Calcul Scientifique Parallèle

Cours AMS301 — Automne 2021 — Cours 5

Résolution de systèmes linéaires : Méthodes itératives (2)

Méthode du gradient conjugué & Introduction aux méthodes de Krylov

Axel Modave

Résolution de systèmes linéaires — Rappels [1/3]

On cherche $\mathbf{x} \in \mathbb{R}^n$ tel que $\boxed{\mathbf{Ax} = \mathbf{b}}$ avec $\mathbf{A} \in \mathbb{R}^{n\times n}$ et $\mathbf{b} \in \mathbb{R}^n$.

Procédures de résolution

- **Méthodes directes :** Factorisation de $\mathbf{A}$ en matrices triangulaires et/ou diagonales (ex. $\mathbf{A} = \mathbf{LU}$) et résolution de problèmes simples.

$$\mathbf{Ax} = \mathbf{b} \quad \Leftrightarrow \quad \mathbf{LUx} = \mathbf{b} \quad \Leftrightarrow \quad \left| \begin{array}{l} \mathbf{Ly} = \mathbf{b} \\ \mathbf{Ux} = \mathbf{y} \end{array} \right.$$

  Avantages : solution exacte obtenue après un nombre fini d'opérations
  Difficultés : coût important *(calcul/mémoire)*, parallélisation difficile

- **Méthodes itératives :** Procédure itérative permettant de minimiser une norme de l'erreur $\|\mathbf{x}^{(k)} - \mathbf{x}_{\text{ref}}\|$ et/ou d'une norme du résidu $\|\mathbf{Ax}^{(k)} - \mathbf{b}\|$.

$$\left| \begin{array}{l} \mathbf{x}^{(0)} = \mathrm{Iter}^{(0)}(\mathbf{A},\mathbf{b}) \\ \mathbf{x}^{(k+1)} = \mathrm{Iter}^{(k+1)}\left(\mathbf{x}^{(k)}, \mathbf{x}^{(k-1)}, \ldots, \mathbf{A}, \mathbf{b}\right), \quad \text{pour } k \geq 0 \end{array} \right.$$

  Avantages : coût limité par itération *(calcul/mémoire)*, parallélisation facile/efficace
  Difficultés : solution approchée, contrôle de la convergence du processus

Résolution de systèmes linéaires — Rappels [2/3]

On cherche $\mathbf{x} \in \mathbb{R}^n$ tel que $\boxed{\mathbf{Ax} = \mathbf{b}}$ avec $\mathbf{A} \in \mathbb{R}^{n\times n}$ et $\mathbf{b} \in \mathbb{R}^n$.

- **Méthodes itératives :** Procédure itérative permettant de minimiser une norme de l'erreur $\|\mathbf{x}^{(k)} - \mathbf{x}_{\text{ref}}\|$ et/ou d'une norme du résidu $\|\mathbf{Ax}^{(k)} - \mathbf{b}\|$.

$$\left| \begin{array}{l} \mathbf{x}^{(0)} = \mathrm{Iter}^{(0)}(\mathbf{A},\mathbf{b}) \\ \mathbf{x}^{(k+1)} = \mathrm{Iter}^{(k+1)}\left(\mathbf{x}^{(k)}, \mathbf{x}^{(k-1)}, \ldots, \mathbf{A}, \mathbf{b}\right), \quad \text{pour } k \geq 0 \end{array} \right.$$

  Avantages : coût limité par itération *(calcul/mémoire)*, parallélisation facile/efficace
  Difficultés : solution approchée, contrôle de la convergence du processus

À la dernière séance, nous avons considéré des *schémas stationnaires* de la forme :

$$\left| \begin{array}{l} \mathbf{x}^{(0)} \text{ donné} \\ \mathbf{x}^{(k+1)} = \mathbf{Bx}^{(k)} + \mathbf{f}, \quad k \geq 0 \end{array} \right.$$

où $\mathbf{B} \in \mathbb{R}^{n\times n}$ est la matrice d'itération et $\mathbf{f} \in \mathbb{R}^n$.

