Parallel Scientific Computing

Course AMS301 — Fall 2023 — Lecture 6

Iterative methods for linear systems (2)

Nonstationary iterative methods

Axel Modave

Find $\mathbf{x} \in \mathbb{R}^n$ such that $\boxed{\mathbf{A}\mathbf{x} = \mathbf{b}}$ with $\mathbf{A} \in \mathbb{R}^{n\times n}$ and $\mathbf{b} \in \mathbb{R}^n$.

Solution procedures

- **Direct methods:** Factorization of $\mathbf{A}$ into triangular and diagonal matrices (ex. $\mathbf{A} = \mathbf{L}\mathbf{U}$) and solution of simpler problems.

$$\mathbf{A}\mathbf{x} = \mathbf{b} \quad \Leftrightarrow \quad \mathbf{L}\mathbf{U}\mathbf{x} = \mathbf{b} \quad \Leftrightarrow \quad \left| \begin{array}{l} \mathbf{L}\mathbf{y} = \mathbf{b} \\ \mathbf{U}\mathbf{x} = \mathbf{y} \end{array} \right.$$

  Advantages: exact solution known after a given number of operations
  Difficulties: heavy computational cost *(operations/memory)*, hard to parallelize

- **Iterative methods:** Iterative procedure to minimizing an error $\|\mathbf{x}^{(\ell)} - \mathbf{x}_{\text{ref}}\|$ and/or a residual $\|\mathbf{A}\mathbf{x}^{(\ell)} - \mathbf{b}\|$.

$$\left| \begin{array}{rl} \mathbf{x}^{(0)} &= \mathrm{Iter}_{(0)}\left(\mathbf{A}, \mathbf{b}\right) \\ \mathbf{x}^{(\ell+1)} &= \mathrm{Iter}^{(\ell+1)}\left(\mathbf{x}^{(\ell)}, \mathbf{x}^{(\ell-1)}, \ldots, \mathbf{A}, \mathbf{b}\right), \quad \text{for } \ell \geq 0 \end{array} \right.$$

  Advantages: limited cost per iteration *(operations/memory)*, easy to parallelize
  Difficulties: approximate solution, control of the convergence of the process

Find $\mathbf{x} \in \mathbb{R}^n$ such that $\boxed{\mathbf{A}\mathbf{x} = \mathbf{b}}$ with $\mathbf{A} \in \mathbb{R}^{n\times n}$ and $\mathbf{b} \in \mathbb{R}^n$.

- **Iterative methods:**

$$\left|\begin{aligned} \mathbf{x}^{(0)} &= \mathrm{Iter}_{(0)}\,(\mathbf{A}, \mathbf{b}) \\ \mathbf{x}^{(\ell+1)} &= \mathrm{Iter}^{(\ell+1)}\left(\mathbf{x}^{(\ell)}, \mathbf{x}^{(\ell-1)}, \ldots, \mathbf{A}, \mathbf{b}\right), \quad \text{for } \ell \geq 0 \end{aligned}\right.$$

The order of the method is the numb. of steps which the current iter. depends on.
Stationary method if the functions $\mathrm{Iter}^{(\ell)}$ are indep. of ℓ, otherwise nonstationary
Linear method if the functions $\mathrm{Iter}^{(\ell)}$ are linear, otherwise nonlinear

In a previous session, we considered statio. linear iterative schemes of first order:

$$\left|\begin{aligned} &\mathbf{x}^{(0)} \text{ given} \\ &\mathbf{x}^{(\ell+1)} = \mathbf{B}\mathbf{x}^{(\ell)} + \mathbf{f}, \quad \ell \geq 0 \end{aligned}\right.$$

where $\mathbf{B} \in \mathbb{R}^{n\times n}$ is the iteration matrix and $\mathbf{f} \in \mathbb{R}^n$ depends on $\mathbf{b}$.

Stationary linear iterative method of first order

Regular decomposition: $\boxed{\mathbf{A} = \mathbf{M} - \mathbf{N}}$ where $\mathbf{M} \in \mathbb{R}^{n\times n}$ is inversible.

