Science & Technology Facilities Council

A Multi-preconditioned GMRES Algorithm

Tyrone Rees

Rutherford Appleton Laboratory

IMA Conference on Numerical Linear Algebra and Optimisation
September 2012

Joint work with:
Chen Greif(UBC), Daniel Szyld (Temple University)

The specific problem we are interested in:

Consider solving

$$\mathcal{A}\mathbf{x} = \mathbf{b}$$

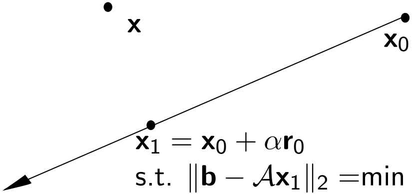
where $\mathcal{A}$ is a large, sparse matrix, using a Krylov subspace method.

Suppose we have two (or more!) possible preconditioners,

$$\mathcal{P}_1 \qquad \text{and} \qquad \mathcal{P}_2,$$

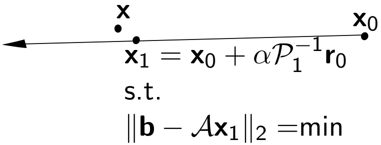
Can we (optimally) combine information from more than one preconditioner?

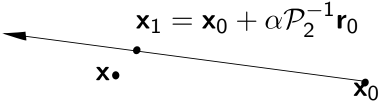
Diagrams



Science & Technology Facilities Council

Diagrams





Science & Technology Facilities Council

Multi-preconditioning

Two relevant methods

Multi-preconditioned conjugate gradients (MPCG): Bridson & Greif (2006)

- Combines multiple preconditioners automatically in a (locally) optimal way
- Requires $\mathcal{A}$ and $\{\mathcal{P}_i\}$ to be symmetric positive definite
- We lose the short-term-recurrence of PCG

Flexible GMRES (FGMRES): Saad (1993)

- Allows variable preconditioners – e.g. $\mathcal{P}_1$ on odd iterations, $\mathcal{P}_2$ on even iterations
- Uses all preconditioners, but nontrivial subspace is being constructed and optimality properties are not fully understood

Science & Technology Facilities Council

Complete Multi-preconditioned Arnoldi

Pick $\mathbf{x}_0$, let $V_1 = \mathbf{r}_0/\|\mathbf{r}_0\|$.
Let $Z_1 = [\mathcal{P}_1^{-1}V_1 \ \cdots \ \mathcal{P}_t^{-1}V_1] \in \mathbb{R}^{n\times t}$

```
for i = 1...max_its
    Q = AZ_i
    for j = 1...i
        H_{j,i} = V_j^T W
        W = W - V_j H_{j,i}
    end
    W = V_{i+1} H_{i+1,i}        (skinny QR factorization)
    Z_{i+1} = [P_1^{-1} V_{i+1} ... P_t^{-1} V_{i+1}]
end
```

Complete Multi-preconditioned Arnoldi

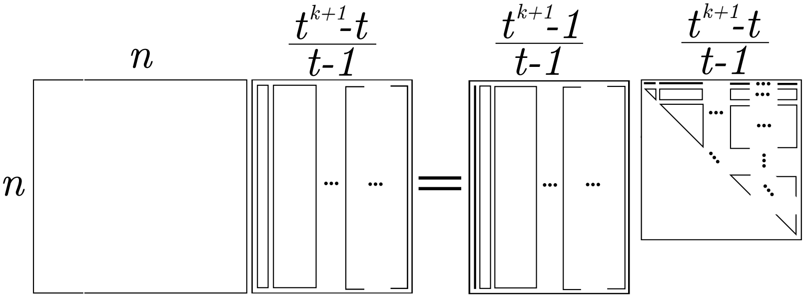
Pick $\mathbf{x}_0$, let $V_1 = \mathbf{r}_0/\|\mathbf{r}_0\|$.
Let $Z_1 = [\mathcal{P}_1^{-1}V_1 \ \cdots \ \mathcal{P}_t^{-1}V_1] \in \mathbb{R}^{n\times t}$
`for` $i = 1 \ldots$ `max_its`
  $Q = AZ_i$
  `for` $j = 1 \ldots i$
    $H_{j,i} = V_j^T W$
    $W = W - V_j H_{j,i}$
  `end`
  $W = V_{i+1}H_{i+1,i}$ (skinny QR factorization)
  $Z_{i+1} = [\mathcal{P}_1^{-1}V_{i+1} \ \cdots \ \mathcal{P}_t^{-1}V_{i+1}]$
`end`