Résolution de systèmes linéaires — Rappels [3/3]

Méthodes stationnaires linéaires du 1^er ordre

Décomposition régulière : $\boxed{\mathbf{A} = \mathbf{M} - \mathbf{N}}$ où $\mathbf{M} \in \mathbb{R}^{n\times n}$ est inversible.

Méthode de relaxation

$\mathbf{x}^{(0)} \in \mathbb{C}^n$
for $k = 0, 1, \ldots$ **do**
  $\mathbf{Mx}^{(k+1)} = \mathbf{Nx}^{(k)} + \mathbf{b}$
end

C'est-à-dire $\boxed{\mathbf{x}^{(k+1)} = \mathbf{Bx}^{(k)} + \mathbf{f}}$
avec $\mathbf{B} = \mathbf{M}^{-1}\mathbf{N}$ et $\mathbf{f} = \mathbf{M}^{-1}\mathbf{b}$

Choix pour les matrices :

	Par points	Par blocs
Jacobi	$\mathbf{M} = \mathbf{D}$	$\mathbf{M} = \mathbf{D}^{\text{blk}}$
Gauss-Seidel	$\mathbf{M} = \mathbf{D} + \mathbf{L}$	$\mathbf{M} = \mathbf{D}^{\text{blk}} + \mathbf{L}^{\text{blk}}$

$$\underbrace{\begin{bmatrix} \times & \times & \times \\ \times & \times & \times \\ \times & \times & \times \end{bmatrix}}_{\mathbf{A}} = \underbrace{\begin{bmatrix} \times & & \\ & \times & \\ & & \times \end{bmatrix}}_{\mathbf{D}} + \underbrace{\begin{bmatrix} & & \\ \times & & \\ \times & \times & \end{bmatrix}}_{\mathbf{L}} + \underbrace{\begin{bmatrix} & \times & \times \\ & & \times \\ & & \end{bmatrix}}_{\mathbf{U}}$$

Résolution de systèmes linéaires — Méthodes instationnaires ...

On cherche $\mathbf{x} \in \mathbb{R}^n$ tel que $\boxed{\mathbf{Ax} = \mathbf{b}}$ avec $\mathbf{A} \in \mathbb{R}^{n\times n}$ et $\mathbf{b} \in \mathbb{R}^n$.

À la dernière séance, nous avons considéré des *schémas stationnaires* de la forme :

$$\left| \begin{array}{l} \mathbf{x}^{(0)} \text{ donné} \\ \mathbf{Mx}^{(k+1)} = \mathbf{Nx}^{(k)} + \mathbf{b}, \quad k \geq 0 \end{array} \right.$$

avec la décomposition régulière $\mathbf{A} = \mathbf{M} - \mathbf{N}$.

Pendant cette séance, nous considérons des *schémas instationnaires* de la forme :

$$\left| \begin{array}{l} \mathbf{x}^{(0)} \text{ donné} \\ \mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + \alpha^{(k)}\mathbf{p}^{(k)}, \quad k \geq 0 \end{array} \right.$$

où le pas $\alpha^{(k)}$ et la direction $\mathbf{p}^{(k)}$ sont à choisir.

Méthodes itératives pour les systèmes linéaires

Méthode du gradient conjugué

Introduction aux méthodes de Krylov

Méthode du gradient conjugué — Principe [1/2]

On cherche $\mathbf{x} \in \mathbb{R}^n$ tel que $\boxed{\mathbf{Ax} = \mathbf{b}}$ avec $\mathbf{A} \in \mathbb{R}^{n\times n}$ et $\mathbf{b} \in \mathbb{R}^n$,
où $\mathbf{A}$ est symétrique définie positive (SDP).

$$\mathbf{A} = \mathbf{A}^\top \quad \text{et} \quad (\mathbf{Av}, \mathbf{v}) > 0, \quad \forall \mathbf{v} \in \mathbb{R}^n \setminus \{0\}$$

Lien avec un problème de minimisation

On considère le problème de minimisation suivant :

Trouver $\mathbf{x} \in \mathbb{R}^n$ qui minimise la fonctionnelle $J(\mathbf{v}) = \frac{1}{2}(\mathbf{Av}, \mathbf{v}) - (\mathbf{b}, \mathbf{v})$.

