Achieving scalability with a Hermitian preconditioner for a class of non-Hermitian matrices

Nicole Spillane

CNRS, École Polytechnique – www.cmap.polytechnique.fr/~spillane

General Problem

Solve $\mathbf{A}\mathbf{x}_* = \mathbf{b}$ by Weighted and Preconditioned GMRES (WP-GMRES) when $\mathbf{A}$ is positive-definite, *i.e.*, $\mathbf{A}$ has positive-definite Hermitian part.

WP-GMRES is GMRES in the W-inner product [1, 3]

Weight matrix $\mathbf{W}$ is hpd and $\langle \mathbf{x}, \mathbf{y}\rangle_{\mathbf{W}} = \mathbf{y}^*\mathbf{W}\mathbf{x}$.
In the right-preconditioned version (by $\mathbf{H}$), the i-th residual $\mathbf{r}_i$ satisfies:

$$\|\mathbf{r}_i\|_{\mathbf{W}} = \min\Big(\underbrace{\|\mathbf{b}-\mathbf{A}\mathbf{x}\|_{\mathbf{W}}}_{\text{Weighted Norm}};\ \mathbf{x} \in \mathbf{x}_0 + \underbrace{\operatorname{span}\{\mathbf{H}\mathbf{r}_0, \mathbf{H}\mathbf{A}\mathbf{H}\mathbf{r}_0, \dots, (\mathbf{H}\mathbf{A})^{i-1}\mathbf{H}\mathbf{r}_0\}}_{\text{Preconditioned Krylov subspace}}\Big)$$

Takeaway

- WP-GMRES converges fast if
  - $\mathbf{H}$ is a good hpd preconditioner for $\mathbf{M}(\mathbf{A}) = \frac{\mathbf{A}+\mathbf{A}^*}{2}$,
  - $\mathbf{A}$ is mildly non-hermitian,
  - $\mathbf{W}$ is set to $\mathbf{W} = \mathbf{H}$.
- WP-GMRES scales if $\mathbf{H}$ is a *scalable preconditioner* for $\mathbf{M}(\mathbf{A}) = \frac{\mathbf{A}+\mathbf{A}^*}{2}$.

Objective of this work

Provide convergence bounds and choices of

- Preconditioner $\mathbf{H}$,
- Weight matrix $\mathbf{W}$

that guarantee fast convergence.

Preprint (arXiv)

Nicole Spillane (2023)
Hermitian Preconditioning for a class of Non-Hermitian Linear Systems.

Bibliography

[1] Xiao-Chuan Cai.
Some domain decomposition algorithms for nonselfadjoint elliptic and parabolic partial differential equations.
PhD thesis, New York University, 1989.

[2] Stanley C. Eisenstat, Howard C. Elman, and Martin H. Schultz.
Variational iterative methods for nonsymmetric systems of linear equations.
SIAM J. Numer. Anal., 20:345–357, 1983.

[3] Azeddine Essai.
Weighted FOM and GMRES for solving nonsymmetric linear systems.
Numer. Algorithms, 18(3-4):277–292, 1998.

[4] Charles R. Johnson.
Inequalities for a complex matrix whose real part is positive definite.
Trans. Am. Math. Soc., 212:149–154, 1975.

[5] N. Spillane, V. Dolean, P. Hauret, F. Nataf, C. Pechstein, and R. Scheichl.
Abstract robust coarse spaces for systems of PDEs via generalized eigenproblems in the overlaps.
Numer. Math., 126(4):741–770, 2014.