$$\mathcal{A}[Z_1 \ \cdots \ Z_k] = [V_1 \ldots V_{k+1}]\widetilde{H}_k$$

Science & Technology Facilities Council

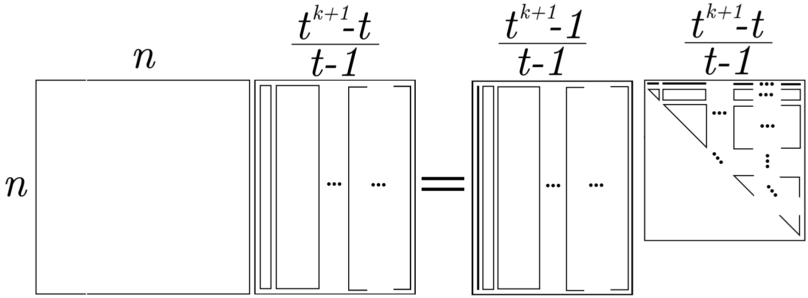
Complete Multipreconditioned Arnoldi

$$\mathcal{A}[Z_1 \ \cdots \ Z_k] = [V_1 \ldots V_{k+1}]\widetilde{H}_k$$



Complete Multipreconditioned Arnoldi

$$\mathcal{A}[Z_1 \ \cdots \ Z_k] = [V_1 \ldots V_{k+1}]\widetilde{H}_k$$



Impractical!

Selective Multi-preconditioned Arnoldi

Pick $\mathbf{x}_0$, let $V_1 = \mathbf{r}_0/\|\mathbf{r}_0\|$.
Let $Z_1 = [\mathcal{P}_1^{-1}V_1 \ \cdots \ \mathcal{P}_t^{-1}V_1] \in \mathbb{R}^{n\times t}$
for $i = 1\ldots$
    $Q = AZ_i$
    for $j = 1 \ldots i$
        $H_{j,i} = V_j^T W$
        $W = W - V_j H_{j,i}$
    end
    $W = V_{i+1}H_{i+1,i}$ (skinny QR factorization)
    $Z_{i+1} = [\mathcal{P}_1^{-1}V_{i+1}^{(1)} \ \cdots \ \mathcal{P}_t^{-1}V_{i+1}^{(t)}]$
end

Selective Multi-preconditioned Arnoldi

Pick $\mathbf{x}_0$, let $V_1 = \mathbf{r}_0/\|\mathbf{r}_0\|$.
Let $Z_1 = [\mathcal{P}_1^{-1}V_1 \ \cdots \ \mathcal{P}_t^{-1}V_1] \in \mathbb{R}^{n\times t}$
for $i = 1 \ldots$
  $Q = AZ_i$
  for $j = 1 \ldots i$
    $H_{j,i} = V_j^T W$
    $W = W - V_j H_{j,i}$
  end
  $W = V_{i+1}H_{i+1,i}$ (skinny QR factorization)
  $Z_{i+1} = [\mathcal{P}_1^{-1}V_{i+1}^{(1)} \ \cdots \ \mathcal{P}_t^{-1}V_{i+1}^{(t)}]$
end