Si $\mathbf{A}$ est une matrice SDP, on a les propriétés suivantes :

[P1] La fonctionnelle $J(\mathbf{v})$ est strictement convexe sur $\mathbb{R}^n$.

[P2] La fonctionnelle $J(\mathbf{v})$ admet un minimum unique.

[P3] Le minimum de $J(\mathbf{v})$, noté $\mathbf{v}_{\text{min}}$, est tel que $\nabla J|_{\mathbf{v}_{\text{min}}} = 0$ et $\mathbf{Av}_{\text{min}} = \mathbf{b}$.

Résoudre le problème de minimisation équivaut à résoudre le système !

Méthode du gradient conjugué — Principe [2/2]

Principe de la méthode du gradient conjugué

- En partant d'un vecteur $\mathbf{x}^{(0)}$, on calcule des vecteurs $\mathbf{x}^{(1)}$, $\mathbf{x}^{(2)}$, ... qui réduisent J.
- À chaque itération, on avance dans la direction $\mathbf{p}^{(k)}$ d'un pas $\alpha^{(k)}$:

$$\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + \alpha^{(k)}\mathbf{p}^{(k)} = \mathbf{x}^{(0)} + \sum_{i=0}^{k} \alpha^{(i)}\mathbf{p}^{(i)}$$

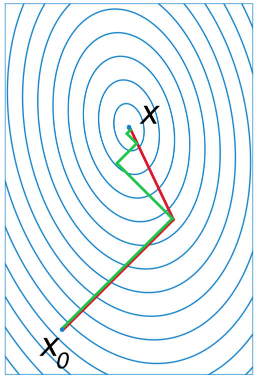

Illustration wikipedia

Méthode du gradient conjugué — Construction [1/3]

Méthode de la plus grande pente

À chaque itération :

- Choix de la direction $\mathbf{p}^{(k)}$ $\longrightarrow$ On prend le gradient : $\mathbf{p}^{(k)} = -\nabla J(\mathbf{x}^{(k)})$
- Choix du pas $\alpha^{(k)}$ $\longrightarrow$ On prend celui qui minimise $J(\mathbf{x}^{(k)} + \alpha^{(k)}\mathbf{p}^{(k)})$

On a

$J(\mathbf{v}) = \frac{1}{2}(\mathbf{A}\mathbf{v},\mathbf{v}) - (\mathbf{b},\mathbf{v})$

$\nabla J(\mathbf{v}) = \mathbf{A}\mathbf{v} - \mathbf{b}$ (Du coup, $\mathbf{p}^{(k)} = \mathbf{b} - \mathbf{A}\mathbf{x}^{(k)}$, qui est le résidu !)

$\min_\alpha J(\mathbf{x}^{(k)} + \alpha\mathbf{p}^{(k)}) \Leftrightarrow \alpha = (\mathbf{b} - \mathbf{A}\mathbf{x}^{(k)},\mathbf{p}^{(k)})/(\mathbf{A}\mathbf{p}^{(k)},\mathbf{p}^{(k)})$

Algorithme de la plus grande pente

$\mathbf{x}^{(0)} \in \mathbb{R}^n$
$\mathbf{p}^{(0)} = \mathbf{b} - \mathbf{A}\mathbf{x}^{(0)}$
for $k = 0, 1, \ldots$ **do**
  $\alpha^{(k)} = (\mathbf{p}^{(k)},\mathbf{p}^{(k)})/(\mathbf{A}\mathbf{p}^{(k)},\mathbf{p}^{(k)})$ Calcul du pas
  $\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + \alpha^{(k)}\mathbf{p}^{(k)}$ Mise à jour
  $\mathbf{p}^{(k+1)} = \mathbf{b} - \mathbf{A}\mathbf{x}^{(k+1)}$ Calcul de la direction/du résidu
  if $\|\mathbf{p}^{(k+1)}\| \le \varepsilon\|\mathbf{p}^{(0)}\|$ **then break**
end

Méthode du gradient conjugué — Construction [2/3]