Analysis and Theoretical results *via* WP-GCR

WP-GCR

$\mathbf{r}_0 = \mathbf{b} - \mathbf{A}\mathbf{x}_0$;
$\mathbf{z}_0 = \mathbf{H}\mathbf{r}_0$;
$\mathbf{p}_0 = \mathbf{z}_0$;
$\mathbf{q}_0 = \mathbf{A}\mathbf{p}_0$;
for $i = 0, 1, \dots$ **do**
  $\delta_i = \langle \mathbf{q}_i, \mathbf{q}_i\rangle_{\mathbf{W}}$;
  $\alpha_i = \frac{\langle \mathbf{q}_i, \mathbf{r}_i\rangle_{\mathbf{W}}}{\delta_i}$;
  $\mathbf{x}_{i+1} = \mathbf{x}_i + \alpha_i \mathbf{p}_i$;
  $\mathbf{r}_{i+1} = \mathbf{r}_i - \alpha_i \mathbf{q}_i$;
  $\mathbf{z}_{i+1} = \mathbf{H}\mathbf{r}_{i+1}$;
  $\mathbf{p}_{i+1} = \mathbf{z}_{i+1}$;
  for $j = 0, 1, \dots, i$ **do**
    $\beta_{i,j} = \frac{\langle \mathbf{q}_j, \mathbf{A}\mathbf{z}_{i+1}\rangle_{\mathbf{W}}}{\delta_j}$;
    $\mathbf{p}_{i+1} = \mathbf{p}_{i+1} - \beta_{i,j}\mathbf{p}_j$;
  end
  $\mathbf{q}_{i+1} = \mathbf{A}\mathbf{p}_{i+1}$;
end

- WP-GCR provides the same approximate solutions as WP-GMRES.
  Since $\mathbf{A}$ is positive-definite, no breakdown, and
  $$\|\mathbf{r}_i\|_{\mathbf{W}} = \min\left(\|\mathbf{b}-\mathbf{A}\mathbf{x}\|_{\mathbf{W}};\ \mathbf{x} \in \mathbf{x}_0 + \mathcal{K}_i\right), \text{ where } \mathcal{K}_i = \operatorname{span}\{\mathbf{H}\mathbf{r}_0, \mathbf{H}\mathbf{A}\mathbf{H}\mathbf{r}_0, \dots, (\mathbf{H}\mathbf{A})^{i-1}\mathbf{H}\mathbf{r}_0\}.$$
- WP-GCR computes an $\mathbf{A}^*\mathbf{W}\mathbf{A}$-orthonormal basis of $\mathcal{K}_i$.
  $$\mathcal{K}_i = \operatorname{span}\{\mathbf{p}_0, \dots, \mathbf{p}_{i-1}\};\ \mathbf{A}\mathcal{K}_i = \operatorname{span}\{\mathbf{q}_0, \dots, \mathbf{q}_{i-1}\} \text{ and } \langle \mathbf{q}_i, \mathbf{q}_j\rangle_{\mathbf{W}} = \delta_{ij} \text{ (Kronecker symbol).}$$

Convergence Theorem

If preconditioner $\mathbf{H}$ is hpd and $\mathbf{W} = \mathbf{H}$, then

$$\frac{\|\mathbf{r}_i\|_{\mathbf{W}}}{\|\mathbf{r}_0\|_{\mathbf{W}}} \le \left[1 - \frac{1}{\kappa(\mathbf{H}\mathbf{M}(\mathbf{A}))} \times \frac{1}{1+\rho(\mathbf{M}(\mathbf{A})^{-1}\mathbf{N}(\mathbf{A}))^2}\right]^{i/2}.$$

- $\mathbf{M}(\mathbf{A}) := \frac{\mathbf{A}+\mathbf{A}^*}{2}$ (Hermitian part)
- $\mathbf{N}(\mathbf{A}) := \frac{\mathbf{A}-\mathbf{A}^*}{2}$ (Skew-Hermitian part)
- $\kappa(\cdot)$: condition number
- $\rho(\cdot)$: spectral radius

