THE UNIVERSITY OF RHODE ISLAND

GMRES - An Iterative Method for Solving Large Linear Systems

Kyle Monette

Department of Mathematics and Applied Mathematical Sciences
University of Rhode Island

Applied Mathematics and Scientific Computing Seminar

April 29, 2024

The Goal

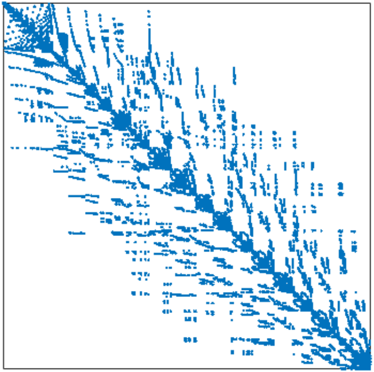
Our goal is to solve a large dimensional linear system:

$$Ax = b, \qquad A \in \mathbb{R}^{n\times n}, \quad b \in \mathbb{R}^{n}, \quad n \text{ is large} \tag{1}$$

The Goal

Our goal is to solve a large dimensional linear system:

$$Ax = b, \qquad A \in \mathbb{R}^{n \times n}, \quad b \in \mathbb{R}^n, \quad n \text{ is large} \tag{1}$$



Number of nonzero elements: 2.71E+07

Size $1,505,785$

Time for A^{-1} is ?????

Time for LU is ?????

Approximation in 1.67 s

Error of 10^{-11}

The Goal

Our goal is to solve a large dimensional linear system:

$$Ax = b, \qquad A \in \mathbb{R}^{n \times n}, \quad b \in \mathbb{R}^n, \quad n \text{ is large} \tag{1}$$

Instead, we solve the following problem:

$$\text{For a given } \varepsilon, \qquad \text{find } \widetilde{x} \text{ such that} \qquad \|A\widetilde{x} - b\|_2 < \varepsilon \tag{2}$$

For approximate $\widetilde{x}$, let $r = A\widetilde{x} - b$ the residual.

The Goal

Our goal is to solve a large dimensional linear system:

$$Ax = b, \qquad A \in \mathbb{R}^{n\times n}, \quad b \in \mathbb{R}^n, \quad n \text{ is large} \tag{1}$$

Instead, we solve the following problem:

$$\text{For a given } \varepsilon, \qquad \text{find } \widetilde{x} \text{ such that} \qquad \|A\widetilde{x} - b\|_2 < \varepsilon \tag{2}$$

For approximate $\widetilde{x}$, let $r = A\widetilde{x} - b$ the residual.

Iterative Methods: use an initial value to generate a sequence of approximations (hopefully!) improving in accuracy.

The Goal

Our goal is to solve a large dimensional linear system:

$$Ax = b, \qquad A \in \mathbb{R}^{n\times n}, \quad b \in \mathbb{R}^n, \quad n \text{ is large} \tag{1}$$

Instead, we solve the following problem:

$$\text{For a given } \varepsilon, \qquad \text{find } \widetilde{x} \text{ such that} \qquad \|A\widetilde{x} - b\|_2 < \varepsilon \tag{2}$$

For approximate $\widetilde{x}$, let $r = A\widetilde{x} - b$ the residual.

Iterative Methods: use an initial value to generate a sequence of approximations (hopefully!) improving in accuracy.

Direct Methods: (exactly) solve the problem with a finite sequence of operations.

Approximations in Subspaces

Consider approximations $\widetilde{x}$ to $Ax = b$ in an affine subspace:

$$\widetilde{x} \in x_0 + \mathcal{S}_m, \qquad \dim \mathcal{S}_m = m,$$

for some initial guess x_0.

Approximations in Subspaces

Consider approximations $\widetilde{x}$ to $Ax = b$ in an affine subspace:

$$\widetilde{x} \in x_0 + \mathcal{S}_m, \qquad \dim \mathcal{S}_m = m,$$

for some initial guess x_0.

We'll require that $b - A\widetilde{x} \perp A\mathcal{S}_m$, for reasons to be seen.

Approximations in Subspaces

Consider approximations $\widetilde{x}$ to $Ax = b$ in an affine subspace:

$$\widetilde{x} \in x_0 + \mathcal{S}_m, \qquad \dim \mathcal{S}_m = m,$$

for some initial guess x_0.

We'll require that $b - A\widetilde{x} \perp A\mathcal{S}_m$, for reasons to be seen.

At each $m = 1, 2, \ldots$ search the subspace $\mathcal{S}_m$ for $\widetilde{x}$ such that

$$\|A\widetilde{x} - b\|_2 < \text{ tolerance.}$$

If satisfied, stop. If not, continue.

Approximations in Subspaces

Consider approximations $\widetilde{x}$ to $Ax = b$ in an affine subspace:

$$\widetilde{x} \in x_0 + \mathcal{S}_m, \qquad \dim \mathcal{S}_m = m,$$

for some initial guess x_0.

We'll require that $b - A\widetilde{x} \perp A\mathcal{S}_m$, for reasons to be seen.

At each $m = 1, 2, \ldots$ search the subspace $\mathcal{S}_m$ for $\widetilde{x}$ such that

$$\|A\widetilde{x} - b\|_2 < \text{ tolerance.}$$

If satisfied, stop. If not, continue.

What subspaces $\mathcal{S}_m$ do we consider?

Background

Theorem 1 (Cayley-Hamilton)

Let $A \in \mathbb{R}^{n\times n}$ and let $q(\lambda)$ denote the characteristic polynomial of A. Then $q(A) = 0_{n\times n}$.

Background

Theorem 1 (Cayley-Hamilton)

Let $A \in \mathbb{R}^{n \times n}$ and let $q(\lambda)$ denote the characteristic polynomial of A. Then $q(A) = 0_{n \times n}$.

If A is nonsingular and $q(\lambda) = \alpha_0 + \alpha_1 \lambda + \cdots + \alpha_{n-1}\lambda^{n-1} + \underbrace{(-1)^n}_{:=\alpha_n} \lambda^n$,

$$0 \quad = \quad \alpha_0 I + \alpha_1 A + \cdots + \alpha_{n-1} A^{n-1} + \alpha_n A^n$$

Background

Theorem 1 (Cayley-Hamilton)

Let $A \in \mathbb{R}^{n\times n}$ and let $q(\lambda)$ denote the characteristic polynomial of A. Then $q(A) = 0_{n\times n}$.

If A is nonsingular and $q(\lambda) = \alpha_0 + \alpha_1\lambda + \cdots + \alpha_{n-1}\lambda^{n-1} + \underbrace{(-1)^n}_{:=\alpha_n}\lambda^n$,

$$
\begin{aligned}
0 &= \alpha_0 I + \alpha_1 A + \cdots + \alpha_{n-1}A^{n-1} + \alpha_n A^n \\
&= \left(\alpha_0 I + \alpha_1 A + \cdots + \alpha_{n-1}A^{n-1} + \alpha_n A^n\right) \cdot A^{-1}
\end{aligned}
$$

Background

Theorem 1 (Cayley-Hamilton)

Let $A \in \mathbb{R}^{n \times n}$ and let $q(\lambda)$ denote the characteristic polynomial of A. Then $q(A) = 0_{n \times n}$.

