The Idea Behind Krylov Methods [1]

Jack Blake

2nd May, 2012

What are they?

Do they make sense?

Can they be practical?

What are they?

Do they make sense?

Can they be practical?

A few definitions before we begin...

The minimal polynomial, $q(t)$, of a matrix A is the unique monic polynomial of minimal degree such that $q(A) = 0$.

A few definitions before we begin...

The minimal polynomial, $q(t)$, of a matrix A is the unique monic polynomial of minimal degree such that $q(A) = 0$.

The index, m_j of an eigenvalue λ_j is the size of the largest associated Jordan block.

A few definitions before we begin...

The minimal polynomial, $q(t)$, of a matrix A is the unique monic polynomial of minimal degree such that $q(A) = 0$.

The index, m_j of an eigenvalue λ_j is the size of the largest associated Jordan block.

A matrix, A, is diagonalisable if there are matrices S and Λ such that $A = S\Lambda S^{-1}$, where Λ is diagonal.

What are Krylov methods?

Want to solve

$$A\mathbf{x} = \mathbf{b}, \quad A \in \mathbb{R}^{n\times n}, \quad \mathbf{x}, \mathbf{b} \in \mathbb{R}^n.$$

What are Krylov methods?

Want to solve

$$A\mathbf{x} = \mathbf{b}, \quad A \in \mathbb{R}^{n \times n}, \quad \mathbf{x}, \mathbf{b} \in \mathbb{R}^n.$$

Iterative method $\quad \rightarrow \quad$ repeatedly approximate $\mathbf{x}$

What are Krylov methods?

Want to solve

$$A\mathbf{x} = \mathbf{b}, \quad A \in \mathbb{R}^{n\times n}, \quad \mathbf{x}, \mathbf{b} \in \mathbb{R}^n.$$

Iterative method $\quad\rightarrow\quad$ repeatedly approximate **x**

Krylov method $\quad\rightarrow\quad$ choose approx$^{\text{n}}$s from a Krylov space.

What are Krylov methods?

Want to solve

$$A\mathbf{x} = \mathbf{b}, \quad A \in \mathbb{R}^{n \times n}, \quad \mathbf{x}, \mathbf{b} \in \mathbb{R}^n.$$

Iterative method $\quad \rightarrow \quad$ repeatedly approximate $\mathbf{x}$

Krylov method $\quad \rightarrow \quad$ choose approx$^{\mathsf{n}}$s from a Krylov space.

"a Krylov method solves $A\mathbf{x} = \mathbf{b}$ by repeatedly performing matrix-vector multiplications involving A" [1]

What is a Krylov space?

Krylov space of dimension k:

$$\mathcal{K}_k\left(A, \mathbf{c}\right) = \operatorname{span}\left\{\mathbf{c}, A\mathbf{c}, \ldots, A^{k-1}\mathbf{c}\right\}.$$

What is a Krylov space?

Krylov space of dimension k:

$$\mathcal{K}_k(A, \mathbf{c}) = \mathrm{span}\left\{\mathbf{c}, A\mathbf{c}, \ldots, A^{k-1}\mathbf{c}\right\}.$$

At iteration k, approximation $\mathbf{x}_k$ is selected from $\mathcal{K}_k(A, \mathbf{c})$.

What is a Krylov space?

Krylov space of dimension k:

$$\mathcal{K}_k(A, \mathbf{c}) = \mathrm{span}\left\{\mathbf{c}, A\mathbf{c}, \ldots, A^{k-1}\mathbf{c}\right\}.$$

At iteration k, approximation $\mathbf{x}_k$ is selected from $\mathcal{K}_k(A, \mathbf{c})$.

Common choice: $\mathbf{c} \equiv \mathbf{r}_0$,
thus: $\mathbf{x}_0 = 0 \Rightarrow \mathbf{c} \equiv \mathbf{b}$

What is a Krylov space?

Krylov space of dimension k:

$$\mathcal{K}_k(A, \mathbf{c}) = \mathrm{span}\left\{\mathbf{c}, A\mathbf{c}, \ldots, A^{k-1}\mathbf{c}\right\}.$$

At iteration k, approximation $\mathbf{x}_k$ is selected from $\mathcal{K}_k(A, \mathbf{c})$.

