Preconditioned GMRES Revisited

Roland Herzog Kirk Soodhalter

TECHNISCHE UNIVERSITÄT CHEMNITZ

UBC (visiting)

RICAM
JOHANN·RADON·INSTITUTE
FOR COMPUTATIONAL AND APPLIED MATHEMATICS
RICAM Linz

Preconditioning Conference 2017
Vancouver

August 01, 2017

Table of Contents

What is GMRES?

Quoting Saad and Schultz (1986):

> *We present an iterative method for solving linear systems, which has the property of minimizing at every step the norm of the residual vector over a Krylov subspace. The algorithm is derived from the Arnoldi process for constructing an ℓ_2-orthogonal basis of Krylov subspaces.*

Quoting Wikipedia:

> *In mathematics, the generalized minimal residual method (usually abbreviated GMRES) is an iterative method for the numerical solution of a nonsymmetric system of linear equations. The method approximates the solution by the vector in a Krylov subspace with minimal residual. The Arnoldi iteration is used to find this vector.*

Some References (Very Incomplete)

contributing to the understanding of GMRES in various situations:

- Starke (1997)
- Chan, Chow, Saad and Yeung (1998)
- Chen, Kincaid and Young (1999)
- Klawonn (1998)
- Ernst (2000); Eiermann and Ernst (2001)
- Sarkis and Szyld (2007)
- Pestana and Wathen (2013)

Some References (Very Incomplete)

contributing to the understanding of GMRES in various situations:

- Starke (1997)
- Chan, Chow, Saad and Yeung (1998)
- Chen, Kincaid and Young (1999)
- Klawonn (1998)
- Ernst (2000); Eiermann and Ernst (2001)
- Sarkis and Szyld (2007)
- Pestana and Wathen (2013)

Some References (Very Incomplete)

contributing to the understanding of GMRES in various situations:

- Starke (1997)
- Chan, Chow, Saad and Yeung (1998)
- Chen, Kincaid and Young (1999)
- Klawonn (1998)
- Ernst (2000); Eiermann and Ernst (2001)
- Sarkis and Szyld (2007)
- Pestana and Wathen (2013)

and in Hilbert space (but with $A \in \mathcal{L}(X)$) in particular:

- Campbell, Ipsen, Kelley and Meyer (1996)
- Moret (1997)
- Calvetti, Lewis and Reichel (2002)
- Gasparo, Papini and Pasquali (2008)

Purpose/Plan of This Talk

- re-derive GMRES with natural algorithmic ingredients
- draw inspirations from a Hilbert space setting (PDE problems)
- obtain 'canonical GMRES'

- locate GMRES variants in the literature in this framework
- to sort out my own personal lack of knowledge/confusion about these variants

- Hilbert space analysis matters
  ($\rightarrow$ recall talk by W. Zulehner for instance)
- even though in this talk everything is finite dimensional

Problem Setting

$Ax = b$ in $\mathbb{R}^n$

$A \in \mathbb{R}^{n \times n}$ (non-singular)

$Ax = b$ in X^*

$A \in \mathcal{L}(X, X^*)$ (non-singular)

X Hilbert space

inner product $(\cdot, \cdot)_M$

Problem Setting

$Ax = b$ in $\mathbb{R}^n$

$A \in \mathbb{R}^{n \times n}$ (non-singular)

$\mathbb{R}^n$ also a Hilbert space

inner product $(\cdot, \cdot)_M$

$(u, v)_M = u^\top M v$

$Ax = b$ in X^*

$A \in \mathcal{L}(X, X^*)$ (non-singular)

X Hilbert space

inner product $(\cdot, \cdot)_M$

Problem Setting

$A\,x = b$ in $\mathbb{R}^n$

$A \in \mathbb{R}^{n\times n}$ (non-singular)

$\mathbb{R}^n$ also a Hilbert space

inner product $(\cdot,\cdot)_M$

$(u,v)_M = u^\top M\,v$

$A\,x = b$ in X^*

$A \in \mathcal{L}(X, X^*)$ (non-singular)

X Hilbert space

inner product $(\cdot,\cdot)_M$

$(u,v)_M = \langle u,\, M\,v\rangle_{X,X^*}$

$$\big(X; \|\cdot\|_M\big) \underset{M}{\overset{M^{-1}}{\rightleftarrows}} \big(X^*; \|\cdot\|_{M^{-1}}\big)$$

$M \in \mathcal{L}(X, X^*)$ Riesz isomorphism

Problem Setting

$Ax = b$ in $\mathbb{R}^n$

$A \in \mathbb{R}^{n \times n}$ (non-singular)

$\mathbb{R}^n$ also a Hilbert space

inner product $(\cdot,\cdot)_M$

$(u,v)_M = u^\top M\, v$

$$\left(\mathbb{R}^n; \|\cdot\|_M\right) \underset{M}{\overset{M^{-1}}{\rightleftharpoons}} \left(\mathbb{R}^n; \|\cdot\|_{M^{-1}}\right)$$

$Ax = b$ in X^*

$A \in \mathcal{L}(X, X^*)$ (non-singular)

X Hilbert space

inner product $(\cdot,\cdot)_M$

$(u,v)_M = \langle u,\, M\, v\rangle_{X,X^*}$

$$\left(X; \|\cdot\|_M\right) \underset{M}{\overset{M^{-1}}{\rightleftharpoons}} \left(X^*; \|\cdot\|_{M^{-1}}\right)$$

$M \in \mathcal{L}(X, X^*)$ Riesz isomorphism

Problem Setting

$A\,x = b$ in $\mathbb{R}^n$

$A \in \mathbb{R}^{n\times n}$ (non-singular)

$\mathbb{R}^n$ also a Hilbert space

inner product $(\cdot,\cdot)_M$

$(u,v)_M = u^\top M\,v$

$A\,x = b$ in X^*

$A \in \mathcal{L}(X, X^*)$ (non-singular)

X Hilbert space

inner product $(\cdot,\cdot)_M$

$(u,v)_M = \langle u,\ M\,v\rangle_{X,X^*}$

$$(\mathbb{R}^n; \|\cdot\|_M) \xrightleftharpoons[M]{M^{-1}} (\mathbb{R}^n; \|\cdot\|_{M^{-1}}) \qquad (X; \|\cdot\|_M) \xrightleftharpoons[M]{M^{-1}} ({\color{red}X^*}; \|\cdot\|_{M^{-1}})$$

$$\left.\begin{array}{l} \text{The inner product } (\cdot,\cdot)_M \text{ in } X, \\ \text{the inner product } (\cdot,\cdot)_{M^{-1}} \text{ in } X^*, \\ \text{the Riesz map } M \in \mathcal{L}(X, X^*), \\ \text{and its inverse } M^{-1} \in \mathcal{L}(X^*, X) \end{array}\right\} \quad \text{uniquely define each other.}$$

Problem Setting

(primal) quantities in X — iterates, errors
(dual) quantities in X^* — rhs, residuals

$A\,x = b$ in $\mathbb{R}^n$
$A \in \mathbb{R}^{n\times n}$ (non-singular)
$\mathbb{R}^n$ also a Hilbert space
inner product $(\cdot,\cdot)_M$
$(u,v)_M = u^\top M\,v$

$A\,x = b$ in X^*
$A \in \mathcal{L}(X, X^*)$ (non-singular)
X Hilbert space
inner product $(\cdot,\cdot)_M$
$(u,v)_M = \langle u,\, M\,v\rangle_{X,X^*}$

$$(\mathbb{R}^n; \|\cdot\|_M) \underset{M}{\overset{M^{-1}}{\rightleftarrows}} (\mathbb{R}^n; \|\cdot\|_{M^{-1}}) \qquad (X; \|\cdot\|_M) \underset{M}{\overset{M^{-1}}{\rightleftarrows}} (X^*; \|\cdot\|_{M^{-1}})$$

$$\left.\begin{array}{l}\text{The inner product } (\cdot,\cdot)_M \text{ in } X,\\ \text{the inner product } (\cdot,\cdot)_{M^{-1}} \text{ in } X^*,\\ \text{the Riesz map } M \in \mathcal{L}(X, X^*),\\ \text{and its inverse } M^{-1} \in \mathcal{L}(X^*, X)\end{array}\right\} \quad \text{uniquely define each other.}$$

Example: Stokes (Saddle-Point Problem)

$$\begin{bmatrix} A & B^\star \\ B & 0 \end{bmatrix} \begin{pmatrix} \boldsymbol{u} \\ p \end{pmatrix} = \begin{pmatrix} \boldsymbol{f} \\ 0 \end{pmatrix} \begin{matrix} \in V^* \\ \in Q^* \end{matrix}$$

$$V = H_0^1(\Omega; \mathbb{R}^3), \quad Q = L_0^2(\Omega)$$

$$A\,\boldsymbol{u} = a(\boldsymbol{u}, \cdot) \in V^*$$

$$B\,\boldsymbol{u} = b(\boldsymbol{u}, \cdot) \in Q^*$$

$$B^\star p = b(\cdot, p) \in V^*$$

$$\underbrace{\int_\Omega \mu \nabla \boldsymbol{u} : \nabla \boldsymbol{v} \,\mathrm{d}x}_{a(\boldsymbol{u},\boldsymbol{v})} \qquad \underbrace{-\int_\Omega p \operatorname{div} \boldsymbol{v} \,\mathrm{d}x}_{b(\boldsymbol{v},p)} = \underbrace{\int_\Omega \boldsymbol{F} \cdot \boldsymbol{v} \,\mathrm{d}x}_{\langle \boldsymbol{f}, \boldsymbol{v} \rangle}$$

$$\underbrace{-\int_\Omega \operatorname{div} \boldsymbol{u}\, q \,\mathrm{d}x}_{b(\boldsymbol{u},q)} = 0$$

Example: Oseen (Saddle-Point Problem)

$$\begin{bmatrix} A+N & B^\star \\ B & 0 \end{bmatrix} \begin{pmatrix} \boldsymbol{u} \\ p \end{pmatrix} = \begin{pmatrix} \boldsymbol{f} \\ 0 \end{pmatrix} \begin{matrix} \in V^* \\ \in Q^* \end{matrix}$$

$$V = H_0^1(\Omega;\mathbb{R}^3),\ Q = L_0^2(\Omega)$$

$$\begin{aligned} A\,\boldsymbol{u} &= a(\boldsymbol{u},\cdot) \in V^* \\ N\,\boldsymbol{u} &= n(\boldsymbol{u},\cdot) \in V^* \\ B\,\boldsymbol{u} &= b(\boldsymbol{u},\cdot) \in Q^* \\ B^\star p &= b(\cdot,p) \in V^* \end{aligned}$$

$$\underbrace{\int_\Omega \mu\,\nabla\boldsymbol{u} : \nabla\boldsymbol{v}\,\mathrm{dx}}_{a(\boldsymbol{u},\boldsymbol{v})} + \underbrace{\int_\Omega (\nabla\boldsymbol{u}\cdot\boldsymbol{U})\cdot\boldsymbol{v}\,\mathrm{dx}}_{n(\boldsymbol{u},\boldsymbol{v})} \underbrace{- \int_\Omega p\,\operatorname{div}\boldsymbol{v}\,\mathrm{dx}}_{b(\boldsymbol{v},p)} = \underbrace{\int_\Omega \boldsymbol{F}\cdot\boldsymbol{v}\,\mathrm{dx}}_{\langle\boldsymbol{f},\boldsymbol{v}\rangle}$$

$$\underbrace{-\int_\Omega \operatorname{div}\boldsymbol{u}\,q\,\mathrm{dx}}_{b(\boldsymbol{u},q)} = 0$$

Ingredients of 'Canonical GMRES'

The residual is always going to be $r := b - Ax \in X^*$.