If A is nonsingular and $q(\lambda) = \alpha_0 + \alpha_1 \lambda + \cdots + \alpha_{n-1}\lambda^{n-1} + \underbrace{(-1)^n}_{:=\alpha_n} \lambda^n$,

$$\begin{aligned} 0 &= \alpha_0 I + \alpha_1 A + \cdots + \alpha_{n-1} A^{n-1} + \alpha_n A^n \\ &= \left(\alpha_0 I + \alpha_1 A + \cdots + \alpha_{n-1} A^{n-1} + \alpha_n A^n\right) \cdot A^{-1} \\ &= \alpha_0 A^{-1} + \alpha_1 I + \ldots \alpha_{n-1} A^{n-2} + \alpha_n A^{n-1} \end{aligned}$$

Theorem 1 (Cayley-Hamilton)

Let $A \in \mathbb{R}^{n \times n}$ and let $q(\lambda)$ denote the characteristic polynomial of A. Then $q(A) = 0_{n \times n}$.

If A is nonsingular and $q(\lambda) = \alpha_0 + \alpha_1 \lambda + \cdots + \alpha_{n-1}\lambda^{n-1} + \underbrace{(-1)^n}_{:=\alpha_n} \lambda^n$,

$$
\begin{aligned}
0 &= \alpha_0 I + \alpha_1 A + \cdots + \alpha_{n-1} A^{n-1} + \alpha_n A^n \\
&= \left(\alpha_0 I + \alpha_1 A + \cdots + \alpha_{n-1} A^{n-1} + \alpha_n A^n\right) \cdot A^{-1} \\
&= \alpha_0 A^{-1} + \alpha_1 I + \ldots \alpha_{n-1} A^{n-2} + \alpha_n A^{n-1} \\
A^{-1} &= -\frac{1}{\alpha_0}\left(\alpha_1 I + \alpha_2 A + \ldots \alpha_{n-1} A^{n-2} + \alpha_n A^{n-1}\right)
\end{aligned}
$$

Background

Theorem 1 (Cayley-Hamilton)

Let $A \in \mathbb{R}^{n \times n}$ and let $q(\lambda)$ denote the characteristic polynomial of A. Then $q(A) = 0_{n \times n}$.

If A is nonsingular and $q(\lambda) = \alpha_0 + \alpha_1 \lambda + \cdots + \alpha_{n-1}\lambda^{n-1} + \underbrace{(-1)^n}_{:=\alpha_n} \lambda^n$,

$$
\begin{aligned}
0 &= \alpha_0 I + \alpha_1 A + \cdots + \alpha_{n-1} A^{n-1} + \alpha_n A^n \\
&= \left(\alpha_0 I + \alpha_1 A + \cdots + \alpha_{n-1} A^{n-1} + \alpha_n A^n\right) \cdot A^{-1} \\
&= \alpha_0 A^{-1} + \alpha_1 I + \ldots \alpha_{n-1} A^{n-2} + \alpha_n A^{n-1} \\
A^{-1} &= -\frac{1}{\alpha_0}\left(\alpha_1 I + \alpha_2 A + \ldots \alpha_{n-1} A^{n-2} + \alpha_n A^{n-1}\right) \\
A^{-1} b &= -\frac{1}{\alpha_0}\left(\alpha_1 b + \alpha_2 A b + \ldots \alpha_{n-1} A^{n-2} b + \alpha_n A^{n-1} b\right)
\end{aligned}
$$

Background

We choose the "search subspaces" $\mathcal{S}_m$ to be

$$\text{span}\left\{b, Ab, A^2b, \ldots, A^{m-1}b\right\}, \quad 1 \leq m \leq n.$$

Or with an initial guess x_0 and residual $r_0 = b - Ax_0$,

$$\text{span}\left\{r_0, Ar_0, A^2r_0, \ldots, A^{m-1}r_0\right\}, \quad 1 \leq m \leq n.$$

Background

We choose the "search subspaces" $\mathcal{S}_m$ to be

$$\operatorname{span}\left\{b, Ab, A^2b, \ldots, A^{m-1}b\right\}, \quad 1 \le m \le n.$$

Or with an initial guess x_0 and residual $r_0 = b - Ax_0$,

$$\operatorname{span}\left\{r_0, Ar_0, A^2r_0, \ldots, A^{m-1}r_0\right\}, \quad 1 \le m \le n.$$

Definition 2 (Krylov Subspace)

For $A_{n\times n}$, the m-th Krylov subspace $\mathcal{K}_m(A, r_0)$ is defined

$$\mathcal{K}_m(A, r_0) = \mathcal{K}_m = \operatorname{span}\left\{r_0, Ar_0, A^2r_0, \ldots, A^{m-1}r_0\right\}.$$

Properties of $\mathcal{K}_m(A, r_0) = \text{span}\{r_0, Ar_0, \ldots, A^{m-1}r_0\}$

- Nested subspaces: $\mathcal{K}_1 \subset \mathcal{K}_2 \subset \cdots \subset \mathcal{K}_n$

Properties of $\mathcal{K}_m(A, r_0) = \text{span}\{r_0, Ar_0, \ldots, A^{m-1}r_0\}$

- Nested subspaces: $\mathcal{K}_1 \subset \mathcal{K}_2 \subset \cdots \subset \mathcal{K}_n$

- $A\mathcal{K}_j \subset \mathcal{K}_{j+1}$

Properties of $\mathcal{K}_m(A, r_0) = \text{span}\left\{r_0, Ar_0, \ldots, A^{m-1}r_0\right\}$

- Nested subspaces: $\mathcal{K}_1 \subset \mathcal{K}_2 \subset \cdots \subset \mathcal{K}_n$
- $A\mathcal{K}_j \subset \mathcal{K}_{j+1}$
- If μ is such that $\mathcal{K}_\mu = \mathcal{K}_{\mu+1}$, then:

Properties of $\mathcal{K}_m(A, r_0) = \text{span}\{r_0, Ar_0, \ldots, A^{m-1}r_0\}$

- Nested subspaces: $\mathcal{K}_1 \subset \mathcal{K}_2 \subset \cdots \subset \mathcal{K}_n$
- $A\mathcal{K}_j \subset \mathcal{K}_{j+1}$
- If μ is such that $\mathcal{K}_\mu = \mathcal{K}_{\mu+1}$, then:
    - $A\mathcal{K}_\mu \subset \mathcal{K}_\mu$ therefore $\mathcal{K}_\mu$ is A-invariant

Properties of $\mathcal{K}_m(A, r_0) = \text{span}\{r_0, Ar_0, \ldots, A^{m-1}r_0\}$

- Nested subspaces: $\mathcal{K}_1 \subset \mathcal{K}_2 \subset \cdots \subset \mathcal{K}_n$

- $A\mathcal{K}_j \subset \mathcal{K}_{j+1}$

- If μ is such that $\mathcal{K}_\mu = \mathcal{K}_{\mu+1}$, then:
  - $A\mathcal{K}_\mu \subset \mathcal{K}_\mu$ therefore $\mathcal{K}_\mu$ is A-invariant
  - $\mathcal{K}_1 \subset \mathcal{K}_2 \subset \cdots \subset \mathcal{K}_\mu = \mathcal{K}_{\mu+1} = \cdots = \mathcal{K}_n$