Common choice: $\mathbf{c} \equiv \mathbf{r}_0$,
thus: $\mathbf{x}_0 = 0 \Rightarrow \mathbf{c} \equiv \mathbf{b}$

Residual: $\mathbf{r}_k \equiv \mathbf{b} - A\mathbf{x}_k$

An example Krylov method

GMRES is a well known Krylov method.

An example Krylov method

GMRES is a well known Krylov method.

Minimises residual $\mathbf{r}_k$ over $\mathcal{K}_k\left(A, \mathbf{b}\right)$:

$$\min_{\mathbf{z}\in\mathcal{K}_k(A,\mathbf{b})} \left\|\mathbf{b} - A\mathbf{z}\right\|$$

An example Krylov method

GMRES is a well known Krylov method.

Minimises residual $\mathbf{r}_k$ over $\mathcal{K}_k(A, \mathbf{b})$:

$$\min_{\mathbf{z} \in \mathcal{K}_k(A,\mathbf{b})} \|\mathbf{b} - A\mathbf{z}\|$$

General Krylov:

number of iterations required depends upon eigenvalues

Questions

Some natural questions arise:

1. Why is $\mathcal{K}_k\left(A, \mathbf{b}\right)$ a good choice?

Questions

Some natural questions arise:

1. Why is $\mathcal{K}_k(A, \mathbf{b})$ a good choice?
2. Why are eigenvalues important?

Questions

Some natural questions arise:

1. Why is $\mathcal{K}_k(A, \mathbf{b})$ a good choice?
2. Why are eigenvalues important?
3. When can we guarantee the solution lies in $\mathcal{K}_k(A, \mathbf{b})$?

Questions

Some natural questions arise:

1. Why is $\mathcal{K}_k(A, \mathbf{b})$ a good choice?
2. Why are eigenvalues important?
3. When can we guarantee the solution lies in $\mathcal{K}_k(A, \mathbf{b})$?

The future of the talk...

What are they?

Do they make sense?

Can they be practical?

Why is $\mathcal{K}_k(A, \mathbf{b})$ a good choice?

Want to solve

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

Take A nonsingular with eigenvalues $\lambda_1, \ldots, \lambda_d$ and respective indices $m_1, \ldots, m_d$.

Why is $\mathcal{K}_k\,(A, \mathbf{b})$ a good choice?

Want to solve

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

Take A nonsingular with eigenvalues $\lambda_1, \ldots, \lambda_d$ and respective indices $m_1, \ldots, m_d$.

Minimal polynomial of A:

$$0 = q(A) = \prod_{j=1}^{d} (A - \lambda_j I)^{m_j}\,.$$

Why is $\mathcal{K}_k(A, \mathbf{b})$ a good choice?

Want to solve

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

Take A nonsingular with eigenvalues $\lambda_1, \ldots, \lambda_d$ and respective indices $m_1, \ldots, m_d$.

Minimal polynomial of A:

$$0 = q(A) = \prod_{j=1}^{d} (A - \lambda_j I)^{m_j} .$$

...BOARD...

Why is $\mathcal{K}_k(A, \mathbf{b})$ a good choice?

Summarising Theorem:

Theorem ([1] p6)

If the minimal polynomial of the nonsingular matrix A has degree m, then the solution to $A\mathbf{x} = \mathbf{b}$ lies in the space $\mathcal{K}_m(A, \mathbf{b})$.

Why is $\mathcal{K}_k(A, \mathbf{b})$ a good choice?

Summarising Theorem:

Theorem ([1] p6)

If the minimal polynomial of the nonsingular matrix A has degree m, then the solution to $A\mathbf{x} = \mathbf{b}$ lies in the space $\mathcal{K}_m(A, \mathbf{b})$.

How could this be better?

Diagonalisable A

Let A be diagonalisable.

Diagonalisable A

Let A be diagonalisable.

$\Rightarrow$ Algebraic multiplicity $=$ geometric multiplicity

Diagonalisable A

Let A be diagonalisable.