$$\mathcal{A}[Z_1 \ \cdots \ Z_k] = [V_1 \ldots V_{k+1}]\widetilde{H}_k$$

Science & Technology Facilities Council

Various possibilities for generating the search directions

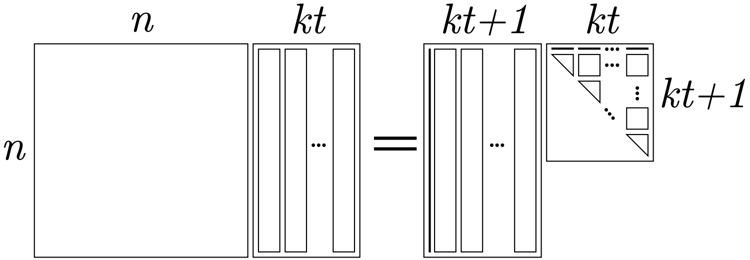
Could replace $Z_{i+1} = [\mathcal{P}_1^{-1} V_{i+1}^{(1)} \ \cdots \ \mathcal{P}_t^{-1} V_{i+1}^{(t)}]$ by taking a mix of all columns, say: $Z_{i+1} = [\mathcal{P}_1^{-1} V_{i+1} \mathbf{1} \ \cdots \ \mathcal{P}_t^{-1} V_{i+1} \mathbf{1}]$, where $\mathbf{1}$ is a vector of all ones.

Practical evidence shows "mixing" is typically more effective; no analytical observations to support this.

Science & Technology Facilities Council

Selective Multi-preconditioned Arnoldi

$$\mathcal{A}[Z_1 \ \cdots \ Z_k] = [V_1 \ldots V_{k+1}]\widetilde{H}_k$$

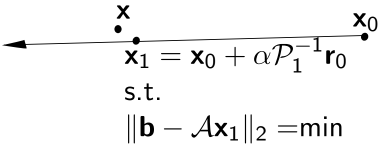


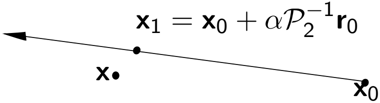
Multi-preconditioned GMRES

Find vector $\mathbf{y}_k$ s.t. $\mathbf{x}_k = \mathbf{x}_0 + Z_k \mathbf{y}_k$
i.e., find $\mathbf{y}_k$ which minimizes

$$\begin{aligned}
\|\mathbf{b} - \mathcal{A}\mathbf{x}_k\|_2 &= \|\mathbf{b} - \mathcal{A}(\mathbf{x}_0 + [Z_1 \cdots Z_k]\mathbf{y}_k)\|_2 \\
&= \|\mathbf{r}_0 - \mathcal{A}[Z_1 \cdots Z_k]\mathbf{y}_k\|_2 \\
&= \|\mathbf{r}_0 - [V_1 \cdots V_{k+1}]\widetilde{H_k}\mathbf{y}_k\|_2 \\
&= \|V_1\|\mathbf{r}_0\|_2 - [V_1 \cdots V_{k+1}]\widetilde{H_k}\mathbf{y}_k\|_2 \\
&= \|[V_1 \cdots V_{k+1}](\|\mathbf{r}_0\|_2\mathbf{e}_1 - \widetilde{H_k}\mathbf{y}_k)\|_2 \\
&= \|\|\mathbf{r}_0\|_2\mathbf{e}_1 - \widetilde{H_k}\mathbf{y}_k\|_2
\end{aligned}$$