Stationary method	
$\mathbf{x}^{(0)} \in \mathbb{C}^n$ **for** $\ell = 0, 1, \dots$ **do** $\quad \mathbf{M}\mathbf{x}^{(\ell+1)} = \mathbf{N}\mathbf{x}^{(\ell)} + \mathbf{b}$ **end**	i.e. $\boxed{\mathbf{x}^{(\ell+1)} = \mathbf{B}\mathbf{x}^{(\ell)} + \mathbf{f}}$ with $\mathbf{B} = \mathbf{M}^{-1}\mathbf{N}$ and $\mathbf{f} = \mathbf{M}^{-1}\mathbf{b}$

Choices

	By points	By blocks
Jacobi	$\mathbf{M} = \mathbf{D}$	$\mathbf{M} = \mathbf{D}^{\mathrm{blk}}$
Gauss-Seidel	$\mathbf{M} = \mathbf{D} + \mathbf{L}$	$\mathbf{M} = \mathbf{D}^{\mathrm{blk}} + \mathbf{L}^{\mathrm{blk}}$

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

Find $\mathbf{x} \in \mathbb{R}^n$ such that $\boxed{\mathbf{A}\mathbf{x} = \mathbf{b}}$ with $\mathbf{A} \in \mathbb{R}^{n \times n}$ and $\mathbf{b} \in \mathbb{R}^n$.

In a previous session, we considered statio. linear iterative schemes of first order:

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

where $\mathbf{B} \in \mathbb{R}^{n \times n}$ is the iteration matrix and $\mathbf{f} \in \mathbb{R}^n$ depends on $\mathbf{b}$.

In this session, we consider nonstationary linear iterative schemes of the form:

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

with the step $\alpha^{(\ell)}$ and the direction $\mathbf{p}^{(\ell)}$ must be chosen.

Nonstationary iterative methods for linear systems

Conjugate gradient method

Interlude on Krylov spaces

Few words about GMRES

Find $\mathbf{x} \in \mathbb{R}^n$ such that $\boxed{\mathbf{A}\mathbf{x} = \mathbf{b}}$ with $\mathbf{A} \in \mathbb{R}^{n \times n}$ and $\mathbf{b} \in \mathbb{R}^n$,
where $\mathbf{A}$ is symmetric positive definite (SPD).

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

Link with a minimization problem

We consider the following minimization problem:

Find $\mathbf{x} \in \mathbb{R}^n$ that minimizes the functional $J(\mathbf{v}) = \dfrac{1}{2}\big(\mathbf{A}\mathbf{v}, \mathbf{v}\big) - \big(\mathbf{b}, \mathbf{v}\big)$.

If $\mathbf{A}$ is an SPD matrix:

- The functional $J(\mathbf{v})$ is strictly convex on $\mathbb{R}^n$.
- The functional $J(\mathbf{v})$ has a unique minium.
- The minimum of $J(\mathbf{v})$, denoted $\mathbf{v}_{\text{min}}$, is such that $\nabla J|_{\mathbf{v}_{\text{min}}} = 0$ and $\mathbf{A}\mathbf{v}_{\text{min}} = \mathbf{b}$.

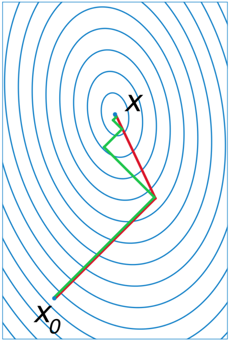
Solving the minimization problem is equivalent to solving the system!

General idea

- Starting from a vector $\mathbf{x}^{(0)}$, we compute vectors $\mathbf{x}^{(1)}$, $\mathbf{x}^{(2)}$, ... such that J is min.
- At each iteration, we take one step $\alpha^{(\ell)}$ along direction $\mathbf{p}^{(\ell)}$:

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

Illustration for a problem with $n = 2$:



(Source: Wikipedia)

Steepest descent method

At each step:

- Choice of direction $\mathbf{p}^{(\ell)}$ $\longrightarrow$ We take the gradient: $\mathbf{p}^{(\ell)} = -\nabla J(\mathbf{x}^{(\ell)})$
- Choice of step $\alpha^{(\ell)}$ $\longrightarrow$ We take the one that minimizes $J(\mathbf{x}^{(\ell)} + \alpha^{(\ell)}\mathbf{p}^{(\ell)})$.