$\Rightarrow$ Algebraic multiplicity $=$ geometric multiplicity

$\Rightarrow$ Jordan form is strictly diagonal

Diagonalisable A

Let A be diagonalisable.

$\Rightarrow$ Algebraic multiplicity $=$ geometric multiplicity

$\Rightarrow$ Jordan form is strictly diagonal

$\Rightarrow$ degree of $q(t)$ $=$ number of distinct eigenvalues, d

Diagonalisable A

Let A be diagonalisable.

$\Rightarrow$ Algebraic multiplicity $=$ geometric multiplicity

$\Rightarrow$ Jordan form is strictly diagonal

$\Rightarrow$ degree of $q(t)$ $=$ number of distinct eigenvalues, d

Ammended Theorem:

Theorem ([1] p6 (adapted))

If matrix A is nonsingular and diagonalisable with d distinct eigenvalues, then the solution to $A\mathbf{x} = \mathbf{b}$ lies in the space $\mathcal{K}_d\left(A, \mathbf{b}\right)$.

Preconditioning

Potential problem $\quad \rightarrow \quad$ d close to system size

Preconditioning

Potential problem $\rightarrow$ d close to system size

Solution $\rightarrow$ preconditioning:

$$PA\mathbf{x} = P\mathbf{b}$$

Preconditioning

Potential problem $\rightarrow$ d close to system size

Solution $\rightarrow$ preconditioning:

$$PA\mathbf{x} = P\mathbf{b}$$

- Unfeasible "best case": $P = A^{-1}$

Preconditioning

Potential problem $\quad\rightarrow\quad$ d close to system size

Solution $\quad\rightarrow\quad$ preconditioning:

$$PA\mathbf{x} = P\mathbf{b}$$

- Unfeasible "best case": $P = A^{-1}$
- Good case: PA diagonalisable with small d

Preconditioning

Potential problem $\rightarrow$ d close to system size

Solution $\rightarrow$ preconditioning:

$$PA\mathbf{x} = P\mathbf{b}$$

- Unfeasible "best case": $P = A^{-1}$
- Good case: PA diagonalisable with small d
- Realistic expectation: PA has a few 'clusters' of eigenvalues

What are they?

Do they make sense?

Can they be practical?

One cluster of eigenvalues [2]

Take nonsingular A with one cluster of eigenvalues, radius ρ, centred at 1. I.e. $|1 - \lambda_i| < \rho, \forall i = 1, \ldots, d$.

One cluster of eigenvalues [2]

Take nonsingular A with one cluster of eigenvalues, radius ρ, centred at 1. I.e. $|1 - \lambda_i| < \rho, \forall i = 1, \ldots, d$.

Recall GMRES: $\|\mathbf{r}_k\| = \min_{\mathbf{z} \in \mathcal{K}_k(A,\mathbf{b})} \|\mathbf{b} - A\mathbf{z}\|$

One cluster of eigenvalues [2]

Take nonsingular A with one cluster of eigenvalues, radius ρ, centred at 1. I.e. $|1 - \lambda_i| < \rho, \forall i = 1, \ldots, d$.

Recall GMRES: $\|\mathbf{r}_k\| = \min_{\mathbf{z} \in \mathcal{K}_k(A,\mathbf{b})} \|\mathbf{b} - A\mathbf{z}\|$

$$\mathbf{z} \in \mathcal{K}_k(A, \mathbf{b}) \quad \Rightarrow \quad \mathbf{z} = \sum_{i=0}^{k-1} \alpha_i A^i \mathbf{b}$$

One cluster of eigenvalues [2]

Take nonsingular A with one cluster of eigenvalues, radius ρ, centred at 1. I.e. $|1-\lambda_i| < \rho, \forall i = 1, \ldots, d$.

Recall GMRES: $\|\mathbf{r}_k\| = \min_{\mathbf{z} \in \mathcal{K}_k(A,\mathbf{b})} \|\mathbf{b} - A\mathbf{z}\|$

$$\mathbf{z} \in \mathcal{K}_k(A, \mathbf{b}) \quad \Rightarrow \quad \mathbf{z} = \sum_{i=0}^{k-1} \alpha_i A^i \mathbf{b}$$

Thus

$$\|\mathbf{r}_k\| = \min_{\alpha_i} \left\| \left(I - \sum_{i=0}^{k-1} \alpha_i A^{i+1} \right) \mathbf{b} \right\|$$

One cluster of eigenvalues [2]

Take nonsingular A with one cluster of eigenvalues, radius ρ, centred at 1. I.e. $|1-\lambda_i| < \rho, \forall i = 1, \ldots, d$.