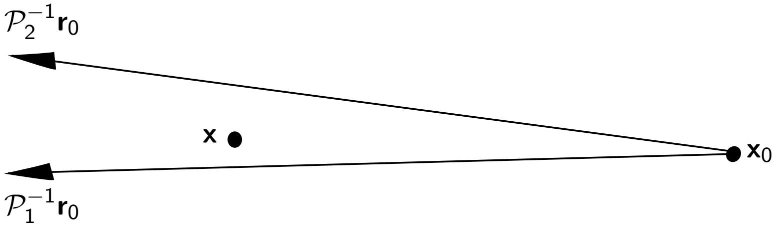
Diagrams





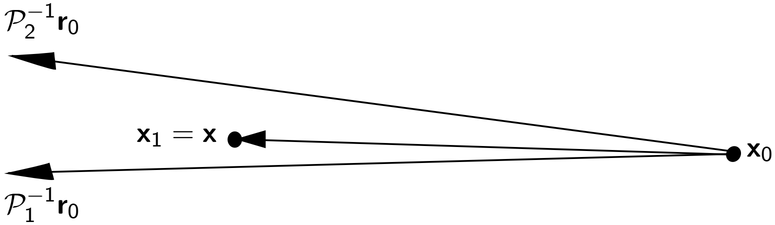
Science & Technology Facilities Council

Diagrams



Science & Technology Facilities Council

Diagrams



Science & Technology Facilities Council

Comparison of costs

Costs at the kth iteration:

	Mat.-vec. prods	inner products	pre. solves
MPGMRES	t^k	$\frac{t^{2k+1}+t^{2k}+t^{k+1}-3t^k}{2(t-1)}$	t^k
tMPGMRES	t	$(k-\frac{1}{2})t^2+\frac{3}{2}t$	t
GMRES	1	k+1	1

Science & Technology Facilities Council

Comparison of costs

Costs at the kth iteration:

	Mat.-vec. prods	inner products	pre. solves
MPGMRES	t^k	$\frac{t^{2k+1}+t^{2k}+t^{k+1}-3t^k}{2(t-1)}$	t^k
tMPGMRES	t	$(k-\frac{1}{2})t^2+\frac{3}{2}t$	t
GMRES	1	k+1	1

Easily parallelized

Characterizing the search space

GMRES preconditioned with a right-preconditioner $\mathcal{P}$ finds the vector that minimizes the 2-norm of the residual over all vectors of the form

$$\mathbf{x}^{(k)} = \mathbf{x}^{(0)} + \mathcal{P}^{-1}\mathbf{y}_k,$$

where $\mathbf{y}_k$ is a member of the Krylov subspace

$$\mathcal{K}_k(\mathcal{A}\mathcal{P}^{-1}, \mathbf{r}^{(0)}) = \mathrm{span}(\mathbf{r}^{(0)}, \mathcal{A}\mathcal{P}^{-1}\mathbf{r}^{(0)}, \ldots, (\mathcal{A}\mathcal{P}^{-1})^{k-1}\mathbf{r}^{(0)}).$$

Science & Technology Facilities Council

Characterizing the search space

The extension for (complete) MPGMRES: the first two iterates satisfy

$$\mathbf{x}^{(1)} - \mathbf{x}^{(0)} \in \operatorname{span}\{\mathcal{P}_1^{-1}\mathbf{r}^{(0)}, \mathcal{P}_2^{-1}\mathbf{r}^{(0)}\}$$

$$\mathbf{x}^{(2)} - \mathbf{x}^{(0)} \in \operatorname{span}\{\mathcal{P}_1^{-1}\mathbf{r}^{(0)}, \mathcal{P}_2^{-1}\mathbf{r}^{(0)}, \mathcal{P}_1^{-1}\mathcal{A}\mathcal{P}_1^{-1}\mathbf{r}^{(0)}, \mathcal{P}_1^{-1}\mathcal{A}\mathcal{P}_2^{-1}\mathbf{r}^{(0)}, \mathcal{P}_2^{-1}\mathcal{A}\mathcal{P}_1^{-1}\mathbf{r}^{(0)}, \mathcal{P}_2^{-1}\mathcal{A}\mathcal{P}_2^{-1}\mathbf{r}^{(0)}\},$$

and the rest follow the same pattern.