We have

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

$$\nabla J(\mathbf{v}) = \mathbf{A}\mathbf{v} - \mathbf{b}$$

(Then, $\mathbf{p}^{(\ell)} = \mathbf{b} - \mathbf{A}\mathbf{x}^{(\ell)}$, which is the residual!)

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

Steepest descent method

$\mathbf{x}^{(0)} \in \mathbb{R}^n$

$\mathbf{p}^{(0)} = \mathbf{b} - \mathbf{A}\mathbf{x}^{(0)}$

for $\ell = 0, 1, \ldots$ **do**

- $\alpha^{(\ell)} = \big(\mathbf{p}^{(\ell)},\mathbf{p}^{(\ell)}\big)/\big(\mathbf{A}\mathbf{p}^{(\ell)},\mathbf{p}^{(\ell)}\big)$ — Comput. of step
- $\mathbf{x}^{(\ell+1)} = \mathbf{x}^{(\ell)} + \alpha^{(\ell)}\mathbf{p}^{(\ell)}$ — Update
- $\mathbf{p}^{(\ell+1)} = \mathbf{b} - \mathbf{A}\mathbf{x}^{(\ell+1)}$ — Comput. of direction/residual
- **if** $\|\mathbf{p}^{(\ell+1)}\| \leq \varepsilon\|\mathbf{p}^{(0)}\|$ **then break**

end

Conjugate gradient method

- We take $\{\mathbf{p}^{(\ell)}\}_{\ell=0\dots n-1}$ such that they form a basis of $\mathbb{R}^n$.
  We take a $\mathbf{A}-$orthogonal basis, i.e. orthogonal with the scalar product $(\mathbf{A}\cdot,\cdot)$:
  $$\left(\mathbf{A}\mathbf{p}^{(i)},\mathbf{p}^{(j)}\right)=0, \quad \forall i\neq j$$
- We take $\{\alpha^{(\ell)}\}_{\ell=0\dots n-1}$ such that
  $$\mathbf{x}=\mathbf{x}^{(0)}+\sum_{i=0}^{n-1}\alpha^{(i)}\mathbf{p}^{(i)}$$
  where $\mathbf{x}$ is the solution of the problem.

Conjugate gradient method

$\mathbf{x}^{(0)}\in\mathbb{R}^n$

$\mathbf{p}^{(0)}=\mathbf{b}-\mathbf{A}\mathbf{x}^{(0)}$

for $\ell=0,1,\dots$ **do**

$\quad\vdots$

$\quad\mathbf{x}^{(\ell+1)}=\mathbf{x}^{(\ell)}+\alpha^{(\ell)}\mathbf{p}^{(\ell)} \qquad = \mathbf{x}^{(0)}+\sum_{i=0}^{\ell}\alpha^{(i)}\mathbf{p}^{(i)}$ Update

$\quad\vdots$

end

Conjugate gradient method *(continuation)*

At each step:

- Choice of direction $\mathbf{p}^{(\ell)}$ $\longrightarrow$ Part of $\mathbf{r}^{(\ell)}$ that is $\mathbf{A}-$orthogonal with $\mathbf{p}^{(\ell-1)}$
- Choice of step $\alpha^{(\ell)}$ $\longrightarrow$ Value that minimizes $J(\mathbf{x}^{(\ell)} + \alpha^{(\ell)}\mathbf{p}^{(\ell)})$