Proof

- From update formula $\mathbf{r}_i = \mathbf{r}_{i+1} + |\alpha_i|\mathbf{q}_i$ with $\mathbf{r}_{i+1} \perp^{\mathbf{W}} \mathbf{q}_i$:
  $$\|\mathbf{r}_i\|_{\mathbf{W}}^2 = \|\mathbf{r}_{i+1}\|_{\mathbf{W}}^2 + |\alpha_i|^2\|\mathbf{q}_i\|_{\mathbf{W}}^2 = \|\mathbf{r}_{i+1}\|_{\mathbf{W}}^2 + \frac{|\langle \mathbf{q}_i, \mathbf{r}_i\rangle_{\mathbf{W}}|^2}{\|\mathbf{q}_i\|_{\mathbf{W}}^4}\|\mathbf{q}_i\|_{\mathbf{W}}^2.$$
- From orthogonalization formula $\mathbf{q}_i = \mathbf{A}\mathbf{z}_i - \sum_{j=0}^{i-1}\beta_{i,j}\mathbf{q}_j$:
  - $\langle \mathbf{q}_i, \mathbf{r}_i\rangle_{\mathbf{W}} = \langle \mathbf{A}\mathbf{z}_i, \mathbf{r}_i\rangle_{\mathbf{W}} - \sum_{j=0}^{i-1}\beta_{i,j}\langle \mathbf{q}_j, \mathbf{r}_i\rangle_{\mathbf{W}} = \langle \mathbf{A}\mathbf{z}_i, \mathbf{r}_i\rangle_{\mathbf{W}} = \langle \mathbf{A}\mathbf{H}\mathbf{r}_i, \mathbf{r}_i\rangle_{\mathbf{W}}$,
  - $\|\mathbf{A}\mathbf{H}\mathbf{r}_i\|_{\mathbf{W}}^2 = \|\mathbf{A}\mathbf{z}_i\|_{\mathbf{W}}^2 = \|\mathbf{q}_i + \sum_{j=0}^{i-1}\beta_{i,j}\mathbf{q}_j\|_{\mathbf{W}}^2 = \|\mathbf{q}_i\|_{\mathbf{W}}^2 + \sum_{j=0}^{i-1}\beta_{i,j}^2\|\mathbf{q}_j\|_{\mathbf{W}}^2 \ge \|\mathbf{q}_i\|_{\mathbf{W}}^2$.
- So:
  $$\frac{\|\mathbf{r}_{i+1}\|_{\mathbf{W}}}{\|\mathbf{r}_i\|_{\mathbf{W}}} \le \left[1 - \frac{\langle \mathbf{A}\mathbf{H}\mathbf{r}_i, \mathbf{r}_i\rangle_{\mathbf{W}}^2}{\|\mathbf{A}\mathbf{H}\mathbf{r}_i\|_{\mathbf{W}}^2\|\mathbf{r}_i\|_{\mathbf{W}}^2}\right]^{1/2} \le \left[1 - \inf_{\mathbf{y}\neq 0}\frac{|\langle \mathbf{A}\mathbf{H}\mathbf{y}, \mathbf{y}\rangle_{\mathbf{W}}|^2}{\|\mathbf{A}\mathbf{H}\mathbf{y}\|_{\mathbf{W}}^2\|\mathbf{y}\|_{\mathbf{W}}^2}\right]^{i/2}$$

$$\begin{aligned}
\inf_{\mathbf{y}\neq 0}\frac{|\langle \mathbf{A}\mathbf{H}\mathbf{y}, \mathbf{y}\rangle_{\mathbf{H}}|^2}{\|\mathbf{A}\mathbf{H}\mathbf{y}\|_{\mathbf{H}}^2\|\mathbf{y}\|_{\mathbf{H}}^2} &= \inf_{\mathbf{y}\neq 0}\frac{|\langle \mathbf{A}\mathbf{y}, \mathbf{y}\rangle|^2}{\langle \mathbf{H}\mathbf{A}\mathbf{y}, \mathbf{A}\mathbf{y}\rangle\langle \mathbf{H}^{-1}\mathbf{y}, \mathbf{y}\rangle} \ge \inf_{\mathbf{y}\neq 0}\frac{|\langle \mathbf{A}\mathbf{y}, \mathbf{y}\rangle|}{\langle \mathbf{H}\mathbf{A}\mathbf{y}, \mathbf{A}\mathbf{y}\rangle} \times \inf_{\mathbf{y}\neq 0}\frac{|\langle \mathbf{A}\mathbf{y}, \mathbf{y}\rangle|}{\langle \mathbf{H}^{-1}\mathbf{y}, \mathbf{y}\rangle}\\
&= \inf_{\mathbf{y}\neq 0}\frac{|\langle \mathbf{A}^{-1}\mathbf{y}, \mathbf{y}\rangle|}{\langle \mathbf{H}\mathbf{y}, \mathbf{y}\rangle} \times \inf_{\mathbf{y}\neq 0}\frac{|\langle \mathbf{A}\mathbf{y}, \mathbf{y}\rangle|}{\langle \mathbf{H}^{-1}\mathbf{y}, \mathbf{y}\rangle} \ge \inf_{\mathbf{y}\neq 0}\frac{\langle \mathbf{M}(\mathbf{A}^{-1})\mathbf{y}, \mathbf{y}\rangle}{\langle \mathbf{H}\mathbf{y}, \mathbf{y}\rangle} \times \inf_{\mathbf{y}\neq 0}\frac{\langle \mathbf{M}(\mathbf{A})\mathbf{y}, \mathbf{y}\rangle}{\langle \mathbf{H}^{-1}\mathbf{y}, \mathbf{y}\rangle}\\
&\ge \underbrace{\inf_{\mathbf{y}\neq 0}\frac{\langle \mathbf{M}(\mathbf{A}^{-1})\mathbf{y}, \mathbf{y}\rangle}{\langle \mathbf{M}(\mathbf{A})^{-1}\mathbf{y}, \mathbf{y}\rangle}}_{=(1+\rho(\mathbf{M}(\mathbf{A})^{-1}\mathbf{N}(\mathbf{A}))^2)^{-1} \text{ by [4]}} \times \underbrace{\inf_{\mathbf{y}\neq 0}\frac{\langle \mathbf{M}(\mathbf{A})^{-1}\mathbf{y}, \mathbf{y}\rangle}{\langle \mathbf{H}\mathbf{y}, \mathbf{y}\rangle} \times \inf_{\mathbf{y}\neq 0}\frac{\langle \mathbf{M}(\mathbf{A})\mathbf{y}, \mathbf{y}\rangle}{\langle \mathbf{H}^{-1}\mathbf{y}, \mathbf{y}\rangle}}_{\text{Rayleigh quotients for } (\mathbf{H}, \mathbf{M}(\mathbf{A}))}.
\end{aligned}$$

