrence sur n.

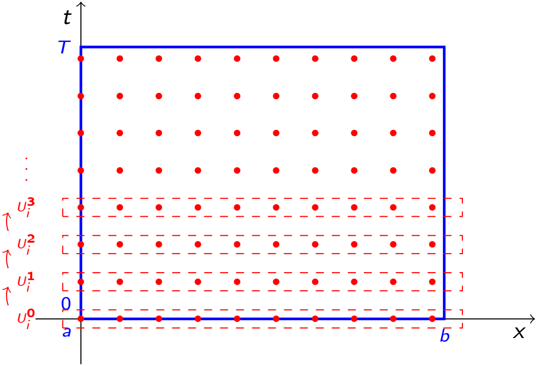
- Les valeurs U_i^0 sont définies par la condition initiale : $U_i^0 = u_0(x_i)$
- Les valeurs U_i^1 sont calculées à partir des valeurs U_i^0
- Les valeurs U_i^2 sont calculées à partir des valeurs U_i^1
- ...



Problème : La méthode des caractéristiques est très coûteuse s'il on veut connaitre la solution u sur un domaine large. Il faut résoudre un très grand nombre d'EDO.

Autre méthode : La méthode des différences finies consiste a calculer directement des approximations notés U_i^n de u sur les points de la grille (x_i, t_n) par récurrence sur n.

- Les valeurs U_i^0 sont définies par la condition initiale : $U_i^0 = u_0(x_i)$
- Les valeurs U_i^1 sont calculées à partir des valeurs U_i^0
- Les valeurs U_i^2 sont calculées à partir des valeurs U_i^1
- ...



On fixe $\Delta x > 0$ et $\Delta t > 0$ et on discrétise le rectangle d'espace-temps par

$$x_i = a + (i-1)\Delta x, \quad 1 \leq i \leq N_x, \qquad t_n = n\Delta t, \quad 0 \leq n \leq N_t.$$

avec $N_x = \frac{(b-a)}{\Delta x} + 1$ et $N_t = \frac{T}{\Delta t}$.

On fixe $\Delta x > 0$ et $\Delta t > 0$ et on discrétise le rectangle d'espace-temps par

$$x_i = a + (i-1)\Delta x, \quad 1 \leq i \leq N_x, \qquad t_n = n\Delta t, \quad 0 \leq n \leq N_t.$$

avec $N_x = \frac{(b-a)}{\Delta x} + 1$ et $N_t = \frac{T}{\Delta t}$.

Définition. Un schéma numérique est une relation (homogène à $\partial_t u$) portant sur les coefficients de (U_i^n) de la forme

$$G_{\Delta x, \Delta t}\left((U^{n+1})_{i=1}^{N_x}, (U_i^n)_{i=1}^{N_x}\right) = 0, \quad 0 \leq n \leq N_t,$$

permettant de calculer ces coefficients par récurrence sur les points de la grille (t_n, x_i) à partir de la donnée initiale

$$U_i^0 = u_0(x_i), \quad 1 \leq i \leq N_x.$$

Discrétisation de l'équation de transport
Schéma numérique, exemple

Afin de discrétiser le problème

$$(\mathcal{T}) \qquad \begin{cases} \partial_t u + c\partial_x u = 0, & \text{sur } \mathbb{R} \times [0, T], \\ u(x,0) = u_0(x), & \forall x \in \mathbb{R}. \end{cases}$$

un exemple simple est le schéma explicite décentré à droite (ED) donné par

$$\text{(ED)} \qquad \frac{U_i^{n+1} - U_i^n}{\Delta t} + c\,\frac{U_{i+1}^n - U_i^n}{\Delta x} = 0.$$