Méthode du gradient conjugué

- On souhaite $\{\mathbf{p}^{(k)}\}_{k=0\ldots n-1}$ tel qu'ils constituent une base de $\mathbb{R}^n$.
  On prend une base $\mathbf{A}$–orthogonale, *i.e.* orthogonale par rapport au produit scalaire $(\mathbf{A}\cdot,\cdot)$:
  $$(\mathbf{A}\mathbf{p}^{(k)},\mathbf{p}^{(l)}) = 0, \quad \forall k \neq l$$
- On souhaite $\{\alpha^{(k)}\}_{k=0\ldots n-1}$ tel que
  $$\mathbf{x} = \mathbf{x}^{(0)} + \sum_{k=0}^{n-1} \alpha^{(k)}\mathbf{p}^{(k)}$$
  où $\mathbf{x}$ est la solution du problème.

Algorithme du gradient conjugué

$\mathbf{x}^{(0)} \in \mathbb{R}^n$
for $k = 0, 1, \ldots$ **do**
  $\vdots$
  $\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + \alpha^{(k)}\mathbf{p}^{(k)}$ $= \mathbf{x}^{(0)} + \sum_{l=0}^{k} \alpha^{(l)}\mathbf{p}^{(l)}$ Mise à jour
  $\vdots$
end

Méthode du gradient conjugué — Construction [3/3]

Méthode du gradient conjugué *(suite)*

À chaque itération :

- Choix de $\mathbf{p}^{(k)}$ $\longrightarrow$ On prend la partie de $\mathbf{r}^{(k)}$ qui est $\mathbf{A}$–orthogonale à $\mathbf{p}^{(k-1)}$.
- Choix du pas $\alpha^{(k)}$ $\longrightarrow$ On prend celui qui minimise $J(\mathbf{x}^{(k)} + \alpha^{(k)}\mathbf{p}^{(k)})$

Algorithme du gradient conjugué

$\mathbf{x}^{(0)} \in \mathbb{R}^n$
$\mathbf{r}^{(0)} = \mathbf{b} - \mathbf{A}\mathbf{x}^{(0)}$
$\mathbf{p}^{(0)} = \mathbf{r}^{(0)}$
for $k = 0, 1, \ldots$ **do**
  $\alpha^{(k)} = (\mathbf{r}^{(k)},\mathbf{p}^{(k)})/(\mathbf{A}\mathbf{p}^{(k)},\mathbf{p}^{(k)})$ Calcul du pas
  $\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + \alpha^{(k)}\mathbf{p}^{(k)}$ $= \mathbf{x}^{(0)} + \sum_{l=0}^{k} \alpha^{(l)}\mathbf{p}^{(l)}$ Mise à jour
  $\mathbf{r}^{(k+1)} = \mathbf{r}^{(k)} - \alpha^{(k)}\mathbf{A}\mathbf{p}^{(k)}$ $= \mathbf{b} + \mathbf{A}\mathbf{x}^{(k+1)}$ Calcul du résidu
  $\beta^{(k)} = -(\mathbf{A}\mathbf{r}^{(k+1)},\mathbf{p}^{(k)})/(\mathbf{A}\mathbf{p}^{(k)},\mathbf{p}^{(k)})$
  $\mathbf{p}^{(k+1)} = \mathbf{r}^{(k+1)} + \beta^{(k)}\mathbf{p}^{(k)}$ Calcul de la direction
  if $\|\mathbf{r}^{(k+1)}\| \le \varepsilon\|\mathbf{r}^{(0)}\|$ **then break**
end

Méthode du gradient conjugué — Discussion

Aspects théoriques *[pour des matrices symétriques définies positives (SDP)]*

- On peut montrer que toutes les directions sont alors $\mathbf{A}$–orthogonales.
- Par construction, convergence en maximum n itérations ! *(en précision ∞)*
- On peut montrer que l'erreur $\mathbf{e}^{(k)} = \mathbf{x} - \mathbf{x}^{(k)}$ vérifie
  $$\sqrt{(\mathbf{A}\mathbf{e}^{(k)},\mathbf{e}^{(k)})} \le \left(\frac{\sqrt{\kappa}-1}{\sqrt{\kappa}+1}\right)^k \sqrt{(\mathbf{A}\mathbf{e}^{(0)},\mathbf{e}^{(0)})}$$
  avec le nombre de conditionnement $\kappa = \|\mathbf{A}\|\,\|\mathbf{A}^{-1}\|$.
  Vu $\mathbf{A}$ SDP : $\kappa = \lambda_{\max}/\lambda_{\min}$. Les valeurs propres max/min influence la vitesse de convergence. Plus κ est proche de 1, plus la convergence est rapide.