1. **Krylov subspaces**

   Due to $A \in \mathcal{L}(X, X^*)$ we cannot form

   $$K_k(A \quad ; r) := \operatorname{span}\{r,\ Ar \quad ,\ A^2 r \quad ,\ \ldots,\ A^{k-1} r \quad \}$$

The residual is always going to be $r := b - Ax \in X^*$.

1. **Krylov subspaces**

   Due to $A \in \mathcal{L}(X, X^*)$ we have to use

   $$K_k(AP^{-1}; r) := \operatorname{span}\left\{r,\ AP^{-1}r,\ (AP^{-1})^2 r,\ \ldots,\ (AP^{-1})^{k-1} r\right\} \subset X^*$$

   where $P \in \mathcal{L}(X, X^*)$ non-singular is a 'preconditioner' (required!).

Ingredients of 'Canonical GMRES'

The residual is always going to be $r := b - Ax \in X^*$.

1. **Krylov subspaces**

   Due to $A \in \mathcal{L}(X, X^*)$ we have to use

   $$K_k(AP^{-1}; r) := \operatorname{span}\{r,\ AP^{-1}r,\ (AP^{-1})^2 r,\ \ldots,\ (AP^{-1})^{k-1} r\} \subset X^*$$

   where $P \in \mathcal{L}(X, X^*)$ non-singular is a 'preconditioner' (required!).

2. **Inner product** $(\cdot,\cdot)_W$ in X for Arnoldi

   Since $K_k(AP^{-1}; r) \subset X^*$ holds, orthonormality is defined w.r.t. W^{-1}.

Ingredients of 'Canonical GMRES'

The residual is always going to be $r := b - Ax \ \in X^*$.

1. **Krylov subspaces**

   Due to $A \in \mathcal{L}(X, X^*)$ we have to use

   $$K_k(AP^{-1}; r) := \operatorname{span}\big\{r,\ AP^{-1}r,\ (AP^{-1})^2 r,\ \ldots,\ (AP^{-1})^{k-1} r\big\} \subset X^*$$

   where $P \in \mathcal{L}(X, X^*)$ non-singular is a 'preconditioner' (required!).

2. **Inner product** $(\cdot,\cdot)_W$ in X for Arnoldi

   Since $K_k(AP^{-1}; r) \subset X^*$ holds, orthonormality is defined w.r.t. W^{-1}.

3. **Inner product** $(\cdot,\cdot)_M$ in X for residual minimization

   Since $r \in X^*$ holds, $\|r\|_{M^{-1}}$ will be the quantity minimized.

The Arnoldi Process

The Arnoldi process (here displayed with modified GS) generates an orthonormal basis of the Krylov subspace $\mathcal{K}_k(A\ ;r_0)$:

Input: Matrix A (or matrix-vector products), initial vector r_0
Output: ONB $v_1, v_2, \ldots$ with $\operatorname{span}\{v_1, \ldots, v_k\} = \mathcal{K}_k(A\ ;r_0)$

Set $v_1 := \dfrac{r_0}{\langle r_0,\ r_0\rangle^{1/2}}$
for $k = 1, 2, \ldots$ **do**
    Set $v_{k+1} := A\, v_k$ // new Krylov vector
    for $j = 1, 2, \ldots, k$ **do**
        Set $h_{j,k} := (v_{k+1},\ v_j)$ // orthogonal. coefficients
        Set $v_{k+1} := v_{k+1} - h_{j,k}\, v_j$
    end for
    Set $h_{k+1,k} := (v_{k+1},\ v_{k+1})^{1/2}$ // $= (v_{k+1},\ v_{k+1})^{1/2}$
    Set $v_{k+1} := \dfrac{v_{k+1}}{h_{k+1,k}}$
end for

The Arnoldi Process

The Arnoldi process (here displayed with modified GS) generates a W^{-1}-orthonormal basis of the Krylov subspace $\mathcal{K}_k(AP^{-1}; r_0)$:

Input: Matrix A, P^{-1}, W^{-1} (or matrix-vector products), initial vector $r_0 \in X^*$
Output: W^{-1}-ONB $v_1, v_2, \ldots \in X^*$ with $\operatorname{span}\{v_1, \ldots, v_k\} = \mathcal{K}_k(AP^{-1}; r_0)$

Set $v_1 := \dfrac{r_0}{\langle r_0,\ r_0\rangle^{1/2}}$
for $k = 1, 2, \ldots$ **do**
    Set $v_{k+1} := A\, v_k$ // new Krylov vector
    for $j = 1, 2, \ldots, k$ **do**
        Set $h_{j,k} := (v_{k+1},\ v_j)$ // orthogonal. coefficients
        Set $v_{k+1} := v_{k+1} - h_{j,k}\, v_j$
    end for
    Set $h_{k+1,k} := (v_{k+1},\ v_{k+1})^{1/2}$ // $= (v_{k+1},\ v_{k+1})^{1/2}$
    Set $v_{k+1} := \dfrac{v_{k+1}}{h_{k+1,k}}$
end for

The Arnoldi Process

The Arnoldi process (here displayed with modified GS) generates a W^{-1}-orthonormal basis of the Krylov subspace $\mathcal{K}_k(AP^{-1}; r_0)$:

Input: Matrix A, P^{-1}, W^{-1} (or matrix-vector products), initial vector $r_0 \in X^*$
Output: W^{-1}-ONB $v_1, v_2, \ldots \in X^*$ with $\operatorname{span}\{v_1, \ldots, v_k\} = \mathcal{K}_k(AP^{-1}; r_0)$

1: Set $z_1 := W^{-1} r_0$ and $v_1 := \dfrac{r_0}{\langle r_0, z_1 \rangle^{1/2}}$ and $z_1 := \dfrac{z_1}{\langle r_0, z_1 \rangle^{1/2}}$
2: **for** $k = 1, 2, \ldots$ **do**
3: Set $v_{k+1} := A\, v_k$ // new Krylov vector
4: **for** $j = 1, 2, \ldots, k$ **do**
5: Set $h_{j,k} := (v_{k+1}, v_j)$ // orthogonal. coefficients
6: Set $v_{k+1} := v_{k+1} - h_{j,k}\, v_j$
7: **end for**
8: Set $h_{k+1,k} := (v_{k+1}, v_{k+1})^{1/2}$ // $= (v_{k+1}, v_{k+1})^{1/2}$
9: Set $v_{k+1} := \dfrac{v_{k+1}}{h_{k+1,k}}$
10: **end for**

Both v_j and $z_j = W^{-1} v_j$ need to be stored due to MGS.

The Arnoldi Process

The Arnoldi process (here displayed with modified GS) generates a W^{-1}-orthonormal basis of the Krylov subspace $\mathcal{K}_k(AP^{-1}; r_0)$:

Input: Matrix A, P^{-1}, W^{-1} (or matrix-vector products), initial vector $r_0 \in X^*$
Output: W^{-1}-ONB $v_1, v_2, \ldots \in X^*$ with $\operatorname{span}\{v_1, \ldots, v_k\} = \mathcal{K}_k(AP^{-1}; r_0)$

```
Set z_1 := W^{-1} r_0 and v_1 := r_0 / ⟨r_0, z_1⟩^{1/2} and z_1 := z_1 / ⟨r_0, z_1⟩^{1/2}
for k = 1, 2, ... do
    Set v_{k+1} := A P^{-1} v_k                                   // new Krylov vector
    for j = 1, 2, ..., k do
        Set h_{j,k} := (v_{k+1}, v_j)                             // orthogonal. coefficients
        Set v_{k+1} := v_{k+1} − h_{j,k} v_j
    end for
    Set h_{k+1,k} := (v_{k+1}, v_{k+1})^{1/2}                     // = (v_{k+1}, v_{k+1})^{1/2}
    Set v_{k+1} := v_{k+1} / h_{k+1,k}
end for
```

Both v_j and $z_j = W^{-1} v_j$ need to be stored due to MGS.

The Arnoldi Process

The Arnoldi process (here displayed with modified GS) generates a W^{-1}-orthonormal basis of the Krylov subspace $\mathcal{K}_k(AP^{-1}; r_0)$:

Input: Matrix A, P^{-1}, W^{-1} (or matrix-vector products), initial vector $r_0 \in X^*$
Output: W^{-1}-ONB $v_1, v_2, \ldots \in X^*$ with $\operatorname{span}\{v_1, \ldots, v_k\} = \mathcal{K}_k(AP^{-1}; r_0)$

Set $z_1 := W^{-1} r_0$ and $v_1 := \dfrac{r_0}{\langle r_0, z_1 \rangle^{1/2}}$ and $z_1 := \dfrac{z_1}{\langle r_0, z_1 \rangle^{1/2}}$
for $k = 1, 2, \ldots$ **do**
    Set $v_{k+1} := AP^{-1} v_k$ // new Krylov vector
    for $j = 1, 2, \ldots, k$ **do**
        Set $h_{j,k} := \langle v_{k+1}, z_j \rangle$ // orthogonal. coefficients
        Set $v_{k+1} := v_{k+1} - h_{j,k}\, v_j$
    end for
    Set $h_{k+1,k} := (v_{k+1}, v_{k+1})^{1/2}$ // $= (v_{k+1}, v_{k+1})^{1/2}$
    Set $v_{k+1} := \dfrac{v_{k+1}}{h_{k+1,k}}$
end for

Both v_j and $z_j = W^{-1} v_j$ need to be stored due to MGS.

The Arnoldi Process

The Arnoldi process (here displayed with modified GS) generates a W^{-1}-orthonormal basis of the Krylov subspace $\mathcal{K}_k(AP^{-1}; r_0)$:

Input: Matrix A, P^{-1}, W^{-1} (or matrix-vector products), initial vector $r_0 \in X^*$
Output: W^{-1}-ONB $v_1, v_2, \ldots \in X^*$ with $\operatorname{span}\{v_1, \ldots, v_k\} = \mathcal{K}_k(AP^{-1}; r_0)$

Set $z_1 := W^{-1} r_0$ and $v_1 := \dfrac{r_0}{\langle r_0,\, z_1\rangle^{1/2}}$ and $z_1 := \dfrac{z_1}{\langle r_0,\, z_1\rangle^{1/2}}$
for $k = 1, 2, \ldots$ **do**
    Set $v_{k+1} := AP^{-1} v_k$ // new Krylov vector
    for $j = 1, 2, \ldots, k$ **do**
        Set $h_{j,k} := \langle v_{k+1},\, z_j\rangle$ // orthogonal. coefficients
        Set $v_{k+1} := v_{k+1} - h_{j,k}\, v_j$
    end for
    Set $h_{k+1,k} := (v_{k+1},\, v_{k+1})^{1/2}$ // $= (v_{k+1},\, v_{k+1})^{1/2}$
    Set $v_{k+1} := \dfrac{v_{k+1}}{h_{k+1,k}}$
end for

Both v_j and $z_j = W^{-1} v_j$ need to be stored due to MGS.