On peut le réécrire

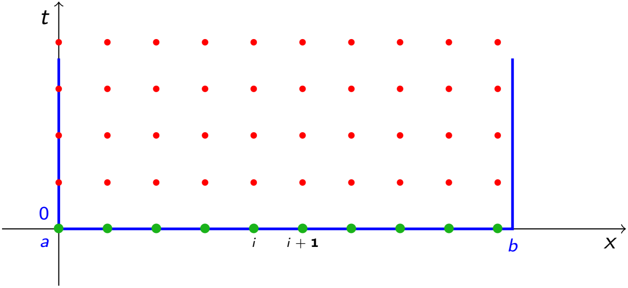
$$U_i^{n+1} = U_i^n + \beta(U_i^n - U_{i+1}^n),$$

avec $\beta = \frac{c\Delta t}{\Delta x}$.

Grâce à la formule

$$U_i^{n+1} = U_i^n + \beta(U_i^n - U_{i+1}^n),$$

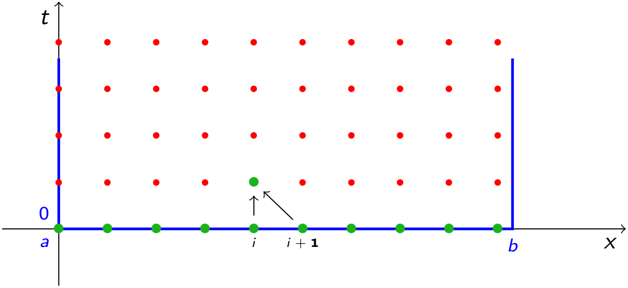
on peut calculer une partie des coefficients



Grâce à la formule

$$U_i^{n+1} = U_i^n + \beta(U_i^n - U_{i+1}^n),$$

on peut calculer une partie des coefficients



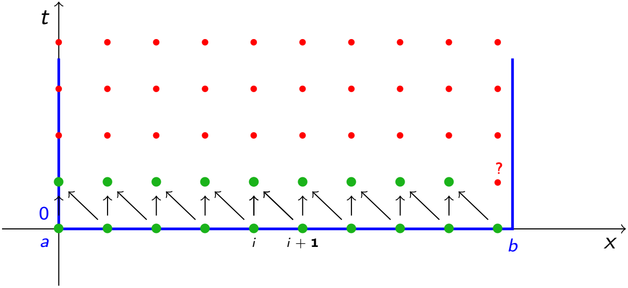
Discrétisation de l'équation de transport

Schéma numérique, exemple

Grâce à la formule

$$U_i^{n+1} = U_i^n + \beta(U_i^n - U_{i+1}^n),$$

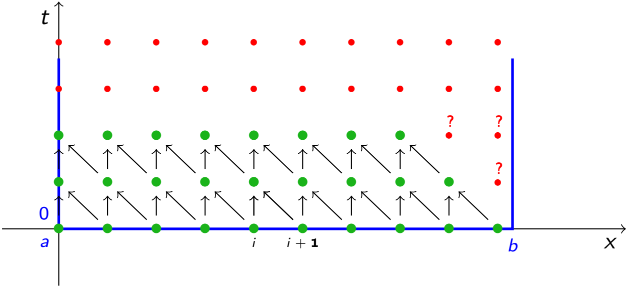
on peut calculer une partie des coefficients



Grâce à la formule

$$U_i^{n+1} = U_i^n + \beta(U_i^n - U_{i+1}^n),$$

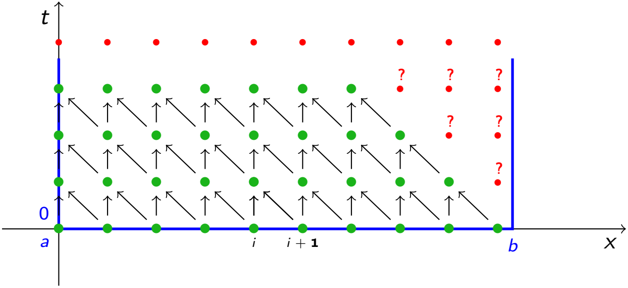
on peut calculer une partie des coefficients