Aspects algorithmiques

- Opérations d'algèbre linéaire (BLAS 1 et 2) $\Rightarrow$ Très facile à paralléliser !
- Calcul de produits scalaires et de normes $\Rightarrow$ Communications collectives.

Méthode de choix pour les matrices SDP !
Extensions pour des matrices plus générales ?

Méthodes itératives pour les systèmes linéaires

Méthode du gradient conjugué

Introduction aux méthodes de Krylov

Espaces de Krylov — Motivation

- La méthode du gradient conjugué repose sur l'*ajout d'un terme de mise à jour* à la solution courante :

$$\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + \alpha^{(k)}\mathbf{p}^{(k)} \quad \Longrightarrow \quad \boxed{\mathbf{x}^{(k+1)} - \mathbf{x}^{(0)} = \sum_{l=0}^{k} \alpha^{(l)}\mathbf{p}^{(l)}}$$

La mise à jour se fait par le choix d'une direction et d'un pas d'avancement.

- On s'intéresse aux *espaces générés par les termes de mise à jour successifs*.

$\longrightarrow$ Pour la méthode du gradient conjugué : $\{\mathbf{p}^l\}_{l=0}^k$.

Ce concept permet d'étudier les méthodes itératives, et de proposer des méthodes valables pour matrices générales non-symétriques/non-hermitiennes.

Définition : Espace de Krylov

L'espace de Krylov d'ordre k associé à $\mathbf{A} \in \mathbb{R}^{n\times n}$ et $\mathbf{v} \in \mathbb{R}^n$, avec $k < n$, est l'espace vectoriel généré par $\mathbf{v}$ et ses $k-1$ produits itérés par $\mathbf{A}$:

$$\mathscr{K}_k(\mathbf{A},\mathbf{v}) := \mathrm{span}(\mathbf{v}, \mathbf{A}\mathbf{v}, \ldots, \mathbf{A}^{k-1}\mathbf{v}).$$

Espaces de Krylov — Définition et propriétés

Définition : Espace de Krylov

L'espace de Krylov d'ordre k associé à $\mathbf{A} \in \mathbb{R}^{n\times n}$ et $\mathbf{v} \in \mathbb{R}^n$, avec $k < n$, est l'espace vectoriel généré par $\mathbf{v}$ et ses $k-1$ produits itérés par $\mathbf{A}$:

$$\mathscr{K}_k(\mathbf{A},\mathbf{v}) := \mathrm{span}(\mathbf{v}, \mathbf{A}\mathbf{v}, \ldots, \mathbf{A}^{k-1}\mathbf{v}).$$

Propriétés

- $\mathscr{K}_k(\mathbf{A},\mathbf{v}) \subseteq \mathscr{K}_{k+l}(\mathbf{A},\mathbf{v}) \subseteq \mathbb{R}^n, \quad \forall l \geq 0$
- $\mathbf{A}\mathscr{K}_k(\mathbf{A},\mathbf{v}) \subseteq \mathscr{K}_{k+1}(\mathbf{A},\mathbf{v}), \quad \forall k$
- $\dim\left(\mathscr{K}_k(\mathbf{A},\mathbf{v})\right) = \min(k,$ degré min. des poly. non-nuls $\mathscr{P}$ tel que $\mathscr{P}(\mathbf{A})\mathbf{v} = 0)$
- La suite $\left(\mathscr{K}_k(\mathbf{A},\mathbf{v})\right)_k$ est strict. croiss. de 1 à $k_{\max}$, puis constante à partir de $k_{\max}$, où $k_{\max}$ est la dimension max. des espaces de Krylov de la suite $\left(\mathscr{K}_k(\mathbf{A},\mathbf{v})\right)_k$.