Conjugate gradient method

$\mathbf{x}^{(0)} \in \mathbb{R}^n$

$\mathbf{r}^{(0)} = \mathbf{b} - \mathbf{A}\mathbf{x}^{(0)}$

$\mathbf{p}^{(0)} = \mathbf{r}^{(0)}$

for $\ell = 0, 1, \ldots$ **do**

- $\alpha^{(\ell)} = \big(\mathbf{r}^{(\ell)}, \mathbf{p}^{(\ell)}\big) / \big(\mathbf{A}\mathbf{p}^{(\ell)}, \mathbf{p}^{(\ell)}\big)$ — Comput. of step
- $\mathbf{x}^{(\ell+1)} = \mathbf{x}^{(\ell)} + \alpha^{(\ell)}\mathbf{p}^{(\ell)}$ $\quad = \mathbf{x}^{(0)} + \sum_{i=0}^{\ell} \alpha^{(i)}\mathbf{p}^{(i)}$ — Update
- $\mathbf{r}^{(\ell+1)} = \mathbf{r}^{(\ell)} - \alpha^{(\ell)}\mathbf{A}\mathbf{p}^{(\ell)}$ $\quad = \mathbf{b} - \mathbf{A}\mathbf{x}^{(\ell+1)}$ — Comput. of residual
- $\beta^{(\ell)} = -\big(\mathbf{A}\mathbf{r}^{(\ell+1)}, \mathbf{p}^{(\ell)}\big) / \big(\mathbf{A}\mathbf{p}^{(\ell)}, \mathbf{p}^{(\ell)}\big)$
- $\mathbf{p}^{(\ell+1)} = \mathbf{r}^{(\ell+1)} + \beta^{(\ell)}\mathbf{p}^{(\ell)}$ — Comput. of direction
- **if** $\big\|\mathbf{r}^{(\ell+1)}\big\| \leq \varepsilon \big\|\mathbf{r}^{(0)}\big\|$ **then break**

end

Conjugate gradient method — Discussion

Theoretical aspects

- Method for symmetric pos. def. (SDP) and Hermitian pos. def. (HPD) matrices
- All the directions are $\mathbf{A}-$orthogonal.
- By construction, convergence with maximum n iterations! *(if ∞ accuracy)*
- The error $\mathbf{e}^{(\ell)} = \mathbf{x} - \mathbf{x}^{(\ell)}$ verifies

$$\sqrt{\left(\mathbf{A}\mathbf{e}^{(\ell)}, \mathbf{e}^{(\ell)}\right)} \leq \left(\frac{\sqrt{\kappa}-1}{\sqrt{\kappa}+1}\right)^{\ell} \sqrt{\left(\mathbf{A}\mathbf{e}^{(0)}, \mathbf{e}^{(0)}\right)}$$

  with the condition number $\kappa = \|\mathbf{A}\| \, \|\mathbf{A}^{-1}\|$.

  Because $\mathbf{A}$ is SPD: $\kappa = \lambda_{\max}/\lambda_{\min}$. The max/min eigenvalues influence the convergence rate. When κ is close to 1, convergence is fast.

Algorithmic aspects

- Linear algebraic operations (BLAS 1 and 2) $\Rightarrow$ Easy for parallel computing
- Computation of scalar products and norms $\Rightarrow$ Collective communications

Best method for systems with SPD matrices!
Extensions for general matrices?

Nonstationary iterative methods for linear systems

Conjugate gradient method

Interlude on Krylov spaces

Few words about GMRES

Krylov spaces — Motivation

The conjugate gradient method relies on adding an update term to the current solution:

$$\mathbf{x}^{(\ell+1)} = \mathbf{x}^{(\ell)} + \alpha^{(\ell)}\mathbf{p}^{(\ell)},$$

with the update direction $\mathbf{p}^{(\ell)}$ and the step $\alpha^{(\ell)}$.

The update term can be rewritten as:

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

It belongs to the subspace:

$$\operatorname{span}\big(\mathbf{p}^{(0)}, \dots, \mathbf{p}^{(\ell)}\big) \subseteq \mathbb{R}^n$$

Toward general iterative methods . . .