Grâce à la formule

$$U_i^{n+1} = U_i^n + \beta(U_i^n - U_{i+1}^n),$$

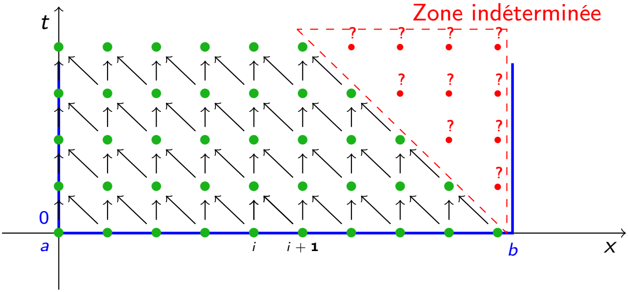
on peut calculer une partie des coefficients



Grâce à la formule

$$U_i^{n+1} = U_i^n + \beta(U_i^n - U_{i+1}^n),$$

on peut calculer une partie des coefficients



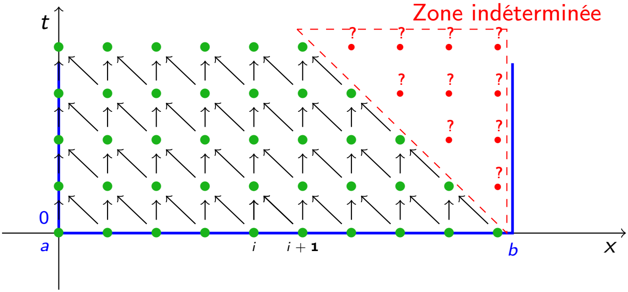
Discrétisation de l'équation de transport

Schéma numérique, exemple

Grâce à la formule

$$U_i^{n+1} = U_i^n + \beta(U_i^n - U_{i+1}^n),$$

on peut calculer une partie des coefficients



Pour gérer la zone indéterminée, on ajoute un condition sur le bord : $U_{N_x}^n = 0$.

Schéma numérique : écriture matricielle

Il faut résoudre N_x équations pour passer des lignes n à $n+1$:

$$\begin{cases} U_i^{n+1} = U_i^n + \beta(U_i^n - U_{i+1}^n), & \forall i \in \{1, \dots, N_x - 1\}, \\ U_{N_x}^{n+1} = U_{N_x}^n + \beta U_{N_x}^n, \end{cases}$$

on calcule matriciellement tout les $\left(U_i^{n+1}\right)_{i=1}^{N_x}$ d'un coup à chaque itération.

Il faut résoudre N_x équations pour passer des lignes n à $n+1$:

$$\begin{cases} U_i^{n+1} = U_i^n + \beta(U_i^n - U_{i+1}^n), & \forall i \in \{1, \ldots, N_x - 1\}, \\ U_{N_x}^{n+1} = U_{N_x}^n + \beta U_{N_x}^n, \end{cases}$$

on calcule matriciellement tout les $\left(U_i^{n+1}\right)_{i=1}^{N_x}$ d'un coup à chaque itération.Posons $\mathrm{U}^n = \left(U_i^{n+1}\right)_{i=1}^{N_x}$ et définissons

$$\mathcal{D} := \begin{bmatrix} 1 & -1 & & \\ & \ddots & \ddots & \\ & & \ddots & -1 \\ & & & 1 \end{bmatrix} \in \mathcal{M}_{N_x}(\mathbb{R}),$$