The Arnoldi Process

The Arnoldi process (here displayed with modified GS) generates a W^{-1}-orthonormal basis of the Krylov subspace $\mathcal{K}_k(AP^{-1}; r_0)$:

Input: Matrix A, P^{-1}, W^{-1} (or matrix-vector products), initial vector $r_0 \in X^*$
Output: W^{-1}-ONB $v_1, v_2, \ldots \in X^*$ with $\operatorname{span}\{v_1, \ldots, v_k\} = \mathcal{K}_k(AP^{-1}; r_0)$

1: Set $z_1 := W^{-1} r_0$ and $v_1 := \dfrac{r_0}{\langle r_0, z_1 \rangle^{1/2}}$ and $z_1 := \dfrac{z_1}{\langle r_0, z_1 \rangle^{1/2}}$
2: **for** $k = 1, 2, \ldots$ **do**
3: Set $v_{k+1} := AP^{-1} v_k$ // new Krylov vector
4: **for** $j = 1, 2, \ldots, k$ **do**
5: Set $h_{j,k} := \langle v_{k+1}, z_j \rangle$ // orthogonal. coefficients
6: Set $v_{k+1} := v_{k+1} - h_{j,k}\, v_j$
7: **end for**
8: Set $z_{k+1} := W^{-1} v_{k+1}$ and $h_{k+1,k} := \langle v_{k+1}, z_{k+1} \rangle^{1/2}$ // $= (v_{k+1}, v_{k+1})_{W^{-1}}^{1/2}$
9: Set $v_{k+1} := \dfrac{v_{k+1}}{h_{k+1,k}}$
10: **end for**

Both v_j and $z_j = W^{-1} v_j$ need to be stored due to MGS.

The Arnoldi Process

The Arnoldi process (here displayed with modified GS) generates a W^{-1}-orthonormal basis of the Krylov subspace $\mathcal{K}_k(AP^{-1}; r_0)$:

Input: Matrix A, P^{-1}, W^{-1} (or matrix-vector products), initial vector $r_0 \in X^*$
Output: W^{-1}-ONB $v_1, v_2, \ldots \in X^*$ with $\operatorname{span}\{v_1, \ldots, v_k\} = \mathcal{K}_k(AP^{-1}; r_0)$

```
Set $z_1 := W^{-1} r_0$ and $v_1 := \dfrac{r_0}{\langle r_0,\, z_1\rangle^{1/2}}$ and $z_1 := \dfrac{z_1}{\langle r_0,\, z_1\rangle^{1/2}}$
for $k = 1, 2, \ldots$ do
    Set $v_{k+1} := AP^{-1} v_k$                                  // new Krylov vector
    for $j = 1, 2, \ldots, k$ do
        Set $h_{j,k} := \langle v_{k+1},\, z_j\rangle$             // orthogonal. coefficients
        Set $v_{k+1} := v_{k+1} - h_{j,k}\, v_j$
    end for
    Set $z_{k+1} := W^{-1} v_{k+1}$ and $h_{k+1,k} := \langle v_{k+1},\, z_{k+1}\rangle^{1/2}$ // $= (v_{k+1},\, v_{k+1})_{W^{-1}}^{1/2}$
    Set $v_{k+1} := \dfrac{v_{k+1}}{h_{k+1,k}}$ and $z_{k+1} := \dfrac{z_{k+1}}{h_{k+1,k}}$
end for
```

Both v_j and $z_j = W^{-1} v_j$ need to be stored due to MGS.

In matrix form the Arnoldi process reads

$$h_{2,1} \begin{bmatrix} | & & \\ v_2 & & \\ | & & \end{bmatrix} = AP^{-1} \begin{bmatrix} | & & \\ v_1 & & \\ | & & \end{bmatrix} - \begin{bmatrix} | & & \\ v_1 & & \\ | & & \end{bmatrix} h_{1,1}$$

where H is the upper Hessenberg matrix

$$H = \begin{bmatrix} (v_1, AP^{-1}v_1)_{W^{-1}} & (v_1, AP^{-1}v_2)_{W^{-1}} & \cdots \\ (v_2, AP^{-1}v_1)_{W^{-1}} & (v_2, AP^{-1}v_2)_{W^{-1}} & \cdots \\ 0 & (v_3, AP^{-1}v_2)_{W^{-1}} & \cdots \\ \vdots & 0 & \\ \vdots & \vdots & \end{bmatrix} \in \mathbb{R}^{(k+1)\times k}.$$

In matrix form the Arnoldi process reads

$$h_{3,2} \begin{bmatrix} | & | & \\ 0 & v_3 & \\ | & | & \end{bmatrix} = AP^{-1} \begin{bmatrix} | & | & \\ v_1 & v_2 & \\ | & | & \end{bmatrix} - \begin{bmatrix} | & | & \\ v_1 & v_2 & \\ | & | & \end{bmatrix} H_{2,2}$$

where H is the upper Hessenberg matrix

$$H = \begin{bmatrix} (v_1, AP^{-1}v_1)_{W^{-1}} & (v_1, AP^{-1}v_2)_{W^{-1}} & \cdots \\ (v_2, AP^{-1}v_1)_{W^{-1}} & (v_2, AP^{-1}v_2)_{W^{-1}} & \cdots \\ 0 & (v_3, AP^{-1}v_2)_{W^{-1}} & \cdots \\ \vdots & 0 & \\ \vdots & \vdots & \end{bmatrix} \in \mathbb{R}^{(k+1)\times k}.$$

In matrix form the Arnoldi process reads

$$h_{k+1,k}\begin{bmatrix} | & & | \\ 0 & \cdots & v_{k+1} \\ | & & | \end{bmatrix} = AP^{-1}\begin{bmatrix} | & & | \\ v_1 & \cdots & v_k \\ | & & | \end{bmatrix} - \begin{bmatrix} | & & | \\ v_1 & \cdots & v_k \\ | & & | \end{bmatrix} H_{k,k}$$

where H is the upper Hessenberg matrix

$$H = \begin{bmatrix} (v_1, AP^{-1}v_1)_{W^{-1}} & (v_1, AP^{-1}v_2)_{W^{-1}} & \cdots \\ (v_2, AP^{-1}v_1)_{W^{-1}} & (v_2, AP^{-1}v_2)_{W^{-1}} & \cdots \\ 0 & (v_3, AP^{-1}v_2)_{W^{-1}} & \cdots \\ \vdots & 0 & \\ \vdots & \vdots & \end{bmatrix} \in \mathbb{R}^{(k+1)\times k}.$$

The Arnoldi Process

In matrix form the Arnoldi process reads

$$h_{k+1,k}\begin{bmatrix} | & & | \\ 0 & \cdots & v_{k+1} \\ | & & | \end{bmatrix} = AP^{-1}\begin{bmatrix} | & & | \\ v_1 & \cdots & v_k \\ | & & | \end{bmatrix} - \begin{bmatrix} | & & | \\ v_1 & \cdots & v_k \\ | & & | \end{bmatrix} H_{k,k}$$

$$\text{or} \quad V_{k+1}H = AP^{-1}V_k$$

where H is the upper Hessenberg matrix

$$H = \begin{bmatrix} (v_1, AP^{-1}v_1)_{W^{-1}} & (v_1, AP^{-1}v_2)_{W^{-1}} & \cdots \\ (v_2, AP^{-1}v_1)_{W^{-1}} & (v_2, AP^{-1}v_2)_{W^{-1}} & \cdots \\ 0 & (v_3, AP^{-1}v_2)_{W^{-1}} & \cdots \\ \vdots & 0 & \\ \vdots & \vdots & \end{bmatrix} \in \mathbb{R}^{(k+1)\times k}.$$

Residual Norm Minimization

$$\text{By design,} \quad x_k - x_0 \in \quad P^{-1} K_k(AP^{-1}; r_0)$$
$$\text{and therefore} \quad r_k - r_0 \in AP^{-1} K_k(AP^{-1}; r_0) = \operatorname{range} AP^{-1} V_k.$$

Hence

$$\begin{aligned}
\|r_k\|^2_{M^{-1}} &= \left\| r_0 - AP^{-1} V_k \vec{y} \right\|^2_{M^{-1}} && \vec{y} \in \mathbb{R}^k \\
&= \left\| r_0 - V_{k+1} H \vec{y} \right\|^2_{M^{-1}} && \text{due to } AP^{-1} V_k = V_{k+1} H \\
&= \left\| \|r_0\|_{W^{-1}} \, v_1 - V_{k+1} H \vec{y} \right\|^2_{M^{-1}} && \text{since } v_1 = r_0 / \|r_0\|_{W^{-1}} \\
&= \left\| V_{k+1} \left(\|r_0\|_{W^{-1}} \, \vec{e}_1 - H \vec{y} \right) \right\|^2_{M^{-1}}
\end{aligned}$$

Residual Norm Minimization

$$\text{By design,} \quad x_k - x_0 \in \quad P^{-1} K_k(AP^{-1}; r_0)$$
$$\text{and therefore} \quad r_k - r_0 \in AP^{-1} K_k(AP^{-1}; r_0) = \text{range}\, AP^{-1} V_k.$$

Hence

$$\begin{aligned}
\|r_k\|^2_{M^{-1}} &= \left\| r_0 - AP^{-1} V_k \vec{y} \right\|^2_{M^{-1}} && \vec{y} \in \mathbb{R}^k \\
&= \left\| r_0 - V_{k+1} H \vec{y} \right\|^2_{M^{-1}} && \text{due to } AP^{-1} V_k = V_{k+1} H \\
&= \left\| \|r_0\|_{W^{-1}}\, v_1 - V_{k+1} H \vec{y} \right\|^2_{M^{-1}} && \text{since } v_1 = r_0 / \|r_0\|_{W^{-1}} \\
&= \left\| V_{k+1} \left(\|r_0\|_{W^{-1}}\, \vec{e}_1 - H \vec{y} \right) \right\|^2_{M^{-1}}
\end{aligned}$$

The Arnoldi-inner product W^{-1} is mathematically irrelevant and can be chosen for algorithmic convenience. Here we are led to choose $W := M$, which from now on we assume.