Properties of $\mathcal{K}_m(A, r_0) = \text{span}\{r_0, Ar_0, \dots, A^{m-1}r_0\}$

- Nested subspaces: $\mathcal{K}_1 \subset \mathcal{K}_2 \subset \cdots \subset \mathcal{K}_n$

- $A\mathcal{K}_j \subset \mathcal{K}_{j+1}$

- If μ is such that $\mathcal{K}_\mu = \mathcal{K}_{\mu+1}$, then:
  - $A\mathcal{K}_\mu \subset \mathcal{K}_\mu$ therefore $\mathcal{K}_\mu$ is A-invariant
  - $\mathcal{K}_1 \subset \mathcal{K}_2 \subset \cdots \subset \mathcal{K}_\mu = \mathcal{K}_{\mu+1} = \cdots = \mathcal{K}_n$
  - Therefore, the exact solution is in $\mathcal{K}_\mu$.

Properties of $\mathcal{K}_m(A, r_0) = \text{span}\{r_0, Ar_0, \ldots, A^{m-1}r_0\}$

- We do not work with $\{r_0, Ar_0, \ldots, A^{m-1}r_0\}$ directly.

Properties of $\mathcal{K}_m(A, r_0) = \text{span}\left\{r_0, Ar_0, \ldots, A^{m-1}r_0\right\}$

- We do not work with $\left\{r_0, Ar_0, \ldots, A^{m-1}r_0\right\}$ directly.
  - Linear independence is not guaranteed (e.g., $A = 3I$)

Properties of $\mathcal{K}_m(A, r_0) = \text{span}\{r_0, Ar_0, \ldots, A^{m-1}r_0\}$

- We do not work with $\{r_0, Ar_0, \ldots, A^{m-1}r_0\}$ directly.
  - Linear independence is not guaranteed (e.g., $A = 3I$)
  - Poorly conditioned vectors: consider $n = 1000$ and $m = 10$.

    $K = \begin{bmatrix} r_0 & Ar_0 & \ldots & A^{m-1}r_0 \end{bmatrix}$ has condition number $\kappa(K) = 6 \times 10^{28}$
  - The vectors lose linear independence

Properties of $\mathcal{K}_m(A, r_0) = \text{span}\{r_0, Ar_0, \ldots, A^{m-1}r_0\}$

- We do not work with $\{r_0, Ar_0, \ldots, A^{m-1}r_0\}$ directly.
  - Linear independence is not guaranteed (e.g., $A = 3I$)
  - Poorly conditioned vectors: consider $n = 1000$ and $m = 10$.

    $K = \begin{bmatrix} r_0 & Ar_0 & \ldots & A^{m-1}r_0 \end{bmatrix}$ has condition number $\kappa(K) = 6 \times 10^{28}$
  - The vectors lose linear independence

Let's construct a basis for $\mathcal{K}_m$.

Overview of (Modified) Gram-Schmidt

- Let $\mathcal{B} = \{u_1, \ldots, u_n\}$ be a basis for a subspace $\mathcal{M}$.
- Create an *orthonormal basis* $\mathcal{O} = \{v_1, \ldots, v_n\}$ for $\mathcal{M}$.

Overview of (Modified) Gram-Schmidt

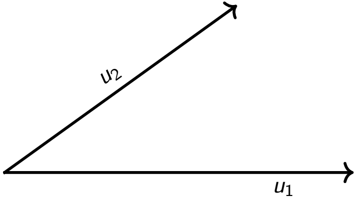
- Let $\mathcal{B} = \{u_1, \ldots, u_n\}$ be a basis for a subspace $\mathcal{M}$.
- Create an *orthonormal basis* $\mathcal{O} = \{v_1, \ldots, v_n\}$ for $\mathcal{M}$.

$n = 2$: $\mathcal{B} = \{u_1, u_2\}$ and $\mathcal{O} = \{v_1, v_2\}$.

Overview of (Modified) Gram-Schmidt

- Let $\mathcal{B} = \{u_1, \ldots, u_n\}$ be a basis for a subspace $\mathcal{M}$.
- Create an *orthonormal basis* $\mathcal{O} = \{v_1, \ldots, v_n\}$ for $\mathcal{M}$.

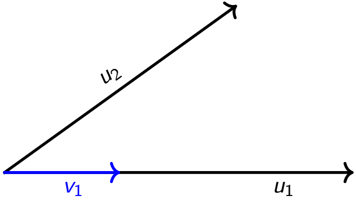
$n = 2$: $\mathcal{B} = \{u_1, u_2\}$ and $\mathcal{O} = \{v_1, v_2\}$.



Overview of (Modified) Gram-Schmidt

- Let $\mathcal{B} = \{u_1, \ldots, u_n\}$ be a basis for a subspace $\mathcal{M}$.
- Create an *orthonormal basis* $\mathcal{O} = \{v_1, \ldots, v_n\}$ for $\mathcal{M}$.

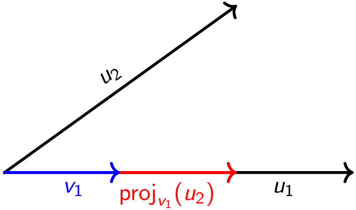
$n = 2$: $\mathcal{B} = \{u_1, u_2\}$ and $\mathcal{O} = \{v_1, v_2\}$.



Overview of (Modified) Gram-Schmidt

- Let $\mathcal{B} = \{u_1, \dots, u_n\}$ be a basis for a subspace $\mathcal{M}$.
- Create an *orthonormal basis* $\mathcal{O} = \{v_1, \dots, v_n\}$ for $\mathcal{M}$.

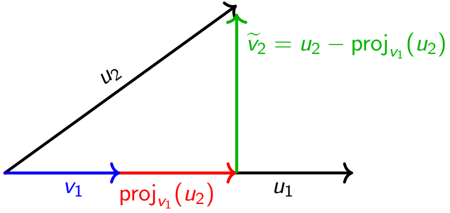
$n = 2$: $\mathcal{B} = \{u_1, u_2\}$ and $\mathcal{O} = \{v_1, v_2\}$.



Overview of (Modified) Gram-Schmidt

- Let $\mathcal{B} = \{u_1, \ldots, u_n\}$ be a basis for a subspace $\mathcal{M}$.
- Create an *orthonormal basis* $\mathcal{O} = \{v_1, \ldots, v_n\}$ for $\mathcal{M}$.

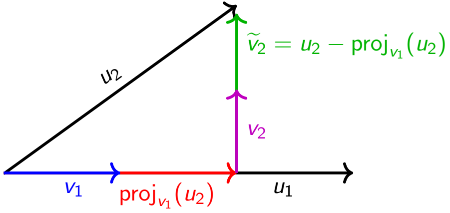
$n = 2$: $\quad \mathcal{B} = \{u_1, u_2\}$ and $\mathcal{O} = \{v_1, v_2\}$.