Recall GMRES: $\|\mathbf{r}_k\| = \min_{\mathbf{z}\in\mathcal{K}_k(A,\mathbf{b})} \|\mathbf{b} - A\mathbf{z}\|$

$$\mathbf{z} \in \mathcal{K}_k(A, \mathbf{b}) \quad \Rightarrow \quad \mathbf{z} = \sum_{i=0}^{k-1} \alpha_i A^i \mathbf{b}$$

Thus

$$\begin{aligned} \|\mathbf{r}_k\| &= \min_{\alpha_i} \left\| \left(I - \sum_{i=0}^{k-1} \alpha_i A^{i+1} \right) \mathbf{b} \right\| \\ &= \min_{p\in\mathcal{P}_k} \|p(A)\mathbf{b}\| \end{aligned}$$

One cluster of eigenvalues [2]

Take nonsingular A with one cluster of eigenvalues, radius ρ, centred at 1. I.e. $|1-\lambda_i| < \rho, \forall i = 1, \ldots, d$.

Recall GMRES: $\|\mathbf{r}_k\| = \min_{\mathbf{z}\in\mathcal{K}_k(A,\mathbf{b})} \|\mathbf{b} - A\mathbf{z}\|$

$$\mathbf{z} \in \mathcal{K}_k(A, \mathbf{b}) \quad \Rightarrow \quad \mathbf{z} = \sum_{i=0}^{k-1} \alpha_i A^i \mathbf{b}$$

Thus

$$\begin{aligned}
\|\mathbf{r}_k\| &= \min_{\alpha_i} \left\| \left(I - \sum_{i=0}^{k-1} \alpha_i A^{i+1} \right) \mathbf{b} \right\| \\
&= \min_{p\in\mathcal{P}_k} \|p(A)\mathbf{b}\| \\
&\leq \|p_{any}(A)\mathbf{b}\|
\end{aligned}$$

One cluster of eigenvalues [2]

Take nonsingular A with one cluster of eigenvalues, radius ρ, centred at 1. I.e. $|1-\lambda_i| < \rho, \forall i = 1, \dots, d$.

Recall GMRES: $\|\mathbf{r}_k\| = \min_{\mathbf{z}\in\mathcal{K}_k(A,\mathbf{b})} \|\mathbf{b} - A\mathbf{z}\|$

$$\mathbf{z} \in \mathcal{K}_k(A, \mathbf{b}) \quad \Rightarrow \quad \mathbf{z} = \sum_{i=0}^{k-1} \alpha_i A^i \mathbf{b}$$

Thus

$$\begin{aligned}
\|\mathbf{r}_k\| &= \min_{\alpha_i} \left\| \left(I - \sum_{i=0}^{k-1} \alpha_i A^{i+1} \right) \mathbf{b} \right\| \\
&= \min_{p\in\mathcal{P}_k} \|p(A)\mathbf{b}\| \\
&\leq \|p_{jack}(A)\mathbf{b}\|
\end{aligned}$$

One cluster of eigenvalues [2]

It is at least true that

$$\|\mathbf{r}_k\| \leq \left\|(I - A)^k \mathbf{b}\right\|$$

One cluster of eigenvalues [2]

It is at least true that

$$\|\mathbf{r}_k\| \leq \left\|(I - A)^k \mathbf{b}\right\| \leq \left\|(I - A)^k\right\| \|\mathbf{b}\| .$$

One cluster of eigenvalues [2]

It is at least true that

$$\|\mathbf{r}_k\| \leq \left\|(I - A)^k \mathbf{b}\right\| \leq \left\|(I - A)^k\right\| \|\mathbf{b}\| .$$

If A diagonalisable:

$$A = S\Lambda S^{-1} \quad \Rightarrow \quad (I - A)^k = S(I - \Lambda)^k S^{-1},$$