Residual Norm Minimization

$$\text{By design,} \quad x_k - x_0 \in \quad P^{-1} K_k(AP^{-1}; r_0)$$
$$\text{and therefore} \quad r_k - r_0 \in AP^{-1} K_k(AP^{-1}; r_0) = \text{range}\, AP^{-1} V_k.$$

Hence

$$\begin{aligned}
\|r_k\|^2_{M^{-1}} &= \left\| r_0 - AP^{-1} V_k \vec{y} \right\|^2_{M^{-1}} && \vec{y} \in \mathbb{R}^k \\
&= \left\| r_0 - V_{k+1} H \vec{y} \right\|^2_{M^{-1}} && \text{due to } AP^{-1} V_k = V_{k+1} H \\
&= \left\| \|r_0\|_{M^{-1}}\, v_1 - V_{k+1} H \vec{y} \right\|^2_{M^{-1}} && \text{since } v_1 = r_0 / \|r_0\|_{M^{-1}} \\
&= \left\| V_{k+1} \left(\|r_0\|_{M^{-1}}\, \vec{e}_1 - H \vec{y} \right) \right\|^2_{M^{-1}} && \\
&= \left\| \|r_0\|_{M^{-1}}\, \vec{e}_1 - H \vec{y} \right\|^2_2 && \text{by } M^{-1}\text{-orthonormality.}
\end{aligned}$$

The Arnoldi-inner product W^{-1} is mathematically irrelevant and can be chosen for algorithmic convenience. Here we are led to choose $W := M$, which from now on we assume.

Short Recurrences

Short recurrences occur when

$$H_{k,k} = \begin{bmatrix} (v_1, AP^{-1}v_1)_{M^{-1}} & (v_1, AP^{-1}v_2)_{M^{-1}} & \cdots \\ (v_2, AP^{-1}v_1)_{M^{-1}} & (v_2, AP^{-1}v_2)_{M^{-1}} & \cdots \\ 0 & (v_3, AP^{-1}v_2)_{M^{-1}} & \cdots \\ \vdots & 0 & \\ \vdots & \vdots & \end{bmatrix} \in \mathbb{R}^{k\times k}$$

is symmetric, i.e., when

$$M^{-1}AP^{-1} = P^{-\top}A^{\top}M^{-1}. \qquad \text{(SRC)}$$

This means that $AP^{-1} \in \mathcal{L}(X^*)$ is $\big(X^*, \|\cdot\|_{M^{-1}}\big)$-Hilbert space self-adjoint.

Short Recurrences

Short recurrences occur when

$$H_{k,k} = \begin{bmatrix} (v_1, AP^{-1}v_1)_{M^{-1}} & (v_1, AP^{-1}v_2)_{M^{-1}} & \cdots \\ (v_2, AP^{-1}v_1)_{M^{-1}} & (v_2, AP^{-1}v_2)_{M^{-1}} & \cdots \\ 0 & (v_3, AP^{-1}v_2)_{M^{-1}} & \cdots \\ \vdots & 0 & \\ \vdots & \vdots & \end{bmatrix} \in \mathbb{R}^{k \times k}$$

is symmetric, i.e., when

$$M^{-1}AP^{-1} = P^{-\top}A^{\top}M^{-1}. \tag{SRC}$$

This means that $AP^{-1} \in \mathcal{L}(X^*)$ is $\big(X^*, \|\cdot\|_{M^{-1}}\big)$-Hilbert space self-adjoint.

Should we call the method MINRES then?

(SRC) and the Normal Equations

We are going to refer to GMRES as just described by

$$\mathrm{GMRES}\big(A,\ P^{-1},\ M^{-1},\ b,\ x_0\big).$$

Recall

$$M^{-1}AP^{-1} = P^{-\top}A^{\top}M^{-1}. \tag{SRC}$$

Given A and M, (SRC) always holds for the choice $P^{-1} := H^{-1}A^{\top}M^{-1}$, where $H = H^{\top}$ is non-singular.

(SRC) and the Normal Equations

We are going to refer to GMRES as just described by

$$\mathrm{GMRES}\big(A,\ P^{-1},\ M^{-1},\ b,\ x_0\big).$$

Recall

$$M^{-1}AP^{-1} = P^{-\top}A^{\top}M^{-1}. \tag{SRC}$$

Given A and M, (SRC) always holds for the choice $P^{-1} := H^{-1}A^{\top}M^{-1}$, where $H = H^{\top}$ is non-singular. This is mathematically equivalent to running GMRES on the normal equations (modified problem)

$$A^{\top}M^{-1}Ax = A^{\top}M^{-1}b$$

with preconditioner H. In other words,

$$\begin{aligned} & \mathrm{GMRES}\big(A,\ H^{-1}A^{\top}M^{-1},\ M^{-1},\ b,\ x_0\big) \\ \Leftrightarrow \quad & \mathrm{GMRES}\big(A^{\top}M^{-1}A,\ H^{-1},\ A^{-1}MA^{-\top},\ A^{\top}M^{-1}b,\ x_0\big) \end{aligned}$$

(SRC) and the Normal Equations

We are going to refer to GMRES as just described by

$$\mathrm{GMRES}\big(A,\ P^{-1},\ M^{-1},\ b,\ x_0\big).$$

Recall

$$M^{-1}AP^{-1} = P^{-\top}A^\top M^{-1}. \tag{SRC}$$

Given A and M, (SRC) always holds for the choice $P^{-1} := H^{-1}A^\top M^{-1}$, where $H = H^\top$ is spd. This is mathematically equivalent to running GMRES on the normal equations (modified problem)

$$A^\top M^{-1}Ax = A^\top M^{-1}b$$

with preconditioner H. In other words,

$$\begin{aligned}
& \mathrm{GMRES}\big(A,\ H^{-1}A^\top M^{-1},\ M^{-1},\ b,\ x_0\big) \\
\Leftrightarrow\quad & \mathrm{GMRES}\big(A^\top M^{-1}A,\ H^{-1},\ A^{-1}MA^{-\top},\ A^\top M^{-1}b,\ x_0\big) \\
\Leftrightarrow\quad & \mathrm{CG}\big(A^\top M^{-1}A,\ H^{-1},\ A^\top M^{-1}b,\ x_0\big).
\end{aligned}$$

Convergence Analysis

In the absence of any particular properties of (A, P, M) we can use the Jordan decomposition (with $U, J \in \mathbb{C}^{n\times n}$)

$$M^{-1/2} A P^{-1} M^{1/2} = U\,J\,U^{-1}.$$

Then for any polynomial p,

$$p(AP^{-1}) = M^{1/2} U\,p(J)\,U^{-1} M^{-1/2}$$

holds and thus

$$\begin{aligned}
\left\|p(AP^{-1})\,r_0\right\|_{M^{-1}}^2 &= \left\|U\,p(J)\,U^{-1} M^{-1/2} r_0\right\|_2^2 \\
&\le \sigma_{\max}^2\big(U\,p(J)\,U^{-1}\big)\,\|r_0\|_{M^{-1}}^2 \\
&= \lambda_{\max}\big(p(AP^{-1})^\top M^{-1} p(AP^{-1});\, M^{-1}\big)\,\|r_0\|_{M^{-1}}^2.
\end{aligned}$$

Convergence Analysis

In the absence of any particular properties of (A, P, M) we can use the Jordan decomposition (with $U, J \in \mathbb{C}^{n\times n}$)

$$M^{-1/2} A P^{-1} M^{1/2} = U\, J\, U^{-1}.$$

Then for any polynomial p,

$$p(AP^{-1}) = M^{1/2} U\, p(J)\, U^{-1} M^{-1/2}$$

holds and thus

$$\begin{aligned}
\left\| p(AP^{-1})\, r_0 \right\|_{M^{-1}}^2 &= \left\| U\, p(J)\, U^{-1} M^{-1/2} r_0 \right\|_2^2 \\
&\le \sigma_{\max}^2\!\left(U\, p(J)\, U^{-1}\right) \|r_0\|_{M^{-1}}^2 \\
&= \lambda_{\max}\!\left(p(AP^{-1})^\top M^{-1} p(AP^{-1}); M^{-1}\right) \|r_0\|_{M^{-1}}^2.
\end{aligned}$$

Alternatively, using the Schur decomposition $M^{-1/2} A P^{-1} M^{1/2} = U\, R\, U^H$ (with $U, R \in \mathbb{C}^{n\times n}$) we arrive at

$$\begin{aligned}
\left\| p(AP^{-1})\, r_0 \right\|_{M^{-1}}^2 &\le \sigma_{\max}^2\!\left(p(R)\right) \|r_0\|_{M^{-1}}^2 \\
&= \lambda_{\max}\!\left(p(AP^{-1})^\top M^{-1} p(AP^{-1}); M^{-1}\right) \|r_0\|_{M^{-1}}^2.
\end{aligned}$$

Convergence Analysis

In the absence of any particular properties of (A, P, M) we can use the Jordan decomposition (with $U, J \in \mathbb{C}^{n\times n}$)

$$M^{-1/2} A P^{-1} M^{1/2} = U\, J\, U^{-1}.$$

Then for any polynomial p,

$$p(AP^{-1}) = M^{1/2} U\, p(J)\, U^{-1} M^{-1/2}$$

holds and thus

$$\begin{aligned} \left\| p(AP^{-1})\, r_0 \right\|_{M^{-1}}^2 &= \left\| U\, p(J)\, U^{-1} M^{-1/2} r_0 \right\|_2^2 \\ &\le \sigma_{\max}^2 \left(U\, p(J)\, U^{-1} \right) \|r_0\|_{M^{-1}}^2 \\ &= \lambda_{\max} \left(p(AP^{-1})^\top M^{-1} p(AP^{-1}); M^{-1} \right) \|r_0\|_{M^{-1}}^2 . \end{aligned}$$

This convergence estimate depends on the complete triplet (A, P, M).

Normality Condition

For $AP^{-1} \in \mathcal{L}(X^*)$, $(AP^{-1})^\top \in \mathcal{L}(X)$ is its adjoint and

$$(AP^{-1})^\circ = M\,(AP^{-1})^\top M^{-1}$$

is its Hilbert space adjoint w.r.t. $\big(X^*, \|\cdot\|_{M^{-1}}\big)$.

Normality Condition

For $AP^{-1} \in \mathcal{L}(X^*)$, $(AP^{-1})^\top \in \mathcal{L}(X)$ is its adjoint and

$$(AP^{-1})^\circ = M\,(AP^{-1})^\top M^{-1}$$

is its Hilbert space adjoint w.r.t. $\big(X^*, \|\cdot\|_{M^{-1}}\big)$.

We have already encountered the (Hilbert space) self-adjointness condition $AP^{-1} = (AP^{-1})^\circ$, responsible for short recursions:

$$M^{-1}AP^{-1} = P^{-\top}A^\top M^{-1}. \tag{SRC}$$

Normality Condition

For $AP^{-1} \in \mathcal{L}(X^*)$, $(AP^{-1})^\top \in \mathcal{L}(X)$ is its adjoint and

$$(AP^{-1})^\circ = M\,(AP^{-1})^\top M^{-1}$$

is its Hilbert space adjoint w.r.t. $\big(X^*, \|\cdot\|_{M^{-1}}\big)$.

We have already encountered the (Hilbert space) self-adjointness condition $AP^{-1} = (AP^{-1})^\circ$, responsible for short recursions:

$$M^{-1}AP^{-1} = P^{-\top}A^\top M^{-1}. \tag{SRC}$$

The normality condition $AP^{-1}(AP^{-1})^\circ = (AP^{-1})^\circ AP^{-1}$ reads

$$AP^{-1}M(AP^{-1})^\top M^{-1} = M\,(AP^{-1})^\top M^{-1}AP^{-1} \tag{NC}$$

or $B\,B^\top = B^\top B$ with $B = M^{-1/2}AP^{-1}M^{1/2}$.