Schéma numérique : écriture matricielle

Il faut résoudre N_x équations pour passer des lignes n à $n+1$:

$$\begin{cases} U_i^{n+1} = U_i^n + \beta(U_i^n - U_{i+1}^n), \qquad \forall i \in \{1, \dots, N_x - 1\}, \\ U_{N_x}^{n+1} = U_{N_x}^n + \beta U_{N_x}^n, \end{cases}$$

on calcule matriciellement tout les $\left(U_i^{n+1}\right)_{i=1}^{N_x}$ d'un coup à chaque itération.Posons $\mathrm{U}^n = \left(U_i^{n+1}\right)_{i=1}^{N_x}$ et définissons

$$\mathcal{D} := \begin{bmatrix} 1 & -1 & & & \\ & \ddots & \ddots & & \\ & & & & \\ & & \ddots & -1 & \\ & & & 1 \end{bmatrix} \in \mathcal{M}_{N_x}(\mathbb{R}),$$

on obtient la suite de vecteurs par recurrence :

$$\begin{cases} \mathrm{U}^{n+1} = (\mathcal{I} + \beta\mathcal{D})\mathrm{U}^n, \\ \mathrm{U}^0 = \left(u_0(x_i)\right)_{i=1}^{N_x}. \end{cases}$$

Remarque la matrice $\mathcal{A} := \mathcal{I} + \beta\mathcal{D}$ est appelé matrice d'itération du schéma.

Pour mesurer la performance d'un schéma numérique pour résoudre une EDP donnée on introduit les même notions que pour l'approximation des EDO :

Pour mesurer la performance d'un schéma numérique pour résoudre une EDP donnée on introduit les même notions que pour l'approximation des EDO :

- L'erreur de consistance est l'erreur commise par le schéma à chaque pas d'espace et de temps par rapport à la solution exacte.

Pour mesurer la performance d'un schéma numérique pour résoudre une EDP donnée on introduit les même notions que pour l'approximation des EDO :

- L'erreur de consistance est l'erreur commise par le schéma à chaque pas d'espace et de temps par rapport à la solution exacte.
- La stabilité contrôle la propagation des erreurs quand n varie.

Pour mesurer la performance d'un schéma numérique pour résoudre une EDP donnée on introduit les même notions que pour l'approximation des EDO :

- L'erreur de consistance est l'erreur commise par le schéma à chaque pas d'espace et de temps par rapport à la solution exacte.
- La stabilité contrôle la propagation des erreurs quand n varie.
- La convergence est encore la conséquence de la consistance et de la stabilité. Il s'agit de la convergence vers zero de l'erreur entre la solution du schéma et les valeurs exactes : $|U_i^n - u(t_n, x_i)| \to 0$ uniformément en i et en n.

Discrétisation de l'équation de transport

Schéma numérique

On note u la solution exacte de l'EDP et $G_{\Delta x, \Delta t}$ un schéma numérique.

Définition. L'erreur de consistance est donnée par

$$\varepsilon_i^n = G_{\Delta x, \Delta t}\big(u(x_i, t_{n+1})_{i=1}^{N_x}, u(x_i, t_n)_{i=1}^{N_x}\big).$$

ÉCOLE CENTRALE LYON

On note u la solution exacte de l'EDP et $G_{\Delta x, \Delta t}$ un schéma numérique.

Définition. L'erreur de consistance est donnée par

$$\varepsilon_i^n = G_{\Delta x, \Delta t}\left(u(x_i, t_{n+1})_{i=1}^{N_x}, u(x_i, t_n)_{i=1}^{N_x}\right).$$

Exemple : Pour le schéma (ED), l'erreur de consistance est

$$\varepsilon_i^n = \frac{u(x_i, t_{n+1}) - u(x_i, t_n)}{\Delta t} + c\,\frac{u(x_{i+1}, t_n) - u(x_i, t_n)}{\Delta x}.$$

Discrétisation de l'équation de transport

Schéma numérique

On note u la solution exacte de l'EDP et $G_{\Delta x,\Delta t}$ un schéma numérique.