One cluster of eigenvalues [2]

It is at least true that

$$\|\mathbf{r}_k\| \leq \left\|(I - A)^k \mathbf{b}\right\| \leq \left\|(I - A)^k\right\| \|\mathbf{b}\| .$$

If A diagonalisable:

$$A = S \Lambda S^{-1} \quad \Rightarrow \quad (I - A)^k = S (I - \Lambda)^k S^{-1},$$

thus

$$\left\|(I - A)^k\right\| \leq \|S\| \, \| (I - \Lambda)^k \| \|S^{-1}\|.$$

One cluster of eigenvalues [2]

It is at least true that

$$\|\mathbf{r}_k\| \leq \left\|(I - A)^k \mathbf{b}\right\| \leq \left\|(I - A)^k\right\| \|\mathbf{b}\| .$$

If A diagonalisable:

$$A = S \Lambda S^{-1} \quad \Rightarrow \quad (I - A)^k = S(I - \Lambda)^k S^{-1},$$

thus

$$\begin{aligned} \left\|(I - A)^k\right\| &\leq \|S\| \, \| (I - \Lambda)^k \| \|S^{-1}\|, \\ &= \|S\| \, \|S^{-1}\| \left(\max_i |1 - \lambda_i| \right)^k . \end{aligned}$$

One cluster of eigenvalues [2]

It is at least true that

$$\|\mathbf{r}_k\| \leq \left\|(I - A)^k \mathbf{b}\right\| \leq \left\|(I - A)^k\right\| \|\mathbf{b}\| .$$

If A diagonalisable:

$$A = S\Lambda S^{-1} \quad \Rightarrow \quad (I - A)^k = S(I - \Lambda)^k S^{-1},$$

thus

$$\begin{aligned} \left\|(I - A)^k\right\| &\leq \|S\| \, \| (I - \Lambda)^k \| \|S^{-1}\|, \\ &= \|S\| \, \|S^{-1}\| \underbrace{\left(\max_i |1 - \lambda_i|\right)^k}_{\leq \rho^k} . \end{aligned}$$

One cluster of eigenvalues [2]

It is at least true that

$$\|\mathbf{r}_k\| \leq \left\|(I-A)^k \mathbf{b}\right\| \leq \left\|(I-A)^k\right\| \|\mathbf{b}\| .$$

If A diagonalisable:

$$A = S\Lambda S^{-1} \quad \Rightarrow \quad (I-A)^k = S(I-\Lambda)^k S^{-1},$$

thus

$$\begin{aligned} \left\|(I-A)^k\right\| &\leq \|S\| \, \|(I-\Lambda)^k\| \|S^{-1}\|, \\ &= \|S\| \, \|S^{-1}\| \underbrace{\left(\max_i |1-\lambda_i|\right)^k}_{\leq \rho^k}. \end{aligned}$$

So

$$\|\mathbf{r}_k\| \leq \|S\| \, \|S^{-1}\| \, \|\mathbf{b}\| \, \rho^k.$$

One cluster of eigenvalues and one outlier [2]

Now let the eigenvalues of A satisfy:

$$|1 - \lambda_i| < \rho, \forall i = 2, \ldots, d, \quad \text{and} \quad |1 - \lambda_1| \gg \rho.$$

One cluster of eigenvalues and one outlier [2]

Now let the eigenvalues of A satisfy:

$$|1-\lambda_i| < \rho, \forall i = 2, \dots, d, \quad \text{and} \quad |1-\lambda_1| \gg \rho.$$

Take

$$p_{jack} = \left(I - \frac{A}{\lambda_1}\right)(I-A)^{k-1}.$$

One cluster of eigenvalues and one outlier [2]

Now let the eigenvalues of A satisfy:

$$|1-\lambda_i| < \rho, \forall i = 2, \ldots, d, \quad \text{and} \quad |1-\lambda_1| \gg \rho.$$

Take

$$p_{jack} = \left(I - \frac{A}{\lambda_1}\right)(I-A)^{k-1}.$$

So

$$\|\mathbf{r}_k\| \leq \left\|\left(I - \frac{A}{\lambda_1}\right)(I-A)^{k-1}\mathbf{b}\right\|.$$