General iterative methods for more general non-symmetric/non-Hermitian matrices can be built by considering Krylov subspaces, e.g.

$$\mathcal{K}_\ell(\mathbf{A}, \mathbf{r}^{(0)}) := \operatorname{span}(\mathbf{r}^{(0)}, \mathbf{A}\mathbf{r}^{(0)}, \dots, \mathbf{A}^{\ell-1}\mathbf{r}^{(0)}).$$

These methods rely on 2 steps:

1. Building a basis $\left\{\mathbf{v}^{(1)}, \dots, \mathbf{v}^{(\ell)}\right\}$ for a Krylov subspace
2. Solving a minimization problem to get the update term towards the *“best solution”*

Krylov spaces — Definition and properties

Definition – Krylov subspace

The order-ℓ Krylov subspace associated to $\mathbf{A} \in \mathbb{R}^{n\times n}$ and $\mathbf{v} \in \mathbb{R}^n$, with $\ell < n$, is the linear subspace spanned by the images of $\mathbf{v}$ under the first ℓ powers of $\mathbf{A}$, starting from $\mathbf{A}^0$:

$$\mathcal{K}_\ell(\mathbf{A}, \mathbf{v}) := \operatorname{span}(\mathbf{v}, \mathbf{A}\mathbf{v}, \ldots, \mathbf{A}^{\ell-1}\mathbf{v}).$$

Properties

- $\mathcal{K}_\ell(\mathbf{A}, \mathbf{v}) \subseteq \mathcal{K}_{\ell+i}(\mathbf{A}, \mathbf{v}) \subseteq \mathbb{R}^n, \quad \forall i \geq 0$
- $\mathbf{A}\mathcal{K}_\ell(\mathbf{A}, \mathbf{v}) \subseteq \mathcal{K}_{\ell+1}(\mathbf{A}, \mathbf{v}), \quad \forall \ell$
- $\dim\left(\mathcal{K}_\ell(\mathbf{A}, \mathbf{v})\right) = \min(\ell,$ min. degree of non-zero poly. $\mathcal{P}$ such that $\mathcal{P}(\mathbf{A})\mathbf{v} = 0)$
- The sequence $\left(\mathcal{K}_\ell(\mathbf{A}, \mathbf{v})\right)_\ell$ is strictly increasing from 1 to $\ell_{\max}$, then it is constant starting from $\ell_{\max}$, where $\ell_{\max} := \operatorname{argmax}_\ell(\dim \mathcal{K}_\ell(\mathbf{A}, \mathbf{v}))$.

Example

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

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

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

Recap:

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

Property: $\boxed{\mathbf{x}^{(\ell+1)} - \mathbf{x}^{(0)} \;\in\; \mathcal{K}_{\ell+1}\big(\mathbf{A}, \mathbf{r}^{(0)}\big)}$

Proof

One has:
$$\begin{aligned}\mathbf{x}^{(\ell+1)} &= \mathbf{x}^{(\ell)} + \alpha^{(\ell)}\mathbf{p}^{(\ell)} \\ &= \mathbf{x}^{(0)} + \textstyle\sum_{i=0}^{\ell} \alpha^{(i)}\mathbf{p}^{(i)}\end{aligned}$$

Initially, one has: $\mathbf{p}^{(0)} = \mathbf{r}^{(0)} \;\Rightarrow\; \mathbf{p}^{(0)} \in \mathcal{K}_1\big(\mathbf{A}, \mathbf{r}^{(0)}\big)$.

At iteration ℓ, one has:
$$\begin{aligned}\mathbf{p}^{(\ell+1)} &= \mathbf{r}^{(\ell+1)} + \beta^{(\ell)}\mathbf{p}^{(\ell)} \\ &= \mathbf{r}^{(\ell)} - \alpha^{(\ell)}\mathbf{A}\mathbf{p}^{(\ell)} + \beta^{(\ell)}\mathbf{p}^{(\ell)} \\ &= \mathbf{r}^{(0)} - \sum_{i=0}^{\ell} \alpha^{(i)}\mathbf{A}\mathbf{p}^{(i)} + \beta^{(\ell)}\mathbf{p}^{(\ell)}\end{aligned}$$

If $\mathbf{p}^{(i)} \in \mathcal{K}_{i+1}\big(\mathbf{A}, \mathbf{r}^{(0)}\big)$ for $\forall i < \ell$, then $\mathbf{p}^{(\ell+1)} \in \mathcal{K}_{\ell+2}\big(\mathbf{A}, \mathbf{r}^{(0)}\big)$.