Convergence Analysis Under Normality

Suppose that (NC) holds. Then with $U, D \in \mathbb{C}^{n\times n}$ unitary/diagonal,

$$B = M^{-1/2} A P^{-1} M^{1/2} = U\,D\,U^H$$

and for any polynomial p,

$$p(AP^{-1}) = M^{1/2} U\,p(D)\,U^H M^{-1/2}$$

holds and thus

$$\begin{aligned} \left\| p(AP^{-1})\, r_0 \right\|_{M^{-1}}^2 &= \left\| U\,p(D)\,U^H M^{-1/2} r_0 \right\|_2^2 \\ &\le \sigma_{\max}^2\big(U\,p(D)\,U^H\big)\,\|r_0\|_{M^{-1}}^2 \\ &= \big[\max_j |p(\lambda_j)|\,\big]^2\,\|r_0\|_{M^{-1}}^2, \end{aligned}$$

where λ_j are the (complex; real under the (SRC)) eigenvalues of AP^{-1}.

Convergence Analysis Under Normality

Suppose that (NC) holds. Then with $U, D \in \mathbb{C}^{n \times n}$ unitary/diagonal,

$$B = M^{-1/2} A P^{-1} M^{1/2} = U\, D\, U^H$$

and for any polynomial p,

$$p(AP^{-1}) = M^{1/2} U\, p(D)\, U^H M^{-1/2}$$

holds and thus

$$\begin{aligned} \big\| p(AP^{-1})\, r_0 \big\|_{M^{-1}}^2 &= \big\| U\, p(D)\, U^H M^{-1/2} r_0 \big\|_2^2 \\ &\le \sigma_{\max}^2\big(U\, p(D)\, U^H\big)\, \|r_0\|_{M^{-1}}^2 \\ &= \big[\max_j |p(\lambda_j)| \big]^2\, \|r_0\|_{M^{-1}}^2, \end{aligned}$$

where λ_j are the (complex; real under the (SRC)) eigenvalues of AP^{-1}.

The resulting convergence estimate depends only on (A, P) and it is independent of the norm M^{-1} we choose for residual minimization.

Intermediate Summary

- We have re-derived 'canonical' GMRES from 'first principles' and from a Hilbert space perspective, carefully distinguishing mapping properties and primal and dual quantities.
- Two choices arise (for the user):
  1. preconditioner $P \in \mathcal{L}(X, X^*)$
  2. inner product $(\cdot,\cdot)_{M^{-1}}$ in X^* for residual minimization
  3. inner product $(\cdot,\cdot)_{W^{-1}}$ in X^* for Arnoldi (non-essential)

  For the latter we fix wlog $W := M$ for algorithmic convenience.
- One application of A, P^{-1} and M^{-1} each is required per iteration.
- (SRC) $\Rightarrow$ (NC) $\Rightarrow$ convergence est. in terms of spectrum of AP^{-1}

Intermediate Summary

- We have re-derived 'canonical' GMRES from 'first principles' and from a Hilbert space perspective, carefully distinguishing mapping properties and primal and dual quantities.
- Two choices arise (for the user):
  1. preconditioner $P \in \mathcal{L}(X, X^*)$
  2. inner product $(\cdot, \cdot)_{M^{-1}}$ in X^* for residual minimization
  3. inner product $(\cdot, \cdot)_{W^{-1}}$ in X^* for Arnoldi (non-essential)

  For the latter we fix wlog $W := M$ for algorithmic convenience.
- One application of A, P^{-1} and M^{-1} each is required per iteration.
- (SRC) $\Rightarrow$ (NC) $\Rightarrow$ convergence est. in terms of spectrum of AP^{-1}
- Not discussed:
  - Solution of least-squares problems (by updated QR factorization).
  - Under particular choices of (A, P, M) one may take advantage of additional structure to streamline the implementation.

[$\rightarrow$ see talk by D. Szyld for an example]

Table of Contents

Dual vs. Primal Krylov Subspaces

'Canonical GMRES' is working with the (dual, residual-based) Krylov subspaces

$$K_k(AP^{-1}; r_0) \subset X^*$$

generated by $AP^{-1} \in \mathcal{L}(X^*)$. It turned out convenient to fix wlog $W := M$.

Dual vs. Primal Krylov Subspaces

'Canonical GMRES' is working with the (dual, residual-based) Krylov subspaces

$$K_k(AP^{-1}; r_0) \subset X^*$$

generated by $AP^{-1} \in \mathcal{L}(X^*)$. It turned out convenient to fix wlog $W := M$.

Equivalently, the algorithm can be written using the (primal) Krylov subspaces (see next two slides)

$$K_k(P^{-1}A; P^{-1}r_0) \subset X.$$

Dual vs. Primal Krylov Subspaces

'Canonical GMRES' is working with the (dual, residual-based) Krylov subspaces

$$K_k(AP^{-1}; r_0) \subset X^*$$

generated by $AP^{-1} \in \mathcal{L}(X^*)$. It turned out convenient to fix wlog $W := M$.

Equivalently, the algorithm can be written using the (primal) Krylov subspaces (see next two slides)

$$P^{-1}K_k(AP^{-1}; r_0) = K_k(P^{-1}A; P^{-1}r_0) \subset X.$$

Dual vs. Primal Krylov Subspaces

'Canonical GMRES' is working with the (dual, residual-based) Krylov subspaces

$$K_k(AP^{-1}; r_0) \subset X^*$$

generated by $AP^{-1} \in \mathcal{L}(X^*)$. It turned out convenient to fix wlog $W := M$.

Equivalently, the algorithm can be written using the (primal) Krylov subspaces (see next two slides)

$$P^{-1}K_k(AP^{-1}; r_0) = K_k(P^{-1}A; P^{-1}r_0) \subset X.$$

'Equivalently' means same iterates x_k, same residuals $r_k = b - Ax_k$, same residual norms $\|r_k\|_{M^{-1}}$.

Dual vs. Primal Krylov Subspaces

'Canonical GMRES' is working with the (dual, residual-based) Krylov subspaces

$$K_k(AP^{-1}; r_0) \subset X^*$$

generated by $AP^{-1} \in \mathcal{L}(X^*)$. It turned out convenient to fix wlog $W := M$.

Equivalently, the algorithm can be written using the (primal) Krylov subspaces (see next two slides)

$$P^{-1}K_k(AP^{-1}; r_0) = K_k(P^{-1}A; P^{-1}r_0) \subset X.$$

'Equivalently' means same iterates x_k, same residuals $r_k = b - Ax_k$, same residual norms $\|r_k\|_{M^{-1}}$.

Caution: Do not confuse with 'right/left preconditioning'. There is no such distinction in 'canonical GMRES'.

Using Primal Krylov Subspaces

The Arnoldi process generates a basis $\{U_k\}$ of $K_k(P^{-1}A; P^{-1}r_0) \subset X$ with is orthonormal w.r.t. the inner product W (mathematically irrelevant).

In matrix form the Arnoldi process reads

$$P^{-1}AU_k = U_{k+1}H,$$

where H is the upper Hessenberg matrix

$$H = \begin{bmatrix} (v_1,\, P^{-1}Av_1)_W & (v_1,\, P^{-1}Av_2)_W & \cdots \\ (v_2,\, P^{-1}Av_1)_W & (v_2,\, P^{-1}Av_2)_W & \cdots \\ 0 & (v_3,\, P^{-1}Av_2)_W & \cdots \\ \vdots & 0 & \\ \vdots & \vdots & \end{bmatrix} \in \mathbb{R}^{(k+1)\times k}.$$

The short recursion condition becomes $WP^{-1}A = (WP^{-1}A)^\top$, which appears to be different from (SRC) …

Using Primal Krylov Subspaces

We still want to minimize

$$
\begin{aligned}
\|r_k\|^2_{M^{-1}} &= \left\|r_0 - AU_k\vec{y}\right\|^2_{M^{-1}} && \vec{y} \in \mathbb{R}^k \\
&= \left\|P^{-1}r_0 - U_{k+1}H\vec{y}\right\|^2_{P^\top M^{-1}P} && P^{-1}AU_k = U_{k+1}H \\
&= \left\|\|P^{-1}r_0\|_W\, u_1 - U_{k+1}H\vec{y}\right\|^2_{P^\top M^{-1}P} && u_1 = P^{-1}r_0/\|P^{-1}r_0\|_W \\
&= \left\|U_{k+1}\left(\|P^{-1}r_0\|_W\,\vec{e}_1 - H\vec{y}\right)\right\|^2_{P^\top M^{-1}P}
\end{aligned}
$$

Using Primal Krylov Subspaces

We still want to minimize

$$
\begin{aligned}
\|r_k\|^2_{M^{-1}} &= \left\|r_0 - AU_k\vec{y}\right\|^2_{M^{-1}} && \vec{y} \in \mathbb{R}^k \\
&= \left\|P^{-1}r_0 - U_{k+1}H\vec{y}\right\|^2_{P^\top M^{-1}P} && P^{-1}AU_k = U_{k+1}H \\
&= \left\|\|P^{-1}r_0\|_W\, u_1 - U_{k+1}H\vec{y}\right\|^2_{P^\top M^{-1}P} && u_1 = P^{-1}r_0/\|P^{-1}r_0\|_W \\
&= \left\|U_{k+1}\left(\|P^{-1}r_0\|_W\, \vec{e}_1 - H\vec{y}\right)\right\|^2_{P^\top M^{-1}P}
\end{aligned}
$$

The Arnoldi-inner product W is mathematically irrelevant and can be chosen for algorithmic convenience. Here we are led to choose $W := P^\top M^{-1} P \in \mathcal{L}(X, X^*)$.

Using Primal Krylov Subspaces

We still want to minimize

$$
\begin{aligned}
\|r_k\|_{M^{-1}}^2 &= \left\| r_0 - A U_k \vec{y} \right\|_{M^{-1}}^2 && \vec{y} \in \mathbb{R}^k \\
&= \left\| P^{-1} r_0 - U_{k+1} H \vec{y} \right\|_{P^\top M^{-1} P}^2 && P^{-1} A U_k = U_{k+1} H \\
&= \left\| \|P^{-1} r_0\|_W \, u_1 - U_{k+1} H \vec{y} \right\|_{P^\top M^{-1} P}^2 && u_1 = P^{-1} r_0 / \|P^{-1} r_0\|_W \\
&= \left\| U_{k+1} \left(\|P^{-1} r_0\|_W \, \vec{e}_1 - H \vec{y} \right) \right\|_{P^\top M^{-1} P}^2 \\
&= \left\| \|r_0\|_{M^{-1}} \, \vec{e}_1 - H \vec{y} \right\|_2^2 && \text{by } W\text{-orthonormality.}
\end{aligned}
$$

The Arnoldi-inner product W is mathematically irrelevant and can be chosen for algorithmic convenience. Here we are led to choose $W := P^\top M^{-1} P \in \mathcal{L}(X, X^*)$.