Science & Technology Facilities Council

Characterizing the search space

The extension for (complete) MPGMRES: the first two iterates satisfy

$$\begin{aligned}
\mathbf{x}^{(1)} - \mathbf{x}^{(0)} &\in \operatorname{span}\{\mathcal{P}_1^{-1}\mathbf{r}^{(0)}, \mathcal{P}_2^{-1}\mathbf{r}^{(0)}\} \\
&= \operatorname{span}\{X\mathbf{e}^{(0)}, Y\mathbf{e}^{(0)}\} \\
\mathbf{x}^{(2)} - \mathbf{x}^{(0)} &\in \operatorname{span}\{\mathcal{P}_1^{-1}\mathbf{r}^{(0)}, \mathcal{P}_2^{-1}\mathbf{r}^{(0)}, \mathcal{P}_1^{-1}\mathcal{A}\mathcal{P}_1^{-1}\mathbf{r}^{(0)}, \mathcal{P}_1^{-1}\mathcal{A}\mathcal{P}_2^{-1}\mathbf{r}^{(0)}, \\
&\qquad \mathcal{P}_2^{-1}\mathcal{A}\mathcal{P}_1^{-1}\mathbf{r}^{(0)}, \mathcal{P}_2^{-1}\mathcal{A}\mathcal{P}_2^{-1}\mathbf{r}^{(0)}\}, \\
&= \operatorname{span}\{X\mathbf{e}^{(0)}, Y\mathbf{e}^{(0)}, X^2\mathbf{e}^{(0)}, XY\mathbf{e}^{(0)} YX\mathbf{e}^{(0)}, Y^2\mathbf{e}^{(0)}\}
\end{aligned}$$

$$X = \mathcal{P}_1^{-1}\mathcal{A}, \;\; Y = \mathcal{P}_2^{-1}\mathcal{A}$$

and the rest follow the same pattern.

Demonstration of the relative richness of the search space

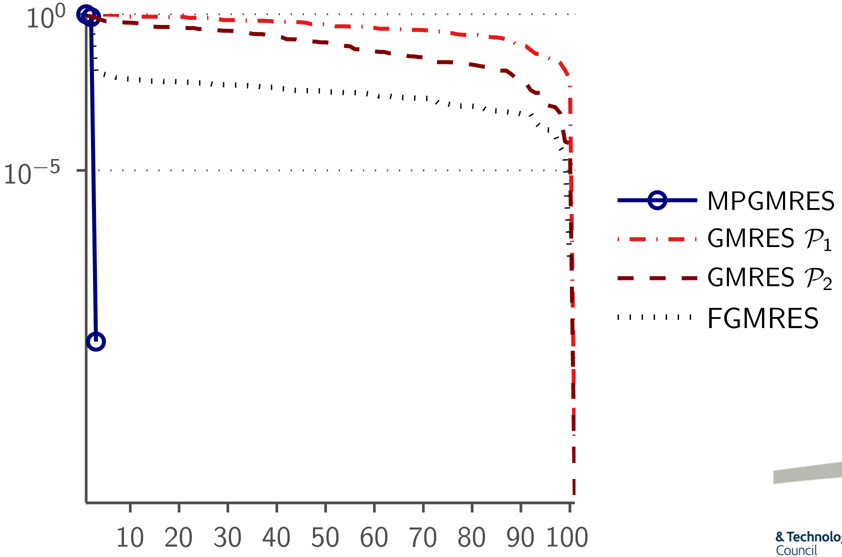
Given $\mathcal{P}_1$ and $\mathcal{P}_2$, take $\mathbf{x}^{(0)} = \mathbf{0}$. Then

$$\mathcal{P}_1^{-1}\mathcal{A}\mathcal{P}_2^{-1}\mathbf{b}$$

lies in the search space after two iterations.

Therefore if $\mathbf{b}$ is an eigenvector of $\mathcal{A}\mathcal{P}_1^{-1}\mathcal{A}\mathcal{P}_2^{-1}$, MPGMRES will converge after two iterations.

Two iterations



& Technology Council

Breakdowns

All breakdowns in standard GMRES are 'lucky'.