Overview of (Modified) Gram-Schmidt

- Let $\mathcal{B} = \{u_1, \ldots, u_n\}$ be a basis for a subspace $\mathcal{M}$.
- Create an *orthonormal basis* $\mathcal{O} = \{v_1, \ldots, v_n\}$ for $\mathcal{M}$.

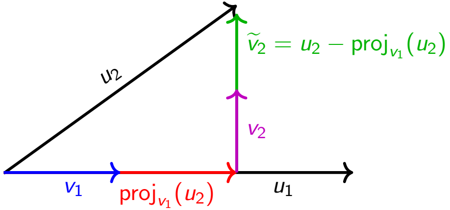
$n = 2$: $\quad \mathcal{B} = \{u_1, u_2\}$ and $\mathcal{O} = \{v_1, v_2\}$.



Overview of (Modified) Gram-Schmidt

- Let $\mathcal{B} = \{u_1, \ldots, u_n\}$ be a basis for a subspace $\mathcal{M}$.
- Create an *orthonormal basis* $\mathcal{O} = \{v_1, \ldots, v_n\}$ for $\mathcal{M}$.

$n = 2$: $\mathcal{B} = \{u_1, u_2\}$ and $\mathcal{O} = \{v_1, v_2\}$.



$$\text{proj}_v(u) = \frac{\langle u, v\rangle v}{\langle v, v\rangle} = \langle u, v\rangle v$$

$$v_1 = \frac{u_1}{\|u_1\|_2}$$

$$v_2 = \frac{u_2 - \text{proj}_{v_1}(u_2)}{\left\|u_2 - \text{proj}_{v_1}(u_2)\right\|_2} = \frac{u_2 - \langle u_2, v_1\rangle v_1}{\| \quad \ldots \quad \|_2}$$

Overview of (Modified) Gram-Schmidt

$$\begin{aligned} v_1 &= \frac{u_1}{\gamma_1} \\ v_2 &= \frac{u_2 - \langle u_2, v_1 \rangle v_1}{\gamma_2} \\ v_3 &= \frac{u_3 - \langle u_3, v_1 \rangle v_1 - \langle u_3, v_2 \rangle v_2}{\gamma_3} \\ &\vdots \\ v_n &= \frac{u_n - \langle u_n, v_1 \rangle v_1 - \cdots - \langle u_n, v_{n-1} \rangle v_{n-1}}{\gamma_k} \end{aligned}$$

where $\gamma_k = \left\| u_k - \sum_{i=1}^{k-1} \langle u_k, v_i \rangle v_i \right\|_2$ for $k = 1, \ldots, n$

Moving Onwards . . .

$$\begin{aligned}
u_1 &= \gamma_1 v_1 \\
u_2 &= \langle u_2, v_1 \rangle v_1 + \gamma_2 v_2 \\
u_3 &= \langle u_3, v_1 \rangle v_1 + \langle u_3, v_2 \rangle v_2 + \gamma_3 v_3 \\
&\vdots \\
u_n &= \langle u_n, v_1 \rangle v_1 + \cdots + \gamma_n v_n
\end{aligned}$$

Moving Onwards ...

$$
\begin{aligned}
u_1 &= \gamma_1 v_1 \\
u_2 &= \langle u_2, v_1 \rangle v_1 + \gamma_2 v_2 \\
u_3 &= \langle u_3, v_1 \rangle v_1 + \langle u_3, v_2 \rangle v_2 + \gamma_3 v_3 \\
&\vdots \\
u_n &= \langle u_n, v_1 \rangle v_1 + \cdots + \gamma_n v_n
\end{aligned}
$$

Let $A = \begin{bmatrix} u_1 & u_2 & \dots & u_n \end{bmatrix}_{m \times n}$ and $\widehat{Q} = \begin{bmatrix} v_1 & v_2 & \dots & v_n \end{bmatrix}_{m \times n}$

Moving Onwards ...

$$
\begin{aligned}
u_1 &= \gamma_1 v_1 \\
u_2 &= \langle u_2, v_1 \rangle v_1 + \gamma_2 v_2 \\
u_3 &= \langle u_3, v_1 \rangle v_1 + \langle u_3, v_2 \rangle v_2 + \gamma_3 v_3 \\
&\vdots \\
u_n &= \langle u_n, v_1 \rangle v_1 + \cdots + \gamma_n v_n
\end{aligned}
$$

Let $A = \begin{bmatrix} u_1 & u_2 & \dots & u_n \end{bmatrix}_{m \times n}$ and $\widehat{Q} = \begin{bmatrix} v_1 & v_2 & \dots & v_n \end{bmatrix}_{m \times n}$

$$
A = \widehat{Q}_{m \times n} \cdot \widehat{R}_{n \times n} = \widehat{Q} \cdot \begin{bmatrix} \gamma_1 & \langle u_2, v_1 \rangle & \langle u_3, v_1 \rangle & \dots & \langle u_n, v_1 \rangle \\ & \gamma_2 & \langle u_3, v_2 \rangle & \dots & \langle u_n, v_2 \rangle \\ & & \gamma_3 & \dots & \langle u_n, v_3 \rangle \\ & & & \ddots & \vdots \\ & & & & \gamma_n \end{bmatrix}.
$$

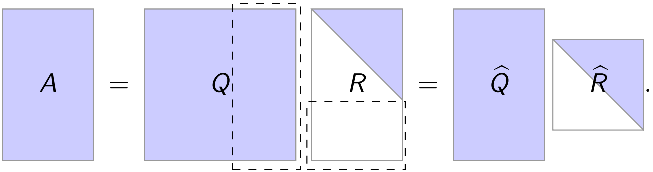
QR Factorization

- $A \in \mathbb{R}^{m\times n}$, $m \geq n$, with linearly independent columns has a unique reduced QR factorization $A = \widehat{Q}_{m\times n} \cdot \widehat{R}_{n\times n}$.

- … and also has a full QR factorization

$$A = Q_{m\times m} \cdot R_{m\times n} = \left[\ \widehat{Q}_{m\times n} \ \middle|\ \widetilde{Q}_{m\times(m-n)}\ \right] \left[\frac{\widehat{R}_{n\times n}}{0_{(m-n)\times n}}\right]$$

where Q is orthogonal and R has positive diagonal entries.