Using Primal Krylov Subspaces

We still want to minimize

$$
\begin{aligned}
\|r_k\|^2_{M^{-1}} &= \left\|r_0 - AU_k\vec{y}\right\|^2_{M^{-1}} && \vec{y} \in \mathbb{R}^k \\
&= \left\|P^{-1}r_0 - U_{k+1}H\vec{y}\right\|^2_{P^\top M^{-1}P} && P^{-1}AU_k = U_{k+1}H \\
&= \left\|\|P^{-1}r_0\|_W\, u_1 - U_{k+1}H\vec{y}\right\|^2_{P^\top M^{-1}P} && u_1 = P^{-1}r_0/\|P^{-1}r_0\|_W \\
&= \left\|U_{k+1}\big(\|P^{-1}r_0\|_W\,\vec{e}_1 - H\vec{y}\big)\right\|^2_{P^\top M^{-1}P} \\
&= \left\|\|r_0\|_{M^{-1}}\,\vec{e}_1 - H\vec{y}\right\|^2_2 && \text{by } W\text{-orthonormality.}
\end{aligned}
$$

The Arnoldi-inner product W is mathematically irrelevant and can be chosen for algorithmic convenience. Here we are led to choose $W := P^\top M^{-1}P \in \mathcal{L}(X, X^*)$.

But then the implementations with primal/dual Krylov subspaces agree: $u_j = P^{-1}v_j$ holds, the Hessenberg matrices are the same, and the short recursion conditions agree. So we continue with the original version, using dual Krylov subspaces.

What if $A \in \mathcal{L}(X)$?

The residual $r := b - A\,x$ belongs to X now.

1. **Krylov subspaces**

   With $A \in \mathcal{L}(X)$ we can form

$$K_k(A\quad; r) := \operatorname{span}\{r,\ Ar\quad,\ A^2 r\quad,\ \ldots,\ A^{k-1} r\quad\} \subset X$$

What if $A \in \mathcal{L}(X)$?

The residual $r := b - A\,x$ belongs to X now.

1. **Krylov subspaces**

   With $A \in \mathcal{L}(X)$ we may still want to use a preconditioner

   $$K_k(AP^{-1}; r) := \operatorname{span}\big\{r,\ AP^{-1}r,\ (AP^{-1})^2 r,\ \ldots,\ (AP^{-1})^{k-1} r\big\} \subset X$$

   but $P \in \mathcal{L}(X)$ is optional now.

What if $A \in \mathcal{L}(X)$?

The residual $r := b - A\,x$ belongs to X now.

1. **Krylov subspaces**

   With $A \in \mathcal{L}(X)$ we may still want to use a preconditioner

   $$K_k(AP^{-1}; r) := \operatorname{span}\{r,\ AP^{-1}r,\ (AP^{-1})^2 r,\ \ldots,\ (AP^{-1})^{k-1} r\} \subset X$$

   but $P \in \mathcal{L}(X)$ is optional now.

2. **Inner product** $(\cdot,\cdot)_W$ in X for Arnoldi

   Since $K_k(AP^{-1}; r) \subset X$ holds, orthonormality is defined w.r.t. W.

What if $A \in \mathcal{L}(X)$?

The residual $r := b - A\,x$ belongs to X now.

1. **Krylov subspaces**

   With $A \in \mathcal{L}(X)$ we may still want to use a preconditioner

   $$K_k(AP^{-1}; r) := \operatorname{span}\{r,\ AP^{-1}r,\ (AP^{-1})^2 r,\ \ldots,\ (AP^{-1})^{k-1} r\} \subset X$$

   but $P \in \mathcal{L}(X)$ is optional now.

2. **Inner product** $(\cdot,\cdot)_W$ in X for Arnoldi

   Since $K_k(AP^{-1}; r) \subset X$ holds, orthonormality is defined w.r.t. W.

3. **Inner product** $(\cdot,\cdot)_M$ in X for residual minimization

   Since $r \in X$ holds, $\|r\|_M$ will be the quantity minimized.

The Arnoldi Process for $A \in \mathcal{L}(X, X^*)$

The Arnoldi process (here displayed with modified GS) generates a W^{-1}-orthonormal basis of the Krylov subspace $\mathcal{K}_k(AP^{-1}; r_0)$:

Input: Matrix A, P^{-1}, W^{-1} (or matrix-vector products), initial vector $r_0 \in X^*$
Output: W^{-1}-ONB $v_1, v_2, \ldots \in X^*$ with $\operatorname{span}\{v_1, \ldots, v_k\} = \mathcal{K}_k(AP^{-1}; r_0)$

Set $z_1 := W^{-1} r_0$ and $v_1 := \dfrac{r_0}{\langle r_0, z_1 \rangle^{1/2}}$ and $z_1 := \dfrac{z_1}{\langle r_0, z_1 \rangle^{1/2}}$
for $k = 1, 2, \ldots$ **do**
  Set $v_{k+1} := AP^{-1} v_k$ // new Krylov vector
  for $j = 1, 2, \ldots, k$ **do**
    Set $h_{j,k} := \langle v_{k+1}, z_j \rangle$ // orthogonal. coefficients
    Set $v_{k+1} := v_{k+1} - h_{j,k}\, v_j$
  end for
  Set $z_{k+1} := W^{-1} v_{k+1}$ and $h_{k+1,k} := \langle v_{k+1}, z_{k+1} \rangle^{1/2}$ // $= (v_{k+1}, v_{k+1})^{1/2}_{W^{-1}}$
  Set $v_{k+1} := \dfrac{v_{k+1}}{h_{k+1,k}}$ and $z_{k+1} := \dfrac{z_{k+1}}{h_{k+1,k}}$
end for

Both v_j and $z_j = W^{-1} v_j$ need to be stored due to MGS.

The Arnoldi Process for $A \in \mathcal{L}(X)$

The Arnoldi process (here displayed with modified GS) generates a W -orthonormal basis of the Krylov subspace $\mathcal{K}_k(AP^{-1}; r_0)$:

Input: Matrix A, P^{-1}, W (or matrix-vector products), initial vector $r_0 \in X$
Output: W- ONB $v_1, v_2, \ldots \in X$ with $\operatorname{span}\{v_1, \ldots, v_k\} = \mathcal{K}_k(AP^{-1}; r_0)$

Set $z_1 := W \; r_0$ and $v_1 := \dfrac{r_0}{\langle r_0, z_1 \rangle^{1/2}}$ and $z_1 := \dfrac{z_1}{\langle r_0, z_1 \rangle^{1/2}}$
for $k = 1, 2, \ldots$ **do**
    Set $v_{k+1} := AP^{-1} v_k$ // new Krylov vector
    for $j = 1, 2, \ldots, k$ **do**
        Set $h_{j,k} := \langle v_{k+1}, z_j \rangle$ // orthogonal. coefficients
        Set $v_{k+1} := v_{k+1} - h_{j,k} \, v_j$
    end for
    Set $z_{k+1} := W \; v_{k+1}$ and $h_{k+1,k} := \langle v_{k+1}, z_{k+1} \rangle^{1/2}$ // $= (v_{k+1}, v_{k+1})_W^{1/2}$
    Set $v_{k+1} := \dfrac{v_{k+1}}{h_{k+1,k}}$ and $z_{k+1} := \dfrac{z_{k+1}}{h_{k+1,k}}$
end for

Both v_j and $z_j = W \; v_j$ need to be stored due to MGS.

Residual Minimization for $A \in \mathcal{L}(X, X^*)$

By design, $x_k - x_0 \in P^{-1} K_k(AP^{-1}; r_0)$

and therefore $r_k - r_0 \in AP^{-1} K_k(AP^{-1}; r_0) = \operatorname{range} AP^{-1} V_k$.

$$
\begin{aligned}
\|r_k\|^2_{M^{-1}} &= \left\| r_0 - AP^{-1} V_k \vec{y} \right\|^2_{M^{-1}} && \vec{y} \in \mathbb{R}^k \\
&= \left\| r_0 - V_{k+1} H \vec{y} \right\|^2_{M^{-1}} && \text{due to } AP^{-1} V_k = V_{k+1} H \\
&= \left\| \|r_0\|_{M^{-1}} \, v_1 - V_{k+1} H \vec{y} \right\|^2_{M^{-1}} && \text{since } v_1 = r_0 / \|r_0\|_{M^{-1}} \\
&= \left\| V_{k+1} \left(\|r_0\|_{M^{-1}} \, \vec{e}_1 - H \vec{y} \right) \right\|^2_{M^{-1}} \\
&= \left\| \|r_0\|_{M^{-1}} \, \vec{e}_1 - H \vec{y} \right\|^2_2 && \text{by } M^{-1}\text{-orthonormality.}
\end{aligned}
$$

Residual Minimization for $A \in \mathcal{L}(X)$

By design, $\quad x_k - x_0 \in \quad P^{-1} K_k(AP^{-1}; r_0)$
and therefore $\quad r_k - r_0 \in AP^{-1} K_k(AP^{-1}; r_0) = \text{range}\, AP^{-1} V_k.$

$$
\begin{aligned}
\|r_k\|_M^2 &= \left\| r_0 - AP^{-1} V_k \vec{y} \right\|_M^2 && \vec{y} \in \mathbb{R}^k \\
&= \left\| r_0 - V_{k+1} H \vec{y} \right\|_M^2 && \text{due to } AP^{-1} V_k = V_{k+1} H \\
&= \left\| \|r_0\|_M \quad v_1 - V_{k+1} H \vec{y} \right\|_M^2 && \text{since } v_1 = r_0 / \|r_0\|_M \\
&= \left\| V_{k+1} \left(\|r_0\|_M \quad \vec{e}_1 - H \vec{y} \right) \right\|_M^2 && \\
&= \left\| \|r_0\|_M \quad \vec{e}_1 - H \vec{y} \right\|_2^2 && \text{by } M \quad \text{-orthonormality.}
\end{aligned}
$$

The Arnoldi-inner product W is mathematically irrelevant and can be chosen for algorithmic convenience. Here we are led — again — to choose $W := M$, which from now on we assume.

Short Rec./Norm. Conditions for $A \in \mathcal{L}(X, X^*)$

Self-adjointness $AP^{-1} = (AP^{-1})^\circ$ in $\left(X^*; \|\cdot\|_{M^{-1}}\right)$ is responsible for short recursions:

$$M^{-1}AP^{-1} = P^{-\top}A^\top M^{-1}. \tag{SRC}$$

The normality condition $AP^{-1}(AP^{-1})^\circ = (AP^{-1})^\circ AP^{-1}$ reads

$$AP^{-1}M \quad (AP^{-1})^\top M^{-1} = M \quad (AP^{-1})^\top M^{-1} AP^{-1}. \tag{NC}$$

Short Rec./Norm. Conditions for $A \in \mathcal{L}(X)$

Self-adjointness $AP^{-1} = (AP^{-1})^\circ$ in $\left(X \; ; \|\cdot\|_M \right)$ is responsible for short recursions:

$$M \quad AP^{-1} = P^{-\top} A^\top M. \tag{SRC}$$

The normality condition $AP^{-1}(AP^{-1})^\circ = (AP^{-1})^\circ AP^{-1}$ reads

$$AP^{-1} M^{-1} (AP^{-1})^\top M \quad = M^{-1} (AP^{-1})^\top M \quad AP^{-1}. \tag{NC}$$

The convergence analysis carries over, *mutatis mutandis*.