This is not the case with MPGMRES....

e.g. if $\mathcal{P}_1 = \mathcal{P}_2$, the matrix $Z_1 = [\mathcal{P}_1^{-1}\mathbf{r}^{(0)}\ \mathcal{P}_2^{-1}\mathbf{r}^{(0)}]$ will be of rank one

Science & Technology Facilities Council

Breakdowns

All breakdowns in standard GMRES are ‘lucky’.

This is not the case with MPGMRES....

e.g. if $\mathcal{P}_1 = \mathcal{P}_2$, the matrix $Z_1 = [\mathcal{P}_1^{-1}\mathbf{r}^{(0)}\ \mathcal{P}_2^{-1}\mathbf{r}^{(0)}]$ will be of rank one

...but not a problem in general – we can monitor the subdiagonal entries of the upper Hessenberg matrix:

- if 0 on subdiagonal, and not converged – must be a linearly dependent vector: discard and all is fine
- if 0 on the subdiagonal, and residual small enough – lucky breakdown!
- if no zero on subdiagonal, no problem!

Examples

Domain decomposition

Consider the advection-diffusion equation on $\Omega = [0,1]^2$:

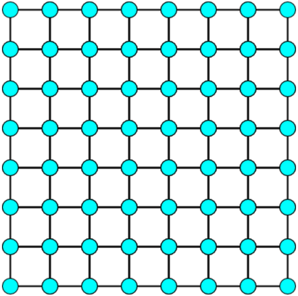
$$
\begin{aligned}
-\nabla^2 u + \boldsymbol{\omega} \cdot \nabla u = f \qquad & \text{in } \Omega \\
u = 0 \qquad & \text{on } \partial\Omega.
\end{aligned}
$$

Upon discretization by finite differences we get the matrix equation

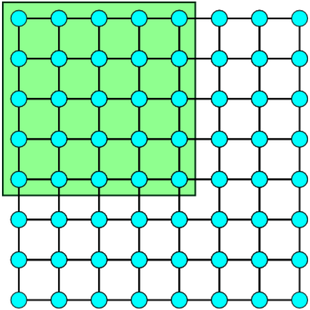
$$A\mathbf{u} = \mathbf{b},$$

where A is a real positive, but nonsymmetric, matrix.

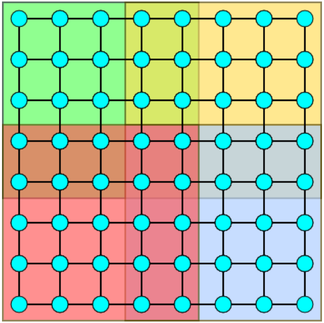
Domain decomposition (cont.)



Domain decomposition (cont.)



Domain decomposition (cont.)



$$A^{(i)} = R_{i,\delta} A R_{i,\delta}^T, i = 1, ..., 4,$$

where $R_{i,\delta}$ is a restriction matrix, and δ denotes the number of nodes overlapping ($\delta = 1$ here).

Science & Technology Facilities Council

Additive Schwarz preconditioner

The additive Schwarz preconditioner has its inverse defined as

$$M^{-1} = \sum_{i=1}^{t} R_{i,\delta}^T (A^{(i)})^{-1} R_{i,\delta}.$$

Additive Schwarz preconditioner

The additive Schwarz preconditioner has its inverse defined as

$$M^{-1} = \sum_{i=1}^{t} R_{i,\delta}^{T} (A^{(i)})^{-1} R_{i,\delta}.$$

Well suited to a multi-preconditioned approach: take each solve on a subdomain as a preconditioner.

Science & Technology Facilities Council

Restricted Additive Schwarz

The restricted Additive Schwarz preconditioner is defined as

$$M^{-1} = \sum_{i=1}^{t} R_{i,0}^{T} (R_{i,\delta} A R_{i,\delta}^{T})^{-1} R_{i,\delta}.$$

This has the effect of removing the overlap in the preconditioner, hence improving convergence.

Also an ideal candidate for multi-preconditioning.