To Recap

- We need to obtain an ONB $\{v_1, \ldots, v_m\}$ for $\mathcal{K}_m$.
- It is evident we should use Gram-Schmidt ...
- ... and store the inner products carefully
- This procedure is called the Arnoldi Algorithm

Arnoldi

$\|v_1\|_2 = 1$
for $j = 1, \ldots, m$ **do**
    $h_{ij} = \langle Av_j, v_i \rangle$, $1 \leq i \leq j$
    $w_j = Av_j - \sum_{i=1}^{j} h_{ij} v_i$
    $h_{j+1,j} = \|w_j\|_2$
    If $h_{j+1,j} = 0$, stop
    $v_{j+1} = w_j / h_{j+1,j}$
end for

Arnoldi

$\|v_1\|_2 = 1$
for $j = 1, \ldots, m$ **do**
  $h_{ij} = \langle Av_j, v_i \rangle$, $1 \leq i \leq j$
  $w_j = Av_j - \sum_{i=1}^{j} h_{ij} v_i$
  $h_{j+1,j} = \|w_j\|_2$
  If $h_{j+1,j} = 0$, stop
  $v_{j+1} = w_j / h_{j+1,j}$
end for

$$
\begin{aligned}
Av_1 &= h_{11} v_1 + h_{21} v_2 \\
Av_2 &= h_{12} v_1 + h_{22} v_2 + h_{32} v_3 \\
Av_3 &= h_{13} v_1 + h_{23} v_2 + h_{33} v_3 + h_{43} v_4 \\
&\vdots \\
Av_m &= h_{1m} v_1 + h_{2m} v_2 + \cdots + h_{m+1,m} v_{m+1}
\end{aligned}
$$

Arnoldi $\quad (Av_m = h_{1m}\, v_1 + h_{2m}\, v_2 + \cdots + h_{m+1,m}\, v_{m+1})$

$$A \begin{bmatrix} | & | & & | \\ v_1 & v_2 & \dots & v_m \\ | & | & & | \end{bmatrix} = \begin{bmatrix} | & | & & | & | \\ v_1 & v_2 & \dots & v_m & v_{m+1} \\ | & | & & | & | \end{bmatrix} \begin{bmatrix} h_{11} & h_{12} & h_{13} & \dots & h_{1m} \\ h_{21} & h_{22} & h_{23} & \dots & h_{2m} \\ & h_{32} & h_{33} & \dots & h_{3m} \\ & & \ddots & \ddots & \vdots \\ & & & & h_{m,m} \\ & & & & h_{m+1,m} \end{bmatrix}$$

$$AV_m = V_{m+1}\widetilde{H}_m$$

- V_m is an ONB for $\mathcal{K}_m$
- $\widetilde{H}_m$ is "upper Hessenberg" — upper triangular + subdiagonal

In Summary

Recall: We are finding an approximate solution $\widetilde{x}$ to $Ax = b$

Idea: Look for $\widetilde{x} \in x_0 + \mathcal{K}_m(A, r_0)$ where $b - A\widetilde{x} \perp A\mathcal{K}_m$

Requirement: $\|b - A\widetilde{x}\| < \varepsilon$ for given tolerance ε

Next Step: Use the ONB from Arnoldi to represent $\widetilde{x}$

Find Approximate Solution

$$x \in x_0 + \mathcal{K}_m(A, r_0) \quad \Rightarrow \quad x = x_0 + V_m y, \qquad y \in \mathbb{R}^m.$$

Recall that $AV_m = V_{m+1}\widetilde{H}_m$.

$$b - Ax \quad = \quad b - A(x_0 + V_m y)$$

Find Approximate Solution

$$x \in x_0 + \mathcal{K}_m(A, r_0) \quad \Rightarrow \quad x = x_0 + V_m y, \qquad y \in \mathbb{R}^m.$$

Recall that $AV_m = V_{m+1}\widetilde{H}_m$.

$$\begin{aligned} b - Ax &= b - A(x_0 + V_m y) \\ &= r_0 - AV_m y \end{aligned}$$

Find Approximate Solution

$$x \in x_0 + \mathcal{K}_m(A, r_0) \quad \Rightarrow \quad x = x_0 + V_m y, \qquad y \in \mathbb{R}^m.$$

Recall that $AV_m = V_{m+1}\widetilde{H}_m$.

$$\begin{aligned} b - Ax &= b - A(x_0 + V_m y) \\ &= r_0 - AV_m y \\ &= \beta v_1 - V_{m+1}\widetilde{H}_m y \qquad\qquad \beta := \|r_0\|_2 \end{aligned}$$

Find Approximate Solution

$$x \in x_0 + \mathcal{K}_m(A, r_0) \quad \Rightarrow \quad x = x_0 + V_m y, \qquad y \in \mathbb{R}^m.$$

Recall that $AV_m = V_{m+1}\widetilde{H}_m$.

$$\begin{aligned} b - Ax &= b - A(x_0 + V_m y) \\ &= r_0 - AV_m y \\ &= \beta v_1 - V_{m+1}\widetilde{H}_m y & \beta := \|r_0\|_2 \\ &= V_{m+1}(\beta e_1 - \widetilde{H}_m y) \end{aligned}$$

Find Approximate Solution

$$x \in x_0 + \mathcal{K}_m(A, r_0) \quad \Rightarrow \quad x = x_0 + V_m y, \qquad y \in \mathbb{R}^m.$$

Recall that $AV_m = V_{m+1}\widetilde{H}_m$.

$$\begin{aligned}
b - Ax &= b - A(x_0 + V_m y) \\
&= r_0 - AV_m y \\
&= \beta v_1 - V_{m+1}\widetilde{H}_m y \qquad\qquad \beta := \|r_0\|_2 \\
&= V_{m+1}(\beta e_1 - \widetilde{H}_m y) \\
\|b - Ax\|_2 &= \left\|\beta e_1 - \widetilde{H}_m y\right\|_2
\end{aligned}$$

Find Approximate Solution

$$x \in x_0 + \mathcal{K}_m(A, r_0) \quad \Rightarrow \quad x = x_0 + V_m y, \qquad y \in \mathbb{R}^m.$$

Recall that $AV_m = V_{m+1}\widetilde{H}_m$.

$$\begin{aligned} b - Ax &= b - A(x_0 + V_m y) \\ &= r_0 - AV_m y \\ &= \beta v_1 - V_{m+1}\widetilde{H}_m y \qquad\qquad \beta := \|r_0\|_2 \\ &= V_{m+1}(\beta e_1 - \widetilde{H}_m y) \\ \|b - Ax\|_2 &= \left\|\beta e_1 - \widetilde{H}_m y\right\|_2 \end{aligned}$$

Therefore,

$$\min_{x \in x_0 + \mathcal{K}_m} \|b - Ax\|_2 = \min_{y \in \mathbb{R}^m} \left\|\beta e_1 - \widetilde{H}_m y\right\|_2.$$

This is a small dimensional problem!

Minimizing the Residual

Recall: We require $r_m = b - Ax_m$ to be orthogonal to $A\mathcal{K}_m$.

Minimizing the Residual

Recall: We require $r_m = b - Ax_m$ to be orthogonal to $A\mathcal{K}_m$.

Theorem 3 (Minimal Residual)

Let $x_m \in x_0 + \mathcal{K}_m$ be an approximate solution to $Ax = b$ with residual $r_m = b - Ax_m$. Then $\|r_m\|$ is minimized over $x_0 + \mathcal{K}_m$ if and only if $r_m \perp A\mathcal{K}_m(A, r_0)$.

Computing y_m

The classical way to minimize $\left\|\beta e_1 - \widetilde{H}_m y\right\|_2$ is using full QR factorization.

Computing y_m

The classical way to minimize $\left\| \beta e_1 - \widetilde{H}_m y \right\|_2$ is using full QR factorization.

Let $\widetilde{H}_m = Q_{(m+1)\times(m+1)} R_{(m+1)\times m}$ be the full QR factorization of $\widetilde{H}_m$.

Computing y_m

The classical way to minimize $\left\|\beta e_1 - \widetilde{H}_m y\right\|_2$ is using full QR factorization.