Remark: Elman estimate (aka FOV* bound) [2]

$$\frac{|\langle \mathbf{A}\mathbf{H}\mathbf{y}, \mathbf{y}\rangle_{\mathbf{W}}|^2}{\|\mathbf{A}\mathbf{H}\mathbf{y}\|_{\mathbf{W}}^2\|\mathbf{y}\|_{\mathbf{W}}^2} = \left[\frac{\langle \mathbf{A}\mathbf{H}\mathbf{y}, \mathbf{y}\rangle_{\mathbf{W}}}{\|\mathbf{y}\|_{\mathbf{W}}^2} \times \frac{\|\mathbf{y}\|_{\mathbf{W}}}{\|\mathbf{A}\mathbf{H}\mathbf{y}\|_{\mathbf{W}}}\right]^2 \ge \left[\frac{\text{distance to zero of } \mathbf{W}\text{-FOV}^* \text{ of } \mathbf{A}\mathbf{H}}{\text{norm of } \mathbf{A}\mathbf{H}}\right]^2$$

* FOV: Field Of Values (or numerical range)

Numerical Results

- Convection-Diffusion-Reaction problem discretized by $\mathbb{P}_1$ finite elements
  Find $\mathbf{u} \in H_0^1(\Omega)$ such that: for every $v \in H_0^1(\Omega)$,
  $$\underbrace{\int_\Omega \left(\left(c_0 + \frac{1}{2}\operatorname{div}\mathbf{a}\right) uv + \nu\nabla u\cdot\nabla v\right)}_{\text{Hermitian (symmetric) part } \mathbf{M}(\mathbf{A})} + \int_\Omega\left(\frac{1}{2}\mathbf{a}\cdot\nabla uv - \frac{1}{2}\mathbf{a}\cdot\nabla vu\right) = \int_\Omega fv.$$
  $f(x,y) = \exp(-10((x-0.5)^2 + (y-0.1)^2))$, $\mathbf{a} = 2\pi[-(y-0.1), x-0.5]$,
- $\mathbf{H}$ is a domain decomposition preconditioner for $\mathbf{M}(\mathbf{A})$: Additive Schwarz with GenEO coarse space [5] , $\kappa(\mathbf{H}\mathbf{M}(\mathbf{A})) \le 10$.
- Implemented in Freefem++ with ffddm
  https://doc.freefem.org/documentation/ffddm/index.html

- Scalability is achieved.

$h = 1/200$

Number of subdomains	4	8	16	32
Iteration count	59	59	58	58

$h = 1/500$

Number of subdomains	4	8	16	32
Iteration count	111	113	114	114

- Convergence improves when problem is mildly non-Hermitian.

$c_0 = \nu$	0.1	1	5	10	only symmetric part
Iteration count	> 200	113	36	23	10

Note: c_0 and ν multiply the terms in $\mathbf{M}(\mathbf{A})$ (Hermitian part of the problem).

- Same number of iterates when $\mathbf{W} = $ Identity (see preprint).