Two subdomains

In the special case where we have two subdomains without overlap, we can show that complete and selective MPGMRES are equivalent.

Two subdomains

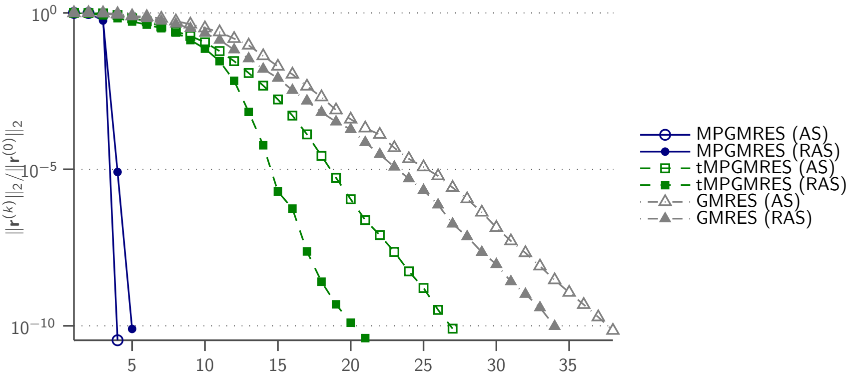
In the special case where we have two subdomains without overlap, we can show that complete and selective MPGMRES are equivalent.

N	sMPGMRES	GMRES	Ratio
2^2	6.59×10^{-3} (5)	9.79×10^{-3} (9)	0.67
2^3	1.86×10^{-2} (8)	1.36×10^{-2} (12)	1.37
2^4	2.79×10^{-2} (11)	4.14×10^{-2} (17)	0.67
2^5	1.08×10^{-1} (16)	1.50×10^{-1} (24)	0.72
2^6	4.66×10^{-1} (19)	7.37×10^{-1} (33)	0.63
2^7	2.78 (25)	4.85 (46)	0.57
2^8	17.1 (30)	34.2 (65)	0.50

Two subdomains

In the special case where we have two subdomains without overlap, we can show that complete and selective MPGMRES are equivalent.

N	sMPGMRES	GMRES	Ratio
2^2	6.59×10^{-3} (5)	9.79×10^{-3} (9)	0.67
2^3	1.86×10^{-2} (8)	1.36×10^{-2} (12)	1.37
2^4	2.79×10^{-2} (11)	4.14×10^{-2} (17)	0.67
2^5	1.08×10^{-1} (16)	1.50×10^{-1} (24)	0.72
2^6	4.66×10^{-1} (19)	7.37×10^{-1} (33)	0.63
2^7	2.78 (25)	4.85 (46)	0.57
2^8	17.1 (30)	34.2 (65)	0.50



Science & Technology Facilities Council

A PDE constrained optimization problem

$$\min_{y,u} \frac{1}{2}||y - \hat{y}||_2^2 + \frac{\beta}{2}||u||_2^2$$

$$\begin{aligned} \text{s.t.} \qquad -\nabla^2 y &= u \text{ in } \Omega \\ y &= f \text{ on } \partial\Omega \end{aligned}$$

Two preconditioners

$$\begin{bmatrix} \beta Q & 0 & -Q \\ 0 & Q & K \\ -Q & K & 0 \end{bmatrix} \begin{bmatrix} \mathbf{u} \\ \mathbf{y} \\ \mathbf{p} \end{bmatrix} = \begin{bmatrix} 0 \\ \mathbf{b} \\ \mathbf{d} \end{bmatrix}$$

$\swarrow$ a choice of preconditioner $\searrow$

$$\mathcal{P}_1 = \begin{bmatrix} \beta Q & 0 & 0 \\ 0 & Q & 0 \\ 0 & 0 & KQ^{-1}K \end{bmatrix}$$