Définition. L'erreur de consistance est donnée par

$$\varepsilon_i^n = G_{\Delta x,\Delta t}\left(u(x_i,t_{n+1})_{i=1}^{N_x}, u(x_i,t_n)_{i=1}^{N_x}\right).$$

Exemple : Pour le schéma (ED), l'erreur de consistance est

$$\varepsilon_i^n = \frac{u(x_i,t_{n+1}) - u(x_i,t_n)}{\Delta t} + c\,\frac{u(x_{i+1},t_n) - u(x_i,t_n)}{\Delta x}.$$

Définition. Le schéma est consistant si

$$\max_{\substack{0\leq i\leq N_x\\ 0\leq n\leq N_t}} |\varepsilon_i^n| \longrightarrow 0 \quad \text{quand} \quad (\Delta x,\Delta t)\to 0.$$

Discrétisation de l'équation de transport

Schéma numérique

On note u la solution exacte de l'EDP et $G_{\Delta x,\Delta t}$ un schéma numérique.

Définition. L'erreur de consistance est donnée par

$$\varepsilon_i^n = G_{\Delta x,\Delta t}\left(u(x_i, t_{n+1})_{i=1}^{N_x}, u(x_i, t_n)_{i=1}^{N_x}\right).$$

Exemple : Pour le schéma (ED), l'erreur de consistance est

$$\varepsilon_i^n = \frac{u(x_i, t_{n+1}) - u(x_i, t_n)}{\Delta t} + c\,\frac{u(x_{i+1}, t_n) - u(x_i, t_n)}{\Delta x}.$$

Définition. Le schéma est consistant si

$$\max_{\substack{0\leq i\leq N_x\\ 0\leq n\leq N_t}} |\varepsilon_i^n| \longrightarrow 0 \quad \text{quand} \quad (\Delta x, \Delta t) \to 0.$$

le schéma est constant d'ordre p en temps et q en espace si

$$\max_{\substack{0\leq i\leq N_x\\ 0\leq n\leq N_t}} |\varepsilon_i^n| = \mathcal{O}\left(\Delta t^p + \Delta x^q\right).$$

Discrétisation de l'équation de transport

Consistance du schéma (ED)

Exemple : Le schéma (ED) est constant d'ordre 1 en espace et 1 en temps.

Exemple : Le schéma (ED) est constant d'ordre 1 en espace et 1 en temps.

Preuve : On développe l'erreur de consistance

$$\varepsilon_i^n = \frac{u(x_i, t_{n+1}) - u(x_i, t_n)}{\Delta t} + c\,\frac{u(x_{i+1}, t_n) - u(x_i, t_n)}{\Delta x}.$$

par des développement de Taylor,

$$\begin{aligned} u(x_i, t_{n+1}) &= u(x_i, t_n) + \Delta t \partial_t u(x_i, t_n) + \mathcal{O}(\Delta t^2), \\ u(x_{i+1}, t_n) &= u(x_i, t_n) + \Delta x \partial_x u(x_i, t_n) + \mathcal{O}(\Delta x^2), \end{aligned}$$

Ainsi

$$\begin{aligned} \varepsilon_i^n &= \partial_t u(x_i, t_n) + \mathcal{O}(\Delta t) + c\partial_x u(x_i, t_n) + \mathcal{O}(\Delta x), \\ &= \mathcal{O}(\Delta t + \Delta x). \end{aligned}$$

□

Discrétisation de l'équation de transport

Stabilité 1

Pour un schéma quelconque écrit sous la forme matricielle

$$(\mathscr{S}) \qquad \begin{cases} \mathsf{U}^{n+1} = \mathcal{A}\mathsf{U}^n, \\ \mathsf{U}^0 \in \mathbb{R}^{N_x}, \end{cases}$$

Pour un schéma quelconque écrit sous la forme matricielle

$$(\mathscr{S}) \qquad \begin{cases} \mathrm{U}^{n+1} = \mathcal{A}\mathrm{U}^n, \\ \mathrm{U}^0 \in \mathbb{R}^{N_x}, \end{cases}$$