Overview of Canonical GMRES

$A \in \mathcal{L}(X, X^*)$

- preconditioner $P \in \mathcal{L}(X, X^*)$
- inner product $M \in \mathcal{L}(X, X^*)$
- Krylov subspaces $K_k(AP^{-1}; r_0) \subset X^*$
- Arnoldi orthonormality w.r.t. M^{-1}, apply M^{-1}
- residual minimization $\|r\|_{M^{-1}}$

$A \in \mathcal{L}(X)$

- preconditioner $P \in \mathcal{L}(X)$
- inner product $M \in \mathcal{L}(X, X^*)$
- Krylov subspaces $K_k(AP^{-1}; r_0) \subset X$
- Arnoldi orthonormality w.r.t. M, apply M
- residual minimization $\|r\|_M$

$$\text{GMRES}\big(A,\ P^{-1},\ M^{-1},\ b,\ x_0\big).$$

$$\text{GMRES}\big(A,\ P^{-1},\ M,\ b,\ x_0\big).$$

GMRES in the Literature I

The very frequent 'right preconditioned' GMRES is motivated by considering the modified problem

$$AP^{-1}u = b.$$

The Arnoldi process for $K_k(AP^{-1}; r_0)$ is typically carried out w.r.t. the Euclidean inner product and $\|r\|_2 = \|b - Ax\|_2$ is minimized.

A non-singular, P non-singular

[see for instance (Saad, 2003, Ch. 9.3), Matlab, PETSc, ...]

GMRES in the Literature I

The very frequent 'right preconditioned' GMRES is motivated by considering the modified problem

$$AP^{-1}u = b.$$

The Arnoldi process for $K_k(AP^{-1}; r_0)$ is typically carried out w.r.t. the Euclidean inner product and $\|r\|_2 = \|b - Ax\|_2$ is minimized.

This corresponds to canonical GMRES with

$$A \text{ non-singular}, \quad P \text{ non-singular}, \quad W = M = \mathrm{id}\,.$$

[see for instance (Saad, 2003, Ch. 9.3), Matlab, PETSc, ...]

GMRES in the Literature I

The very frequent 'right preconditioned' GMRES is motivated by considering the modified problem

$$AP^{-1}u = b.$$

The Arnoldi process for $K_k(AP^{-1}; r_0)$ is typically carried out w.r.t. the Euclidean inner product and $\|r\|_2 = \|b - Ax\|_2$ is minimized.

This corresponds to canonical GMRES with

$$A \text{ non-singular}, \quad P \text{ non-singular}, \quad W = M = \mathrm{id}\,.$$

- 👍 The user does not need to know whether $A \in \mathcal{L}(X, X^*)$ or $A \in \mathcal{L}(X)$ since $M = M^{-1} = \mathrm{id}$. (This also saves one operation per iteration.)
- 👎 The user cannot control which norm of the residual is being minimized. This appears to be a restriction, which cannot be compensated for by the preconditioner. Opportunities to satisfy (SRC) and (NC) are limited.

[see for instance (Saad, 2003, Ch. 9.3), Matlab, PETSc, ...]

GMRES in the Literature II

A small number of papers consider

$$Ax = b \quad \text{with } A \in \mathcal{L}(X, X^*)$$

where the preconditioner $P := M$ (spd) is chosen. The Arnoldi process for $K_k(AP^{-1}; r_0)$ is carried out w.r.t. the inner product $W^{-1} = M^{-1}$.

[see for instance (Starke, 1997, Sect. 3), (Chan et al., 1998, Alg. 2.3), (Ernst, 2000, Ch. 9.3), . . .]

GMRES in the Literature II

A small number of papers consider

$$Ax = b \quad \text{with } A \in \mathcal{L}(X, X^*)$$

where the preconditioner $P := M$ (spd) is chosen. The Arnoldi process for $K_k(AP^{-1}; r_0)$ is carried out w.r.t. the inner product $W^{-1} = M^{-1}$.

This corresponds to canonical GMRES with

$$A \text{ non-singular}, \quad P = M, \quad W = M.$$

[see for instance (Starke, 1997, Sect. 3), (Chan et al., 1998, Alg. 2.3), (Ernst, 2000, Ch. 9.3), ...]

GMRES in the Literature II

A small number of papers consider

$$Ax = b \quad \text{with } A \in \mathcal{L}(X, X^*)$$

where the preconditioner $P := M$ (spd) is chosen. The Arnoldi process for $K_k(AP^{-1}; r_0)$ is carried out w.r.t. the inner product $W^{-1} = M^{-1}$.

This corresponds to canonical GMRES with

$$A \text{ non-singular}, \quad P = M, \quad W = M.$$

- $P = M$ saves one operation per iteration.
- Coupling the choice of residual norm and preconditioner appears to be a restriction compared to full 'canonical GMRES'. Opportunities to satisfy (SRC) — which amounts to $A = A^\top$ — and (NC) are limited.

[see for instance (Starke, 1997, Sect. 3), (Chan et al., 1998, Alg. 2.3), (Ernst, 2000, Ch. 9.3), ...]

GMRES in the Literature III

Bramble & Pasciak derived a method (BPCG)—here adjusted to fit our setting—for self-adjoint, indefinite problems with $A \in \mathcal{L}(X, X^*)$ in saddle-point form

$$A = \begin{bmatrix} A_1 & B^\top \\ B & -A_2 \end{bmatrix}$$

with $A_1 \succ 0$ (spd), $A_2 \succeq 0$ (spsd).

[Bramble and Pasciak (1988), . . .]

GMRES in the Literature III

Bramble & Pasciak derived a method (BPCG)—here adjusted to fit our setting—for self-adjoint, indefinite problems with $A \in \mathcal{L}(X, X^*)$ in saddle-point form with a particular preconditioner P and inner product M:

$$A = \begin{bmatrix} A_1 & B^\top \\ B & -A_2 \end{bmatrix}, \quad P = \begin{bmatrix} \widehat{A}_1 & B^\top \\ 0 & -\operatorname{id} \end{bmatrix}, \quad M = \begin{bmatrix} A_1 - \widehat{A}_1 & 0 \\ 0 & \operatorname{id} \end{bmatrix}$$

with $A_1 \succ 0$ (spd), $A_2 \succeq 0$ (spsd).

[Bramble and Pasciak (1988), ...]

GMRES in the Literature III

Bramble & Pasciak derived a method (BPCG)—here adjusted to fit our setting—for self-adjoint, indefinite problems with $A \in \mathcal{L}(X, X^*)$ in saddle-point form with a particular preconditioner P and inner product M:

$$A = \begin{bmatrix} A_1 & B^\top \\ B & -A_2 \end{bmatrix}, \quad P = \begin{bmatrix} \widehat{A}_1 & B^\top \\ 0 & -\operatorname{id} \end{bmatrix}, \quad M = \begin{bmatrix} A_1 - \widehat{A}_1 & 0 \\ 0 & \operatorname{id} \end{bmatrix}$$

with $A_1 \succ 0$ (spd), $A_2 \succeq 0$ (spsd). The (SRC) holds by construction. Moreover, under appropriate assumptions, the eigenvalues of AP^{-1} are not only real but positive.

- 👍 One obtains a CG-like convergence result.
- 👎 $\|r\|_{M^{-1}} = \|e\|_{AM^{-1}A}$ is non-trivial to interpret.

[Bramble and Pasciak (1988), . . .]

GMRES in the Literature IV

Klawonn extended the analysis of Bramble & Pasciak and employed canonical GMRES for self-adjoint, indefinite problems with $A \in \mathcal{L}(X, X^*)$ in saddle-point form

$$A = \begin{bmatrix} A_1 & B^\top \\ B & -t^2 A_2 \end{bmatrix}$$

with $A_1, A_2 \succ 0$ (spd).

[see for instance Klawonn (1998) and also Simoncini (2004), ...]

GMRES in the Literature IV

Klawonn extended the analysis of Bramble & Pasciak and employed canonical GMRES for self-adjoint, indefinite problems with $A \in \mathcal{L}(X, X^*)$ in saddle-point form with a particular preconditioner P and inner product M:

$$A = \begin{bmatrix} A_1 & B^\top \\ B & -t^2 A_2 \end{bmatrix}, \quad P = \begin{bmatrix} \widehat{A}_1 & B^\top \\ 0 & -\widehat{A}_2 \end{bmatrix}, \quad M = \begin{bmatrix} A_1 - \widehat{A}_1 & 0 \\ 0 & \widehat{A}_2 \end{bmatrix}$$

with $A_1, A_2 \succ 0$ (spd).

[see for instance Klawonn (1998) and also Simoncini (2004), ...]

GMRES in the Literature IV

Klawonn extended the analysis of Bramble & Pasciak and employed canonical GMRES for self-adjoint, indefinite problems with $A \in \mathcal{L}(X, X^*)$ in saddle-point form with a particular preconditioner P and inner product M:

$$A = \begin{bmatrix} A_1 & B^\top \\ B & -t^2 A_2 \end{bmatrix}, \quad P = \begin{bmatrix} \widehat{A}_1 & B^\top \\ 0 & -\widehat{A}_2 \end{bmatrix}, \quad M = \begin{bmatrix} A_1 - \widehat{A}_1 & 0 \\ 0 & \widehat{A}_2 \end{bmatrix}$$

with $A_1, A_2 \succ 0$ (spd). The (SRC) holds (although it does not seem to have been used in the numerical experiments). Moreover, under appropriate assumptions, the eigenvalues of AP^{-1} are not only real but positive.

[see for instance Klawonn (1998) and also Simoncini (2004), ...]

GMRES in the Literature IV

Klawonn extended the analysis of Bramble & Pasciak and employed canonical GMRES for self-adjoint, indefinite problems with $A \in \mathcal{L}(X, X^*)$ in saddle-point form with a particular preconditioner P and inner product M:

$$A = \begin{bmatrix} A_1 & B^\top \\ B & -t^2 A_2 \end{bmatrix}, \quad P = \begin{bmatrix} \widehat{A}_1 & B^\top \\ 0 & -\widehat{A}_2 \end{bmatrix}, \quad M = \begin{bmatrix} A_1 - \widehat{A}_1 & 0 \\ 0 & \widehat{A}_2 \end{bmatrix}$$

with $A_1, A_2 \succ 0$ (spd). The (SRC) holds (although it does not seem to have been used in the numerical experiments). Moreover, under appropriate assumptions, the eigenvalues of AP^{-1} are not only real but positive.

- 👍 One obtains a CG-like convergence result, robustly in $t \in \mathbb{R}$.
- 👎 $\|r\|_{M^{-1}} = \|e\|_{AM^{-1}A}$ is non-trivial to interpret.