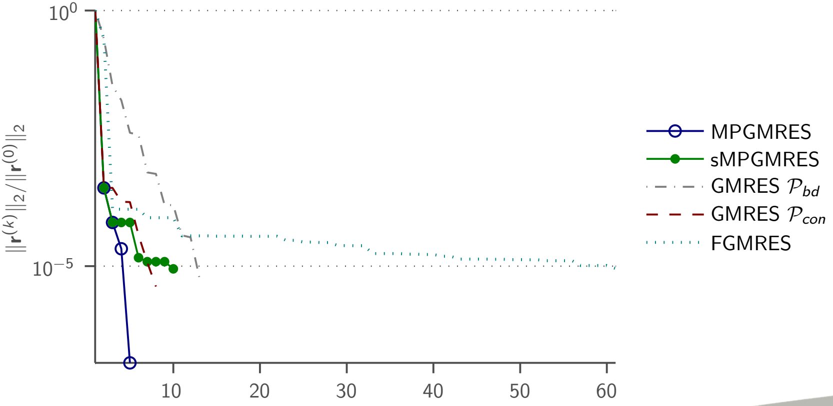
block diagonal

$$\mathcal{P}_2 = \begin{bmatrix} 0 & 0 & -Q \\ 0 & \beta KQ^{-1}K & K \\ -Q & K & 0 \end{bmatrix}$$

constraint

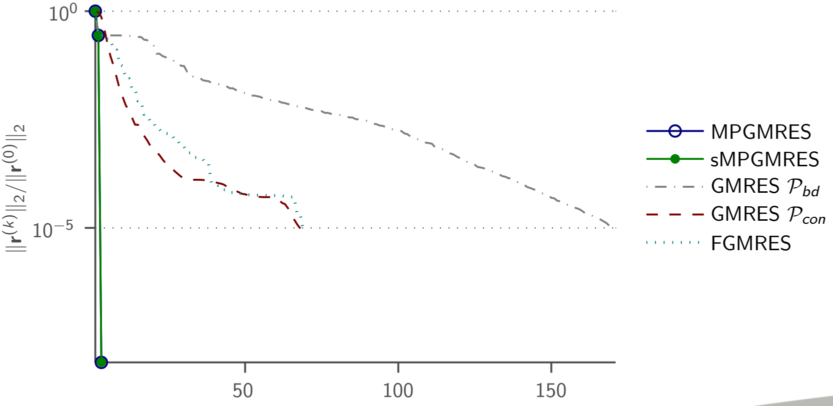
[Rees, Dollar & Wathen (2010)]

$\beta = 10^{-4}$



Science & Technology Facilities Council

$\beta = 10^{-8}$


10^0
10^{-5}
$\|\mathbf{r}^{(k)}\|_2/\|\mathbf{r}^{(0)}\|_2$
50
100
150
MPGMRES
sMPGMRES
GMRES $\mathcal{P}_{bd}$
GMRES $\mathcal{P}_{con}$
FGMRES

Timings

	$\beta = 10^{-4}$	$\beta = 10^{-8}$
complete MPGMRES	10.8	2.1
selective MPGMRES	12.3	1.4
GMRES, $\mathcal{P}_{bd}$	3.5	53.5
GMRES, $\mathcal{P}_{con}$	3.1	28.4
FGMRES	26.4	33.7

Science & Technology Facilities Council

Timings

	$\beta = 10^{-4}$	$\beta = 10^{-8}$
complete MPGMRES	10.8	2.1
selective MPGMRES	12.3	1.4
GMRES, $\mathcal{P}_{bd}$	3.5	53.5
GMRES, $\mathcal{P}_{con}$	3.1	28.4
FGMRES	26.4	33.7

MPGMRES is an extension of the standard preconditioned GMRES which allows us to use more than one preconditioner.

The method:

- seems to work well when we have non-ideal preconditioners which complement each other
- can handle any number of candidate preconditioners
- can be parallelized, obtaining potential computational gains

Paper and MATLAB code available at

`www.numerical.rl.ac.uk/people/rees/`

Fortran 95 code (`HSL_MI28`) under development