On perturbe le schéma par une suite d'erreurs $\boldsymbol{\mu}^n \in \mathbb{R}^{N_x}$:

$$(\mathscr{S}_{\mathrm{pert}}) \qquad \begin{cases} \mathrm{V}^{n+1} = \mathcal{A}\mathrm{V}^n + \boldsymbol{\mu}^n, \\ \mathrm{V}^0 = \mathrm{U}^0. \end{cases}$$

Pour un schéma quelconque écrit sous la forme matricielle

$$(\mathscr{S}) \qquad \begin{cases} \mathrm{U}^{n+1} = \mathcal{A}\mathrm{U}^n, \\ \mathrm{U}^0 \in \mathbb{R}^{N_x}, \end{cases}$$

On perturbe le schéma par une suite d'erreurs $\boldsymbol{\mu}^n \in \mathbb{R}^{N_x}$:

$$(\mathscr{S}_{\text{pert}}) \qquad \begin{cases} \mathrm{V}^{n+1} = \mathcal{A}\mathrm{V}^n + \boldsymbol{\mu}^n, \\ \mathrm{V}^0 = \mathrm{U}^0. \end{cases}$$

Définition. Le schéma $(\mathscr{S})$ est stable pour la norme ∞ s'il existe $C > 0$ tel que pour toute perturbation $\boldsymbol{\mu}^n$ on a

$$\max_{0 \leq n \leq N_t} \|\mathrm{V}^n - \mathrm{U}^n\|_\infty \leq C \sum_{n=0}^{N_t - 1} \|\boldsymbol{\mu}^n\|_\infty .$$

Discrétisation de l'équation de transport

Stabilité 2

proposition. Si la matrice d'itération du $\mathcal{A}$ du schéma vérifie $|||\mathcal{A}|||_\infty \leq 1$, alors le schéma est stable en norme ∞.

Preuve :

proposition. Si la matrice d'itération du $\mathcal{A}$ du schéma vérifie $|||\mathcal{A}|||_\infty \leq 1$, alors le schéma est stable en norme ∞.

Preuve : On écrit simplement par différence que

$$\mathrm{V}^{n+1} - \mathrm{U}^{n+1} = \mathcal{A}(\mathrm{V}^n - \mathrm{U}^n) + \boldsymbol{\mu}^n.$$

Ainsi,

$$\begin{aligned}\left\|\mathrm{V}^{n+1} - \mathrm{U}^{n+1}\right\|_\infty &\leq \left\|\mathcal{A}(\mathrm{V}^n - \mathrm{U}^n)\right\|_\infty + \left\|\boldsymbol{\mu}^n\right\|_\infty, \\ &\leq |||\mathcal{A}|||_\infty \left\|\mathrm{V}^n - \mathrm{U}^n\right\|_\infty + \left\|\boldsymbol{\mu}^n\right\|_\infty, \\ &\leq \left\|\mathrm{V}^n - \mathrm{U}^n\right\|_\infty + \left\|\boldsymbol{\mu}^n\right\|_\infty.\end{aligned}$$

En sommant, il vient que

$$\left\|\mathrm{V}^n - \mathrm{U}^n\right\|_\infty \leq \sum_{n=0}^{N_t-1} \left\|\boldsymbol{\mu}^n\right\|_\infty, \quad \forall n \leq N_t.$$

$\square$

Le schéma (ED) s'ecrit $\mathrm{U}^{n+1} = \mathcal{A}\mathrm{U}^n$ avec $\mathcal{A} = \mathcal{I} + \beta\mathcal{D}$.

Question : Peut-on contrôler $|||\mathcal{A}|||_\infty$?

Le schéma (ED) s'ecrit $\mathrm{U}^{n+1} = \mathcal{A}\mathrm{U}^n$ avec $\mathcal{A} = \mathcal{I} + \beta\mathcal{D}$.