Exemple

$$\mathbf{A} = \begin{bmatrix} 0 & 1 & 1 \\ 1 & 4 & -2 \\ 2 & 2 & -1 \end{bmatrix} \qquad \mathbf{v} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \qquad \mathbf{A}\mathbf{v} = \begin{bmatrix} 0 \\ 1 \\ 2 \end{bmatrix} \qquad \mathbf{A}^2\mathbf{v} = \begin{bmatrix} 3 \\ 0 \\ 0 \end{bmatrix}$$

$$\mathscr{K}_1(\mathbf{A},\mathbf{v}) = \mathrm{span}(\mathbf{v}) \qquad \mathscr{K}_2(\mathbf{A},\mathbf{v}) = \mathrm{span}(\mathbf{v},\mathbf{A}\mathbf{v}) \qquad \mathscr{K}_3(\mathbf{A},\mathbf{v}) = \mathrm{span}(\mathbf{v},\mathbf{A}\mathbf{v},\mathbf{A}^2\mathbf{v})$$

$$\dim(\mathscr{K}_1(\mathbf{A},\mathbf{v})) = 1 \qquad \dim(\mathscr{K}_2(\mathbf{A},\mathbf{v})) = 2 \qquad \dim(\mathscr{K}_3(\mathbf{A},\mathbf{v})) = 2$$

Espaces de Krylov — Analyse de la méthode du gradient conjugué

$$\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + \alpha^{(k)}\mathbf{p}^{(k)}$$
$$\mathbf{r}^{(k+1)} = \mathbf{r}^{(k)} - \alpha^{(k)}\mathbf{A}\mathbf{p}^{(k)}$$
$$\mathbf{p}^{(k+1)} = \mathbf{r}^{(k+1)} + \beta^{(k)}\mathbf{p}^{(k)}$$

Propriété : $\boxed{\mathbf{x}^{(k+1)} - \mathbf{x}^{(0)} \in \mathscr{K}_{k+1}\left(\mathbf{A},\mathbf{r}^{(0)}\right)}$

<u>Preuve</u> *(pour information)*

On a :
$$\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + \alpha^{(k)}\mathbf{p}^{(k)} = \mathbf{x}^{(0)} + \textstyle\sum_{i=0}^{k}\alpha^{(i)}\mathbf{p}^{(i)}$$

À l'initialisation, on a : $\mathbf{p}^{(0)} = \mathbf{r}^{(0)} \Rightarrow \mathbf{p}^{(0)} \in \mathscr{K}_1(\mathbf{A},\mathbf{r}^{(0)})$.

À l'itération k, on a :
$$\begin{aligned}\mathbf{p}^{(k+1)} &= \mathbf{r}^{(k+1)} + \beta^{(k)}\mathbf{p}^{(k)} \\ &= \mathbf{r}^{(k)} - \alpha^{(k)}\mathbf{A}\mathbf{p}^{(k)} + \beta^{(k)}\mathbf{p}^{(k)} \\ &= \mathbf{r}^{(0)} - \sum_{i=0}^{k}\alpha^{(i)}\mathbf{A}\mathbf{p}^{(i)} + \beta^{(k)}\mathbf{p}^{(k)}\end{aligned}$$

Si $\mathbf{p}^{(i)} \in \mathscr{K}_{i+1}(\mathbf{A},\mathbf{r}^{(0)})$ pour $\forall i < k$, alors $\mathbf{p}^{(k+1)} \in \mathscr{K}_{k+2}(\mathbf{A},\mathbf{r}^{(0)})$.