One cluster of eigenvalues and one outlier [2]

Now let the eigenvalues of A satisfy:

$$|1-\lambda_i| < \rho, \forall i = 2, \ldots, d, \quad \text{and} \quad |1-\lambda_1| \gg \rho.$$

Take

$$p_{jack} = \left(I - \frac{A}{\lambda_1}\right)(I-A)^{k-1}.$$

So

$$\|\mathbf{r}_k\| \leq \left\| \underbrace{\left(I - \frac{A}{\lambda_1}\right)(I-A)^{k-1}\mathbf{b}}_{\text{...BOARD...}} \right\|.$$

One cluster of eigenvalues and one outlier [2]

Now let the eigenvalues of A satisfy:

$$|1-\lambda_i| < \rho, \forall i = 2,\ldots,d, \quad \text{and} \quad |1-\lambda_1| \gg \rho.$$

Take

$$p_{jack} = \left(I - \frac{A}{\lambda_1}\right)(I-A)^{k-1}.$$

So

$$\|\mathbf{r}_k\| \;\leq\; \left\|\left(I - \frac{A}{\lambda_1}\right)(I-A)^{k-1}\,\mathbf{b}\right\|,$$

$$\leq\; C\rho^{k-1}$$

Numerical Example

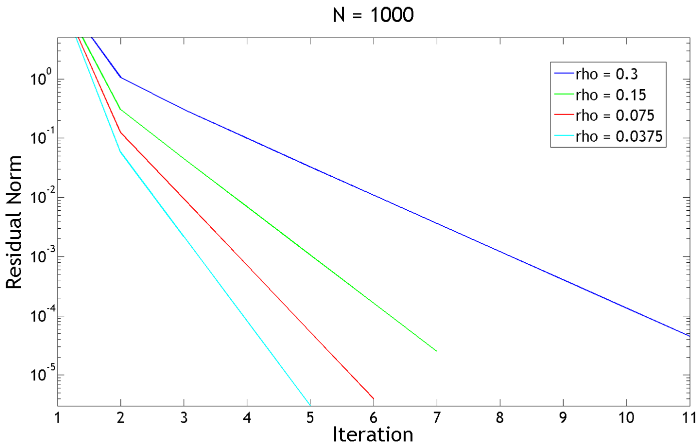
Take

$$A = \begin{pmatrix} 1 & -\rho & & & \\ -\rho & 1 & -\rho & & \\ & -\rho & 1 & -\rho & \\ & & \ddots & \ddots & \ddots \end{pmatrix}^{N \times N}$$

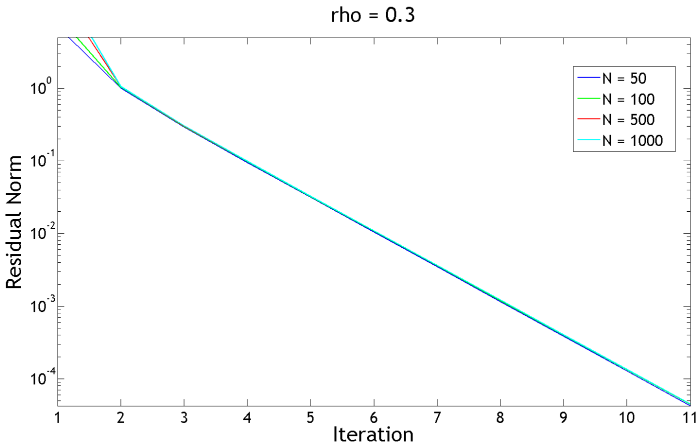
which has eigenvalues

$$\lambda_i = 2\cos\left(\frac{i\pi}{N+1}\right)$$

Numerical Example



Numerical Example



I. C. F. Ipsen and C. D. Meyer.
The idea behind Krylov methods.
Amer. Math. Monthly, 105(10):889–899, 1998.

S. L. Campbell, I. C. F. Ipsen, C. T. Kelley, and C. D. Meyer.
Gmres and the minimal polynomial.
BIT, 36:32–43, 1996.