Question : Peut-on contrôler $|||\mathcal{A}|||_\infty$?

$$\mathcal{A} := \begin{bmatrix} 1+\beta & -\beta & & \\ & \ddots & \ddots & \\ & & \ddots & -\beta \\ & & & 1+\beta \end{bmatrix}.$$

Ainsi $|||\mathcal{A}||| = \max_i \sum_j |\mathcal{A}_{ij}| = |1+\beta| + |\beta|$.

On vérifie alors aisément que $|1+\beta| + |\beta| \leq 1$ si et seulement si $\beta \in [-1, 0]$.

Le schéma (ED) s'ecrit $\mathrm{U}^{n+1} = \mathcal{A}\mathrm{U}^n$ avec $\mathcal{A} = \mathcal{I} + \beta\mathcal{D}$.

Question : Peut-on contrôler $|||\mathcal{A}|||_\infty$?

$$\mathcal{A} := \begin{bmatrix} 1+\beta & -\beta & & \\ & \ddots & \ddots & \\ & & \ddots & -\beta \\ & & & 1+\beta \end{bmatrix}.$$

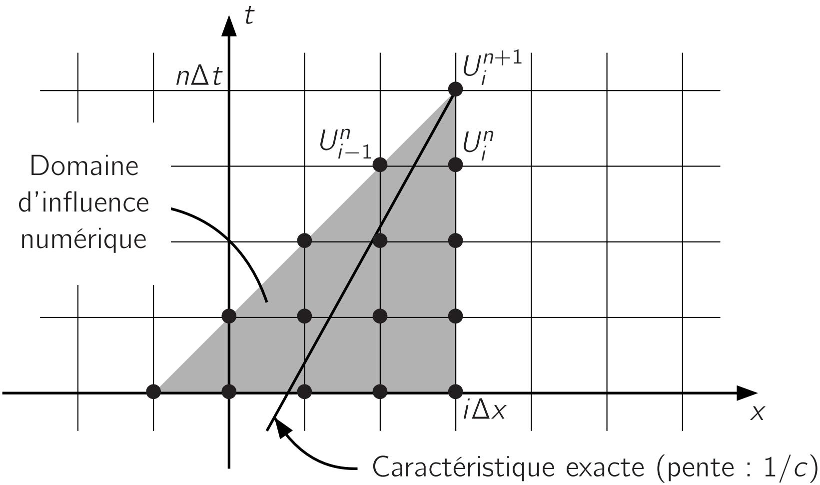
Ainsi $|||\mathcal{A}||| = \max_i \sum_j |\mathcal{A}_{ij}| = |1+\beta| + |\beta|$.

On vérifie alors aisément que $|1+\beta| + |\beta| \leq 1$ si et seulement si $\beta \in [-1, 0]$.

Proposition. Le schéma (ED) est stable si

$$-1 \leq \frac{c\Delta t}{\Delta x} \leq 0$$

Cette condition s'appelle condition CFL (Courant–Friedrichs–Lewy).


t
$n\Delta t$
U_i^{n+1}
U_{i-1}^n
U_i^n
Domaine d'influence numérique
$i\Delta x$
x
Caractéristique exacte (pente : $1/c$)

Définition. Le schéma $(\mathscr{S})$ est convergent pour la norme ∞ si

$$\max_{0 \leq n \leq N_t} \left\| \left(\mathrm{u}(x_i, t_n)\right)_{i=1}^{N_x} - \mathrm{U}^n \right\|_\infty \to 0 \quad \text{quand} \quad (\Delta t, \Delta x) \to 0.$$

Si de plus on a

$$\max_{0 \leq n \leq N_t} \left\| \left(\mathrm{u}(x_i, t_n)\right)_{i=1}^{N_x} - \mathrm{U}^n \right\|_\infty = \mathcal{O}(\Delta t^p + \Delta x^q).$$

le schéma est convergent d'ordre p en temps et q en espace pour la norme ∞.