Donc : $\mathbf{x}^{(k+1)} - \mathbf{x}^{(0)} = \sum_{i=0}^{k}\alpha^{(i)}\mathbf{p}^{(i)} \in \mathscr{K}_{k+1}\left(\mathbf{A},\mathbf{r}^{(0)}\right)$ $\square$

Méthodes de Krylov — Principe

Propriétés de la méthode du gradient conjugué *(suite)*

Propriété : $\boxed{\mathbf{x}^{(k)} - \mathbf{x}^{(0)} \in \mathscr{K}_k\left(\mathbf{A}, \mathbf{r}^{(0)}\right)}$

Propriété : $\boxed{\mathbf{x}^{(k)} = \mathbf{x}^{(0)} + \underset{\mathbf{y} \in \mathscr{K}_k(\mathbf{A}, \mathbf{r}^{(0)})}{\operatorname{argmin}} J\left(\mathbf{x}^{(0)} + \mathbf{y}\right)}$

À chaque itération k, on a la meilleure solution $\mathbf{x}^{(k)}$ au sens *"elle minimise J"* tel que $\mathbf{x}^{(k)} - \mathbf{x}^{(0)}$ appartient à l'espace $\mathscr{K}_k\left(\mathbf{A}, \mathbf{r}^{(0)}\right)$.

... méthode limitée aux matrices SDP.

Principe des méthodes de Krylov

On cherche une méthode qui donne

$\boxed{\mathbf{x}^{(k)} - \mathbf{x}^{(0)} = \mathscr{P}_{k-1}(\mathbf{A})\, \mathbf{r}^{(0)}}$ où $\mathscr{P}_{k-1}(\cdot)$ est un polynôme de degré $k-1$

tel que $\mathbf{x}^{(k)}$ est la *"meilleure solution"* avec $\mathbf{x}^{(k)} - \mathbf{x}^{(0)} \in \mathscr{K}_k(\mathbf{A}, \mathbf{r}^{(0)})$.

La méthode du GMRES *(generalized minimal residual)* est une méthode de Krylov basée sur la minimisation du résidu à chaque itération. *(Il y a d'autres méthodes de Krylov.)*

Méthodes de Krylov — GMRES [1/3]

La méthode du GMRES *(generalized minimal residual)* repose sur 2 étapes, effectuée à chaque itération k :

- **Étape 1 — Construction rapide d'une base orthonormée pour $\mathscr{K}_k(\mathbf{A}, \mathbf{r}^{(0)})$**
  On construit une base orthonormée $\{\mathbf{v}^{(i)}\}_{i=1\dots k}$ pour

$$\mathscr{K}_k\left(\mathbf{A}, \mathbf{r}^{(0)}\right) := \operatorname{span}\left(\mathbf{r}^{(0)}, \mathbf{A}\mathbf{r}^{(0)}, \dots, \mathbf{A}^{k-1}\mathbf{r}^{(0)}\right) = \operatorname{span}\left(\{\mathbf{v}^{(i)}\}_{i=1\dots k}\right).$$

- **Étape 2 — Résolution rapide d'un prob. de min. pour avoir la *"meilleure sol."***
  On cherche la solution $\mathbf{x}^{(k)}$ qui minimise la norme 2 du résidu $\mathbf{r}^{(k)} = \mathbf{b} - \mathbf{A}\mathbf{x}^{(k)}$ avec le terme de mise à jour dans l'espace $\mathscr{K}_k(\mathbf{A}, \mathbf{r}^{(0)})$:

$$\boxed{\mathbf{x}^{(k)} = \mathbf{x}^{(0)} + \operatorname{argmin}_{\mathbf{y} \in \mathscr{K}_k(\mathbf{A}, \mathbf{r}^{(0)})} \|\mathbf{b} - \mathbf{A}(\mathbf{x}^{(0)} + \mathbf{y})\|_2}$$

  Résoudre ce problème revient à résoudre un problème aux moindres carrés.
  Astuce GMRES *(pour information)* : Facto. QR de la matrice rect. du système du prob. aux moindres carrés. Utilisation de matrices de Givens pour réso. rapide.