Let $\widetilde{H}_m = Q_{(m+1)\times(m+1)} R_{(m+1)\times m}$ be the full QR factorization of $\widetilde{H}_m$.

$$\begin{aligned}
\left\|\beta e_1 - \widetilde{H}_m y\right\|_2^2 &= \|\beta e_1 - QRy\|_2^2 \\
&= \left\|QQ^T \beta e_1 - QRy\right\|_2^2 && \text{since } QQ^T = I \\
&= \left\|Q(Q^T \beta e_1 - Ry)\right\|_2^2 \\
&= \left\|\beta Q^T e_1 - Ry\right\|_2^2 && Q \text{ is orthogonal}
\end{aligned}$$

Computing y_m

The classical way to minimize $\left\|\beta e_1 - \widetilde{H}_m y\right\|_2$ is using full QR factorization.

Let $\widetilde{H}_m = Q_{(m+1)\times(m+1)} R_{(m+1)\times m}$ be the full QR factorization of $\widetilde{H}_m$.

$$
\begin{aligned}
\left\|\beta e_1 - \widetilde{H}_m y\right\|_2^2 &= \|\beta e_1 - QRy\|_2^2 \\
&= \left\|QQ^T \beta e_1 - QRy\right\|_2^2 && \text{since } QQ^T = I \\
&= \left\|Q(Q^T \beta e_1 - Ry)\right\|_2^2 \\
&= \left\|\beta Q^T e_1 - Ry\right\|_2^2 && Q \text{ is orthogonal} \\
&= \left\|\beta Q^T e_1 - \begin{bmatrix} \widehat{R}y \\ 0_{1\times 1} \end{bmatrix}\right\|_2^2 && \widehat{R} \in \mathbb{R}^m \\
&= \left\|\begin{bmatrix} z_{1:m} \\ z_{m+1} \end{bmatrix} - \begin{bmatrix} \widehat{R}y \\ 0 \end{bmatrix}\right\|_2^2
\end{aligned}
$$

Where $z = \beta Q^T e_1 = \begin{bmatrix} z_{1:m} \\ z_{m+1} \end{bmatrix} \in \mathbb{R}^{m+1}$

$$
\begin{aligned}
\left\|\beta e_1 - \widetilde{H}_m y\right\|_2^2 &= \left\|\begin{bmatrix} z_{1:m} - \widehat{R}y \\ z_{m+1} \end{bmatrix}\right\|_2^2 \\
&= \left\|z_{1:m} - \widehat{R}y\right\|_2^2 + \|z_{m+1}\|_2^2 \\
&= \left\|z_{1:m} - \widehat{R}y\right\|_2^2 + |z_{m+1}|^2
\end{aligned}
$$

$$
\begin{aligned}
\left\|\beta e_1 - \widetilde{H}_m y\right\|_2^2 &= \left\|\begin{bmatrix} z_{1:m} - \widehat{R}y \\ z_{m+1} \end{bmatrix}\right\|_2^2 \\
&= \left\|z_{1:m} - \widehat{R}y\right\|_2^2 + \|z_{m+1}\|_2^2 \\
&= \left\|z_{1:m} - \widehat{R}y\right\|_2^2 + |z_{m+1}|^2
\end{aligned}
$$

Solving $\widehat{R}y = z_{1:m}$ yields the following:

$$
\begin{aligned}
y_m &= \underset{y \in \mathbb{R}^m}{\operatorname{argmin}} \left\|\beta e_1 - \widetilde{H}_m y\right\|_2 = \widehat{R}^{-1} z_{1:m} \\
& \min \left\|\beta e_1 - \widetilde{H}_m y\right\|_2 = |z_{m+1}| = \left|e_{m+1}^T \beta Q^T e_1\right|
\end{aligned}
$$

Norm of the residual is $\|Ax_m - b\|_2 = |z_{m+1}|$ and approximate solution is

$$
x_m = x_0 + V_m y_m, \qquad y_m = \widehat{R}^{-1} z_{1:m}.
$$

Generalized Minimal RESiduals (GMRES)

Given A, b, and an initial guess x_0, choose size m of the Krylov subspace.

$[x_m, r_m, \|r_m\|] = \text{GMRES}(A, b, x_0, m, tol)$

1: Compute $r_0 = b - Ax_0$, $\beta = \|r_0\|_2$, $v_1 = r_0/\beta$
2: **for** $j = 1, \ldots, m$ **do**
3: $\quad w_j = Av_j$
4: $\quad$ **for** $i = 1, \ldots, j$ **do**
5: $\quad\quad h_{ij} = \langle w_j, v_i \rangle$
6: $\quad\quad w_j = w_j - h_{ij} v_i$
7: $\quad$ **end for**
8: $\quad h_{j+1,j} = \|w_j\|_2$. If $h_{j+1,j} = 0$ set $m = j$ and go to 11
9: $\quad v_{j+1} = w_j / h_{j+1,j}$
10: **end for**
11: Compute $y_m = \underset{y \in \mathbb{R}^m}{\operatorname{argmin}} \left\| \beta e_1 - \widetilde{H}_m y \right\|_2$ and $x_m = x_0 + V_m y_m$

Generalized Minimal RESiduals (GMRES)

Given A, b, and an initial guess x_0, choose size m of the Krylov subspace.

$[x_m, r_m, \|r_m\|] = \text{GMRES}(A, b, x_0, m, tol)$

1: Compute $r_0 = b - Ax_0$, $\beta = \|r_0\|_2$, $v_1 = r_0/\beta$
2: **for** $j = 1, \ldots, m$ **do**
3: $\quad w_j = Av_j$
4: $\quad$ **for** $i = 1, \ldots, j$ **do**
5: $\quad\quad h_{ij} = \langle w_j, v_i \rangle$
6: $\quad\quad w_j = w_j - h_{ij} v_i$
7: $\quad$ **end for**
8: $\quad h_{j+1,j} = \|w_j\|_2$. If $h_{j+1,j} = 0$ set $m = j$ and go to 11
9: $\quad v_{j+1} = w_j / h_{j+1,j}$
10: **end for**
11: Compute $y_m = \underset{y \in \mathbb{R}^m}{\text{argmin}} \left\| \beta e_1 - \widetilde{H}_m y \right\|_2$ and $x_m = x_0 + V_m y_m$

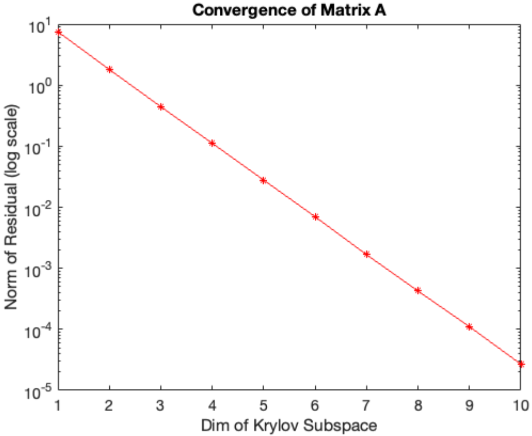
As m increases, $\|b - Ax_m\|_2$ is nonincreasing!