Then: $\mathbf{x}^{(\ell+1)} - \mathbf{x}^{(0)} = \sum_{i=0}^{\ell} \alpha^{(i)}\mathbf{p}^{(i)} \;\in\; \mathcal{K}_{\ell+1}\big(\mathbf{A}, \mathbf{r}^{(0)}\big)$ $\square$

Properties of the CG method *(continuation)*

Property: $\boxed{\mathbf{x}^{(\ell)} - \mathbf{x}^{(0)} \in \mathcal{K}_\ell\big(\mathbf{A}, \mathbf{r}^{(0)}\big)}$

Property: $\boxed{\mathbf{x}^{(\ell)} - \mathbf{x}^{(0)} = \underset{\mathbf{y} \in \mathcal{K}_\ell(\mathbf{A}, \mathbf{r}^{(0)})}{\arg\min} J\left(\mathbf{x}^{(0)} + \mathbf{y}\right)}$

At each iteration ℓ, one has the best solution $\mathbf{x}^{(\ell)}$ in the sense *"J is minimum"* such that $\mathbf{x}^{(\ell)} - \mathbf{x}^{(0)}$ belongs to the subspace $\mathcal{K}_\ell\left(\mathbf{A}, \mathbf{r}^{(0)}\right)$.

Unfortunately, the CG method is limited to SPD/HPD matrices.

Toward general iterative methods for general matrices . . .

We seek for a method that gives . . .

$\boxed{\mathbf{x}^{(\ell)} - \mathbf{x}^{(0)} = \mathcal{P}_{\ell-1}(\mathbf{A})\, \mathbf{r}^{(0)}}$ where $\mathcal{P}_{\ell-1}(\,\cdot\,)$ is a polynomial of degree $\ell - 1$

such that $\mathbf{x}^{(\ell)}$ is the *"best solution"* with $\mathbf{x}^{(\ell)} - \mathbf{x}^{(0)} \in \mathcal{K}_\ell(\mathbf{A}, \mathbf{r}^{(0)})$.

The GMRES *(generalized minimal residual)* method is a Krylov method based on the minimization of the residual at each iteration. *(There are other Krylov methods.)*

Nonstationary iterative methods for linear systems

Conjugate gradient method

Interlude on Krylov spaces

Few words about GMRES

GMRES — Principle

The GMRES *(Generalized Minimal Residual)* method relies on 2 steps performed at each iteration ℓ:

1. Building an orthonormal basis $\{\mathbf{v}^{(1)}, \dots, \mathbf{v}^{(\ell)}\}$ for the Krylov subspace $\mathcal{K}_\ell(\mathbf{A}, \mathbf{r}^{(0)})$

$$\boxed{\mathcal{K}_\ell\big(\mathbf{A}, \mathbf{r}^{(0)}\big) \;:=\; \mathrm{span}\big(\mathbf{r}^{(0)}, \mathbf{A}\mathbf{r}^{(0)}, \dots, \mathbf{A}^{\ell-1}\mathbf{r}^{(0)}\big) \;=\; \mathrm{span}\big(\mathbf{v}^{(1)}, \dots, \mathbf{v}^{(\ell)}\big)}$$

   With GMRES:
   - The basis vectors are built by using Arnoldi iterations.
   - Only the additional vector $\mathbf{v}^{(\ell)}$ is computed at iteration ℓ.