Méthodes de Krylov — GMRES [2/3]

Algorithme du GMRES *(étapes principales, sans détails)*

$\mathbf{x}^{(0)} \in \mathbb{R}^n$
$\mathbf{r}^{(0)} = \mathbf{b} - \mathbf{A}\mathbf{x}^0$
$\mathbf{v}^{(1)} = \mathbf{r}^{(0)} / \|\mathbf{r}^{(0)}\|$
for $k = 1, 2 \dots$ **do**
  // Construction de la base orthonormée
  $\mathbf{w}^{(k)} = \mathbf{A}\mathbf{v}^{(k)}$
  for $i = 1, \dots, k$ **do**
    $\mathbf{w}^{(k)} = \mathbf{w}^{(k)} - (\mathbf{w}^{(k)}, \mathbf{v}^{(i)})\mathbf{v}^{(i)}$
  end
  $\mathbf{v}^{(k+1)} = \mathbf{w}^{(k)} / \|\mathbf{w}^{(k)}\|_2$
  // Résolution du problème de minimisation
  $\mathbf{z}^{(k)} = \operatorname{argmin}_{\mathbf{z} \in \mathbb{R}^k} \|\mathbf{b} - \mathbf{A}\mathbf{V}^{(k)}\mathbf{z}\|_2$
  // Mise à jour de la solution
  $\mathbf{x}^{(k)} = \mathbf{x}^{(0)} + \mathbf{V}^{(k)}\mathbf{z}^{(k)}$
end

$\mathbf{V}^{(k)}$ est une matrice rect. dont les colonnes sont les vecteurs de base $\{\mathbf{v}^{(i)}\}_{i=1\dots k}$.

Méthodes de Krylov — GMRES [3/3]

Aspects théoriques

- Par construction, convergence en maximum n itérations ! *(en précision ∞)*
- Si moins d'itérations sont nécessaires *(breakdown)*, arrêt pendant la contruction de la base.

Aspects algorithmiques

- *Le coût de l'algorithme augmente à chaque itération en $\mathcal{O}(k^2)$!*
  - Stockage d'un vecteur de base supplémentaire.
  - Orthogonalisation par rapport à un vecteur supplémentaire.
  - Résolution du problème des min. et mise à jour de la sol. plus coûteuses.

  Pour limiter le coût, on utilise l'algorithme *avec restart* : redémarrage de la procédure en utilisant la solution courante comme solution initiale.
- *Algorithme facile à paralléliser !*
  - Opérations d'algèbre linéaire (BLAS 1 et 2) $\Rightarrow$ Très facile à paralléliser !
  - Calcul de produits scalaires et de normes $\Rightarrow$ Communications collectives.

Méthode de choix pour les matrices non-symétriques. Très utilisée !
*Nécéssité de limiter le nombre d'itérations ⇒ **Préconditionnement ...***

Summary

- **Stationary methods** $\left(\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + \mathbf{M}^{-1}\mathbf{r}^{(k)}\right)$
  - Jacobi and Gauss-Seidel (G.-S.) methods
  - Improvements: *"relaxation" (JOR and SOR)* and *"by block"* approaches
  - Algorithmic aspects:
    - Matrix-vector products and linear combinations
    - Parallelism easy for Jacobi, a bit more complicated for G.-S.
    - Finite difference problem → *red/black approach for G.-S.*
- **Unstationary methods** $\left(\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + \alpha^{(k)}\mathbf{p}^{(k)}\right)$
  - Steepest descent and Conjuguate Gradient (CG) methods
  - If $\mathbf{A}$ SDP: link with quadratic optimisation, conv. in max. n iterations
  - Algorithmic aspects: matrix-vector prod., lin. comb., scalar products
- **Krylov methods** $\left(\mathbf{x}^{(k)} = \mathbf{x}^{(0)} + \mathbf{V}^{(k)}\mathbf{z}^{(k)}\right)$ which are unstationary methods
  - GMRES
  - For nonsingular $\mathbf{A}$: conv. in max. n iterations
  - Algorithmic aspects: iterations of increasing cost, parallelism is easy

Quelques ressources

- *Méthodes Numériques : Algorithmes, analyse et applications*
  A. Quarteroni, R. Sacco, F. Saleri (2007), Springer
- *Calcul scientifique parallèle*
  F. Magoulès et F.-X. Roux (2017), Dunod
- *Calcul scientifique parallèle*
  P. Ciarlet et E. Jamelot, polycopié de cours
- M. H. Gutknecht. *"A Brief Introduction to Krylov Space Methods for Solving Linear Systems"*, Proc. of the Int. Symp. on Front. of Comput. Sci. (2005) [Preprint]