Example 1

$A = 2I + \frac{1}{2\sqrt{n}}\text{randn}(n), \qquad n = 1000, \qquad m = 10, \qquad \text{tol.} = 10^{-10}$

Example 1

$A = 2I + \frac{1}{2\sqrt{n}}\text{randn}(n), \qquad n = 1000, \qquad m = 10, \qquad \text{tol.} = 10^{-10}$



How do we get the residual at each subspace dimension?

Example 1

$A = 2I + \frac{1}{2\sqrt{n}}\text{randn}(n), \qquad m = 10, \qquad \text{tol.} = 10^{-10}$

n	$\|r_m\|_2$	Timing for x_m	Timing for LU
10^2	10^{-3}	8×10^{-3}	7×10^{-3}
10^3	10^{-5}	7×10^{-3}	3×10^{-2}
10^4	10^{-5}	2×10^{-1}	7×10^{1}

Givens Rotations

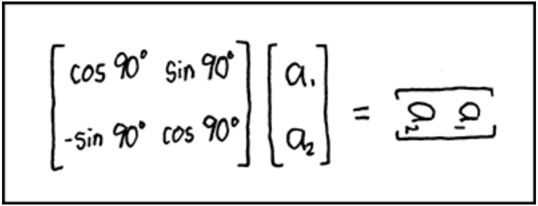
Method: Apply a sequence of orthogonal matrices, converting the matrix to an upper triangular form (creating full *QR* factorization).

Why? Any nonzero $x \in \mathbb{R}^n$ can be rotated to the i-th coordinate axis by a sequence of $n-1$ plane rotation matrices.

Givens Rotations

Method: Apply a sequence of orthogonal matrices, converting the matrix to an upper triangular form (creating full QR factorization).

Why? Any nonzero $x \in \mathbb{R}^n$ can be rotated to the i-th coordinate axis by a sequence of $n-1$ plane rotation matrices.



(XKCD Comic)

Givens Rotations

Method: Apply a sequence of orthogonal matrices, converting the matrix to an upper triangular form (creating full QR factorization).

Why? Any nonzero $x \in \mathbb{R}^n$ can be rotated to the i-th coordinate axis by a sequence of $n-1$ plane rotation matrices.

Given a vector $\begin{bmatrix} a \\ b \end{bmatrix}$, $a \neq 0$, choose $c, s \in \mathbb{R}$ such that $c^2 + s^2 = 1$ and

$$\begin{bmatrix} c & s \\ -s & c \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} \alpha \\ 0 \end{bmatrix}, \qquad \alpha = \sqrt{a^2 + b^2}.$$

Givens Rotations

Therefore $ca + sb = \alpha$ and $-sa + cb = 0$, so

$$\begin{bmatrix} a & b \\ b & -a \end{bmatrix} \begin{bmatrix} c \\ s \end{bmatrix} = \begin{bmatrix} \alpha \\ 0 \end{bmatrix}.$$

Computing row operations,

$$s = \frac{b}{\alpha}, \qquad c = \frac{a}{\alpha}. \tag{3}$$

To annihilate b in $\begin{bmatrix} a \\ b \end{bmatrix}$, choose s and c as in (3).

Givens Rotations

$$A = \begin{bmatrix} 1 & 3 & 1 & 6 \\ 3 & 9 & 3 & 2 \\ 0 & 3 & 1 & 0 \\ 2 & 1 & 5 & 2 \end{bmatrix} \qquad G_1 = \frac{1}{\sqrt{10}} \begin{bmatrix} 1 & 3 \\ -3 & 1 \end{bmatrix} \qquad \Omega_1 = \begin{bmatrix} G_1 & 0 \\ 0 & I_2 \end{bmatrix}$$

Givens Rotations

$$A = \begin{bmatrix} 1 & 3 & 1 & 6 \\ 3 & 9 & 3 & 2 \\ 0 & 3 & 1 & 0 \\ 2 & 1 & 5 & 2 \end{bmatrix} \qquad G_1 = \frac{1}{\sqrt{10}} \begin{bmatrix} 1 & 3 \\ -3 & 1 \end{bmatrix} \qquad \Omega_1 = \begin{bmatrix} G_1 & 0 \\ 0 & I_2 \end{bmatrix}$$

$$\Omega_1 A = \begin{bmatrix} \star & * & * & * \\ 0 & * & * & * \\ 0 & 3 & 1 & 0 \\ 2 & 1 & 5 & 2 \end{bmatrix}$$

Givens Rotations

$$A = \begin{bmatrix} 1 & 3 & 1 & 6 \\ 3 & 9 & 3 & 2 \\ 0 & 3 & 1 & 0 \\ 2 & 1 & 5 & 2 \end{bmatrix} \qquad G_1 = \frac{1}{\sqrt{10}} \begin{bmatrix} 1 & 3 \\ -3 & 1 \end{bmatrix} \qquad \Omega_1 = \begin{bmatrix} G_1 & 0 \\ 0 & I_2 \end{bmatrix}$$

$$\Omega_1 A = \begin{bmatrix} \star & * & * & * \\ 0 & * & * & * \\ 0 & 3 & 1 & 0 \\ 2 & 1 & 5 & 2 \end{bmatrix} \quad G_2 = \frac{1}{\sqrt{\star^2 + 2^2}} \begin{bmatrix} \star & 2 \\ -2 & \star \end{bmatrix} \quad \Omega_2 = \begin{bmatrix} \star & 0 & 0 & 2 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ -2 & 0 & 0 & \star \end{bmatrix}$$

Givens Rotations

$$A = \begin{bmatrix} 1 & 3 & 1 & 6 \\ 3 & 9 & 3 & 2 \\ 0 & 3 & 1 & 0 \\ 2 & 1 & 5 & 2 \end{bmatrix} \qquad G_1 = \frac{1}{\sqrt{10}} \begin{bmatrix} 1 & 3 \\ -3 & 1 \end{bmatrix} \qquad \Omega_1 = \begin{bmatrix} G_1 & 0 \\ 0 & I_2 \end{bmatrix}$$

$$\Omega_1 A = \begin{bmatrix} \star & * & * & * \\ 0 & * & * & * \\ 0 & 3 & 1 & 0 \\ 2 & 1 & 5 & 2 \end{bmatrix} \quad G_2 = \frac{1}{\sqrt{\star^2 + 2^2}} \begin{bmatrix} \star & 2 \\ -2 & \star \end{bmatrix} \quad \Omega_2 = \begin{bmatrix} \star & 0 & 0 & 2 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ -2 & 0 & 0 & \star \end{bmatrix}$$

$$\Omega_2 \Omega_1 A = \begin{bmatrix} \diamond & \diamond & \diamond & \diamond \\ 0 & * & * & * \\ 0 & 3 & 1 & 0 \\ 0 & \diamond & \diamond & \diamond \end{bmatrix}$$