2. Solving a minimization problem to get the *"best solution"*

$$\boxed{\mathbf{x}^{(\ell)} - \mathbf{x}^{(0)} = \underset{\mathbf{y} \in \mathcal{K}_\ell(\mathbf{A}, \mathbf{r}^{(0)})}{\arg\min} \|\mathbf{b} - \mathbf{A}(\mathbf{x}^{(0)} + \mathbf{y})\|_2}$$

   The solution $\mathbf{x}^{(\ell)}$ is such that:
   - The update belongs to the Krylov subspace: $\mathbf{x}^{(\ell)} - \mathbf{x}^{(0)} \in \mathcal{K}_\ell\big(\mathbf{A}, \mathbf{r}^{(0)}\big)$
   - The 2-norm of the residual is minimum: $\|\mathbf{r}^{(\ell)}\|_2$ is minimum

   With GMRES:
   - Solving this problem is equivalent to solving a least square problem.
   - The least square problem can be solved with a QR factorization.
   - The QR factorization can be computed rapidly thanks to Givens matrices.

GMRES — Algorithm

GMRES algorithm *(main steps)*

$\mathbf{x}^{(0)} \in \mathbb{R}^n$

$\mathbf{r}^{(0)} = \mathbf{b} - \mathbf{A}\mathbf{x}^0$

$\mathbf{v}^{(1)} = \mathbf{r}^{(0)} / \|\mathbf{r}^{(0)}\|$

for $\ell = 1, 2 \ \ldots$ **do**

  // Building the orthonormal basis

  $\mathbf{w}^{(\ell)} = \mathbf{A}\mathbf{v}^{(\ell)}$

  for $i = 1, \ \ldots \ , \ell$ **do**

    $\mathbf{w}^{(\ell)} = \mathbf{w}^{(\ell)} - (\mathbf{w}^{(\ell)}, \mathbf{v}^{(i)})\mathbf{v}^{(i)}$

  end

  If $\|\mathbf{w}^{(\ell)}\|_2 = 0 \rightarrow$ Stop.

  $\mathbf{v}^{(\ell+1)} = \mathbf{w}^{(\ell)} / \|\mathbf{w}^{(\ell)}\|_2$

  // Solving the minimization problem

  $\mathbf{z}^{(\ell)} = \arg\min_{\mathbf{z} \in \mathbb{R}^{\ell+1}} \left\|\mathbf{b} - \mathbf{A}\mathbf{V}^{(\ell)}\mathbf{z}\right\|_2$

  $\mathbf{x}^{(\ell)} = \mathbf{x}^{(0)} + \mathbf{V}^{(\ell)}\mathbf{z}^{(\ell)}$

end

$\mathbf{V}^{(\ell)}$ is a $\ell \times (\ell + 1)$ matrix which the columns are the basis vectors $\left\{\mathbf{v}^{(i)}\right\}_{i=1\ldots\ell+1}$.

GMRES — Discussion

Theoretical aspects

- Method for general nonsingular matrices
- By construction, convergence with maximum n iterations. *(with ∞ accuracy)*
- If less iterations are required, procedure stopped during the construction of the basis. $\rightarrow$ Breakdown

Algorithmic aspects

- **The computational cost increases $\mathcal{O}(\ell^2)$ at each iteration.**
  - Storage of an additional basis vector and larger matrices
  - Orthogonalization by an additional basis vector
  - Solution of a larger minimization problem

  To reduce the cost, the process can be restarted by using the current solution as initial solution. $\rightarrow$ Restarted GMRES
- **The algorithm is easy to parallelize.**
  - Basic linear algebra operations (BLAS 1 and 2) $\Rightarrow$ Easy for parallel computing
  - Computation of scalar product and norms $\Rightarrow$ Collective communications

Standard approach for nonsymmetric matrices. Widely used!
Need to limit the number of iterations $\Rightarrow$ ***Preconditioning . . .***

Summary

- **Stationary methods** $\left(\mathbf{M}\mathbf{x}^{(k+1)} = \mathbf{N}\mathbf{x}^{(k)} + \mathbf{b}\right)$
  - 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 $\rightarrow$ *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

Resources

- *Numerical Linear Algebra*
  L.N. Trefethen, D. Bau III (1997), SIAM
- *Iterative Methods for Sparse Linear Systems, 2nd edition*
  Y. Saad (2003), SIAM
- *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]