[see for instance Klawonn (1998) and also Simoncini (2004), ...]

GMRES in the Literature V

The term 'preconditioned MINRES' seems to be reserved for the setting

$$A = A^\top, \quad P \text{ spd}, \quad M = P.$$

Certainly the (SRC) holds in this setting:

$$M^{-1} A P^{-1} = P^{-\top} A^\top M^{-1} \tag{SRC}$$

but it is also valid in many other situations with or without $A = A^\top$.

[Elman et al. (2014); Günnel et al. (2014)]

GMRES in the Literature VI

The very frequent 'left preconditioned' GMRES is motivated by considering the modified problem

$$P_L^{-1} A x = P_L^{-1} b.$$

The Arnoldi process for $K_k(P_L^{-1}A; P_L^{-1}r_0)$ is typically carried out w.r.t. the Euclidean inner product and $\|P_L^{-1} r\|_2 = \|b - Ax\|_{P_L^{-\top} P_L^{-1}}$ is minimized.

This cannot be modeled in canonical GMRES except as a modified problem:

$$P_L^{-1} A \text{ non-singular}, \quad P = \mathrm{id}, \quad W = M = \mathrm{id}\,.$$

[see for instance (Saad, 2003, Ch. 9.3), Matlab, PETSc, ...]

GMRES in the Literature VI

The very frequent 'left preconditioned' GMRES is motivated by considering the modified problem

$$P_L^{-1} A x = P_L^{-1} b.$$

The Arnoldi process for $K_k(P_L^{-1} A; P_L^{-1} r_0)$ is typically carried out w.r.t. the Euclidean inner product and $\|P_L^{-1} r\|_2 = \|b - Ax\|_{P_L^{-\top} P_L^{-1}}$ is minimized.

This cannot be modeled in canonical GMRES except as a modified problem:

$$P_L^{-1} A \text{ non-singular}, \quad P = \text{id}, \quad W = M = \text{id}.$$

- The user does not need to know whether A, P belong to $\mathcal{L}(X, X^*)$ or $\mathcal{L}(X)$ since the choice $P = P^{-1} = \text{id}$ and $M = M^{-1} = \text{id}$ holds. Moreover, this choice reduces the cost per iteration.
- The user may not be able to factorize the desired residual metric. Not using P nor M appears to be a restriction, which cannot be compensated for by P_L. Opportunities for (SRC) and (NC) are limited.

[see for instance (Saad, 2003, Ch. 9.3), Matlab, PETSc, ...]

GMRES in the Literature VII

Pestana & Wathen also consider a 'left preconditioned' GMRES

$$P_L^{-1}Ax = P_L^{-1}b.$$

where $A \in \mathcal{L}(X)$. Both A and the 'left preconditioner' $P_L \in \mathcal{L}(X)$ are assumed to be $(X, \|\cdot\|_H)$-Hilbert space self-adjoint w.r.t. a reference inner product $H \in \mathcal{L}(X, X^*)$: $HA = A^\top H$ and $HP_L = P_L^\top H$.

This corresponds to canonical GMRES with

$$P_L^{-1}A \text{ non-singular}, \quad P = \mathrm{id}$$

[(Pestana and Wathen, 2013, preprint version)]

GMRES in the Literature VII

Pestana & Wathen also consider a 'left preconditioned' GMRES

$$P_L^{-1} A x = P_L^{-1} b.$$

where $A \in \mathcal{L}(X)$. Both A and the 'left preconditioner' $P_L \in \mathcal{L}(X)$ are assumed to be $(X, \|\cdot\|_H)$-Hilbert space self-adjoint w.r.t. a reference inner product $H \in \mathcal{L}(X, X^*)$: $HA = A^\top H$ and $HP_L = P_L^\top H$. The Arnoldi process for $K_k(P_L^{-1} A; P_L^{-1} r_0)$ is carried out w.r.t. the inner product $W = HP_L$ and $\|P_L^{-1} r\|_{HP_L} = \|b - Ax\|_{P_L^{-\top} H}$ is minimized.

This corresponds to canonical GMRES with

$$P_L^{-1} A \text{ non-singular}, \quad P = \mathrm{id}\,, \quad W = M = HP_L.$$

[(Pestana and Wathen, 2013, preprint version)]

GMRES in the Literature VII

Pestana & Wathen also consider a 'left preconditioned' GMRES

$$P_L^{-1} A x = P_L^{-1} b.$$

where $A \in \mathcal{L}(X)$. Both A and the 'left preconditioner' $P_L \in \mathcal{L}(X)$ are assumed to be $(X, \|\cdot\|_H)$-Hilbert space self-adjoint w.r.t. a reference inner product $H \in \mathcal{L}(X, X^*)$: $HA = A^\top H$ and $HP_L = P_L^\top H$. The Arnoldi process for $K_k(P_L^{-1} A; P_L^{-1} r_0)$ is carried out w.r.t. the inner product $W = HP_L$ and $\|P_L^{-1} r\|_{HP_L} = \|b - Ax\|_{P_L^{-\top} H}$ is minimized.

This corresponds to canonical GMRES with

$$P_L^{-1} A \text{ non-singular}, \quad P = \mathrm{id}\,, \quad W = M = HP_L.$$

- The (SRC) holds by construction.
- The requirements appear to constitute a restriction compared to full 'canonical GMRES'.

[(Pestana and Wathen, 2013, preprint version)]

Table of Contents

Concluding Remarks

- We have re-derived 'canonical GMRES' from a Hilbert space perspective for $Ax = b$ in X^* (and $Ax = b$ in X).
- It appears natural that the user
  1. first chooses the inner product $(\cdot,\cdot)_{M^{-1}}$ in X^* for residual minimization,
  2. then the preconditioner $P \in \mathcal{L}(X, X^*)$, so that ideally (SRC) or at least (NC) holds.
- If (NC) is achieved, then P is responsible for fast convergence (eigenvalues of AP^{-1}), while M is responsible for being able to observe a meaningful norm of the residual.
- Don't use Matlab or PETSc implementations of GMRES unless you are determined to measure (and stop on) the residual norm $\|r\|_2$, since they make the choice $M = \mathrm{id}$ for you.
- Should we say MINRES rather than GMRES when (SRC) hold?

Concluding Remarks

- We have re-derived 'canonical GMRES' from a Hilbert space perspective for $Ax = b$ in X^* (and $Ax = b$ in X).
- It appears natural that the user
  1. first chooses the inner product $(\cdot, \cdot)_{M^{-1}}$ in X^* for residual minimization,
  2. then the preconditioner $P \in \mathcal{L}(X, X^*)$, so that ideally (SRC) or at least (NC) holds.
- If (NC) is achieved, then P is responsible for fast convergence (eigenvalues of AP^{-1}), while M is responsible for being able to observe a meaningful norm of the residual.
- Don't use Matlab or PETSc implementations of GMRES unless you are determined to measure (and stop on) the residual norm $\|r\|_2$, since they make the choice $M = \mathrm{id}$ for you.
- Should we say MINRES rather than GMRES when (SRC) hold?

Thank You

References I

Bramble, J. and Pasciak, J. (1988). A preconditioning technique for indefinite systems resulting from mixed approximations of elliptic problems. *Mathematics of Computation* 50: 1–17.

Calvetti, D., Lewis, B. and Reichel, L. (2002). On the regularizing properties of the GMRES method. *Numerische Mathematik* 91: 605–625, doi: 10.1007/s002110100339.

Campbell, S. L., Ipsen, I. C. F., Kelley, C. T. and Meyer, C. D. (1996). GMRES and the minimal polynomial. *BIT. Numerical Mathematics* 36: 664–675, doi: 10.1007/BF01733786.

Chan, T. F., Chow, E., Saad, Y. and Yeung, M. C. (1998). Preserving symmetry in preconditioned Krylov subspace methods. *SIAM Journal on Scientific Computing* 20: 568–581, doi: 10.1137/S1064827596311554.

Chen, J.-Y., Kincaid, D. R. and Young, D. M. (1999). Generalizations and modifications of the GMRES iterative method. *Numerical Algorithms* 21: 119–146, doi: 10.1023/A:1019105328973, numerical methods for partial differential equations (Marrakech, 1998).

Eiermann, M. and Ernst, O. G. (2001). Geometric aspects of the theory of Krylov subspace methods. *Acta Numerica* 10: 251–312, doi: 10.1017/S0962492901000046.

Elman, H., Silvester, D. and Wathen, A. (2014). *Finite elements and fast iterative solvers: with applications in incompressible fluid dynamics*. Numerical Mathematics and Scientific Computation. Oxford University Press, 2nd ed.

References II

Ernst, O. (2000). Minimal and orthogonal residual methods and their generalizations for solving linear operator equations. Habilitation Thesis, Faculty of Mathematics and Computer Sciences, TU Bergakademie Freiberg.

Gasparo, M. G., Papini, A. and Pasquali, A. (2008). Some properties of GMRES in Hilbert spaces. *Numerical Functional Analysis and Optimization. An International Journal* 29: 1276–1285, doi: `10.1080/01630560802580786`.

Günnel, A., Herzog, R. and Sachs, E. (2014). A note on preconditioners and scalar products in Krylov subspace methods for self-adjoint problems in Hilbert space. *Electronic Transactions on Numerical Analysis* 41: 13–20.

Klawonn, A. (1998). Block-triangular preconditioners for saddle point problems with a penalty term. *SIAM Journal on Scientific Computing* 19: 172–184 (electronic), doi: `10.1137/S1064827596303624`, special issue on iterative methods (Copper Mountain, CO, 1996).

Moret, I. (1997). A note on the superlinear convergence of GMRES. *SIAM Journal on Numerical Analysis* 34: 513–516, doi: `10.1137/S0036142993259792`.

Pestana, J. and Wathen, A. J. (2013). On the choice of preconditioner for minimum residual methods for non-Hermitian matrices. *Journal of Computational and Applied Mathematics* 249: 57–68, doi: `10.1016/j.cam.2013.02.020`.

Saad, Y. (2003). *Iterative Methods for Sparse Linear Systems*. Philadelphia: SIAM, 2nd ed.

References III

Saad, Y. and Schultz, M. H. (1986). GMRES: a generalized minimal residual algorithm for solving nonsymmetric linear systems. *Society for Industrial and Applied Mathematics. Journal on Scientific and Statistical Computing* 7: 856–869, doi: 10.1137/0907058.

Sarkis, M. and Szyld, D. B. (2007). Optimal left and right additive Schwarz preconditioning for minimal residual methods with Euclidean and energy norms. *Computer Methods in Applied Mechanics and Engineering* 196: 1612–1621, doi: 10.1016/j.cma.2006.03.027.

Simoncini, V. (2004). Block triangular preconditioners for symmetric saddle-point problems. *Applied Numerical Mathematics. An IMACS Journal* 49: 63–80, doi: 10.1016/j.apnum.2003.11.012.

Starke, G. (1997). Field-of-values analysis of preconditioned iterative methods for nonsymmetric elliptic problems. *Numerische Mathematik* 78: 103–117, doi: 10.1007/s002110050306.