Définition. Le schéma $(\mathscr{S})$ est convergent pour la norme ∞ si

$$\max_{0\leq n\leq N_t} \left\| \left(\mathrm{u}(x_i, t_n)\right)_{i=1}^{N_x} - \mathrm{U}^n \right\|_\infty \to 0 \quad \text{quand} \quad (\Delta t, \Delta x) \to 0.$$

Si de plus on a

$$\max_{0\leq n\leq N_t} \left\| \left(\mathrm{u}(x_i, t_n)\right)_{i=1}^{N_x} - \mathrm{U}^n \right\|_\infty = \mathcal{O}(\Delta t^p + \Delta x^q).$$

le schéma est convergent d'ordre p en temps et q en espace pour la norme ∞.

Théorème. (Lax) Consistance et Stabilité $\Rightarrow$ convergence.

Discrétisation de l'équation de transport

Preuve de Lax pour (ED)

Preuve. pour le schéma

$$(ED) : \quad \frac{U_i^{n+1} - U_i^n}{\Delta t} + c\,\frac{U_{i+1}^n - U_i^n}{\Delta x} = 0,$$

d'écriture matricielle $\mathrm{U}^{n+1} = \mathcal{A}\mathrm{U}^n$ l'erreur de consistance est

$$\varepsilon_i^n = \frac{u(x_i, t_{n+1}) - u(x_i, t_n)}{\Delta t} + c\,\frac{u(x_{i+1}, t_n) - u(x_i, t_n)}{\Delta x}.$$

On pose $V_i^n = u(t_n, x_i)$ qui vérifie le schéma perturbé

$$(ED_{\mathrm{pert}}) : \quad \frac{V_i^{n+1} - V_i^n}{\Delta t} + c\,\frac{V_{i+1}^n - V_i^n}{\Delta x} = \varepsilon_i^n,$$

d'écriture matricielle $\mathrm{V}^{n+1} = \mathcal{A}\mathrm{V}^n + \Delta t \boldsymbol{\varepsilon}^n$ avec $\boldsymbol{\varepsilon}^n = (\varepsilon_i^n)_{i=1}^{N_x}$. On utilise la stabilité du schéma :

$$\begin{aligned}\max_{0 \le n \le N_t} \|\mathrm{V}^n - \mathrm{U}^n\|_\infty &\le C\Delta t \sum_{n=0}^{N_t - 1} \|\boldsymbol{\varepsilon}^n\|_\infty \le C\Delta t N_t \max_{0 \le n \le N_t} \|\boldsymbol{\varepsilon}^n\|_\infty , \\ &\le CT \max_{0 \le n \le N_t} \|\boldsymbol{\varepsilon}^n\|_\infty .\end{aligned}$$

On conclut en utilisant la consistance du schéma. $\square$

D'autres schémas

Le schéma décentré à droite requiert la CFL $c\frac{\Delta t}{\Delta x} \in [-1, 0]$ (qui impose $c < 0$).

Le schéma décentré à droite requiert la CFL $c\frac{\Delta t}{\Delta x} \in [-1, 0]$ (qui impose $c < 0$).

- Le schéma décentré à gauche :

$$\frac{U_i^{n+1} - U_i^n}{\Delta t} + c\,\frac{U_i^n - U_{i-1}^n}{\Delta x} = 0.$$

est adapté au cas $c > 0$.

$\Longrightarrow$ décentrement « amont ».

Le schéma décentré à droite requiert la CFL $c\frac{\Delta t}{\Delta x} \in [-1, 0]$ (qui impose $c < 0$).

- Le schéma décentré à gauche :

$$\frac{U_i^{n+1} - U_i^n}{\Delta t} + c\frac{U_i^n - U_{i-1}^n}{\Delta x} = 0.$$

est adapté au cas $c > 0$.

$\Longrightarrow$ décentrement « amont ».

- Que dire du schéma centré ?

$$\frac{U_i^{n+1} - U_i^n}{\Delta t} + c\frac{U_{i+1}^n - U_{i-1}^n}{2\Delta x} = 0.$$