Givens Rotations

$$A = \begin{bmatrix} 1 & 3 & 1 & 6 \\ 3 & 9 & 3 & 2 \\ 0 & 3 & 1 & 0 \\ 2 & 1 & 5 & 2 \end{bmatrix} \qquad G_1 = \frac{1}{\sqrt{10}} \begin{bmatrix} 1 & 3 \\ -3 & 1 \end{bmatrix} \qquad \Omega_1 = \begin{bmatrix} G_1 & 0 \\ 0 & I_2 \end{bmatrix}$$

$$\Omega_1 A = \begin{bmatrix} \star & * & * & * \\ 0 & * & * & * \\ 0 & 3 & 1 & 0 \\ 2 & 1 & 5 & 2 \end{bmatrix} \quad G_2 = \frac{1}{\sqrt{\star^2 + 2^2}} \begin{bmatrix} \star & 2 \\ -2 & \star \end{bmatrix} \quad \Omega_2 = \begin{bmatrix} \star & 0 & 0 & 2 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ -2 & 0 & 0 & \star \end{bmatrix}$$

$$\Omega_2 \Omega_1 A = \begin{bmatrix} \diamond & \diamond & \diamond & \diamond \\ 0 & * & * & * \\ 0 & 3 & 1 & 0 \\ 0 & \diamond & \diamond & \diamond \end{bmatrix}$$

$$\Omega_\ell \dots \Omega_3 \Omega_2 \Omega_1 A = R \quad \Rightarrow \quad A = \underbrace{\Omega_1^T \Omega_2^T \Omega_3^T \dots \Omega_\ell^T}_{Q} R$$

Givens Rotations Applied to GMRES

Instead of running Arnoldi all the way to m, maybe we can stop earlier with a satisfactory residual.

Givens Rotations Applied to GMRES

Instead of running Arnoldi all the way to m, maybe we can stop earlier with a satisfactory residual.

Define $\beta = \|r_0\|_2$, $v_1 = r_0/\beta$, and $\widetilde{g}_0 = \beta e_1 = \begin{bmatrix} \beta \\ 0 \end{bmatrix}$.

From Arnoldi, we obtain $V_1 = \begin{bmatrix} v_1 & v_2 \end{bmatrix}$ and $\widetilde{H}_1 = \begin{bmatrix} h_{11} \\ h_{21} \end{bmatrix}$.

Givens Rotations Applied to GMRES

Instead of running Arnoldi all the way to m, maybe we can stop earlier with a satisfactory residual.

Define $\beta = \|r_0\|_2$, $v_1 = r_0/\beta$, and $\widetilde{g}_0 = \beta e_1 = \begin{bmatrix} \beta \\ 0 \end{bmatrix}$.

From Arnoldi, we obtain $V_1 = \begin{bmatrix} v_1 & v_2 \end{bmatrix}$ and $\widetilde{H}_1 = \begin{bmatrix} h_{11} \\ h_{21} \end{bmatrix}$. Define

$\Omega_1 = \begin{bmatrix} c_1 & s_1 \\ -s_1 & c_1 \end{bmatrix}$ to annihilate h_{21}, creating

$$\widetilde{R}_1 = \Omega_1 \widetilde{H}_1 = \begin{bmatrix} h_{11}^{(1)} \\ 0 \end{bmatrix}, \qquad \widetilde{g}_1 = \Omega_1 \widetilde{g}_0 = \begin{bmatrix} \beta c_1 \\ -\beta s_1 \end{bmatrix} := \begin{bmatrix} \gamma_1 \\ \gamma_2 \end{bmatrix}.$$

Givens Rotations Applied to GMRES

Instead of running Arnoldi all the way to m, maybe we can stop earlier with a satisfactory residual.

Define $\beta = \|r_0\|_2$, $v_1 = r_0/\beta$, and $\widetilde{g}_0 = \beta e_1 = \begin{bmatrix} \beta \\ 0 \end{bmatrix}$.

From Arnoldi, we obtain $V_1 = \begin{bmatrix} v_1 & v_2 \end{bmatrix}$ and $\widetilde{H}_1 = \begin{bmatrix} h_{11} \\ h_{21} \end{bmatrix}$. Define

$\Omega_1 = \begin{bmatrix} c_1 & s_1 \\ -s_1 & c_1 \end{bmatrix}$ to annihilate h_{21}, creating

$$\widetilde{R}_1 = \Omega_1 \widetilde{H}_1 = \begin{bmatrix} h_{11}^{(1)} \\ 0 \end{bmatrix}, \qquad \widetilde{g}_1 = \Omega_1 \widetilde{g}_0 = \begin{bmatrix} \beta c_1 \\ -\beta s_1 \end{bmatrix} := \begin{bmatrix} \gamma_1 \\ \gamma_2 \end{bmatrix}.$$

The norm of the residual is $|\gamma_2| = |\beta s_1|$.

If small enough, set $R_1 = [h_{11}]$ and $g_1 = [\beta c_1]$ (remove last row), and

$$y_1 = R_1^{-1} g_1, \qquad x_1 = x_0 + V_1 y.$$

With one more Arnoldi step, $V_2 = \begin{bmatrix} v_1 & v_2 & v_3 \end{bmatrix}$ and $h = \begin{bmatrix} h_{21} \\ h_{22} \\ h_{32} \end{bmatrix}$ is the last column of $\widetilde{H}_2$.

With one more Arnoldi step, $V_2 = \begin{bmatrix} v_1 & v_2 & v_3 \end{bmatrix}$ and $h = \begin{bmatrix} h_{21} \\ h_{22} \\ h_{32} \end{bmatrix}$ is the last column of $\widetilde{H}_2$.

Apply previous rotations, Ω_1, to h and append to $\widetilde{R}_1$ with a row of zeros:

$$\widetilde{R}_2 = \left[\begin{array}{c|c} h_{11}^{(1)} & h_{21}^{(2)} \\ 0 & h_{22}^{(2)} \\ \hline 0 & h_{32} \end{array}\right], \qquad \widetilde{g}_2 = \begin{bmatrix} \gamma_1 \\ \gamma_2 \\ 0 \end{bmatrix}.$$

With one more Arnoldi step, $V_2 = \begin{bmatrix} v_1 & v_2 & v_3 \end{bmatrix}$ and $h = \begin{bmatrix} h_{21} \\ h_{22} \\ h_{32} \end{bmatrix}$ is the last column of $\widetilde{H}_2$.

Apply previous rotations, Ω_1, to h and append to $\widetilde{R}_1$ with a row of zeros:

$$\widetilde{R}_2 = \left[\begin{array}{c|c} h_{11}^{(1)} & h_{21}^{(2)} \\ 0 & h_{22}^{(2)} \\ \hline 0 & h_{32} \end{array}\right], \qquad \widetilde{g}_2 = \begin{bmatrix} \gamma_1 \\ \gamma_2 \\ 0 \end{bmatrix}.$$

Create Ω_2, now 3×3, to annihilate h_{32}. Apply also to $\widetilde{g}_2$:

$$\widetilde{R}_2 \leftarrow \begin{bmatrix} h_{11}^{(1)} & h_{21}^{(2)} \\ 0 & h_{22}^{(2)} \\ 0 & 0 \end{bmatrix}, \qquad \widetilde{g}_2 \leftarrow \Omega_2 \widetilde{g}_2 = \begin{bmatrix} \gamma_1 \\ c_2\gamma_2 \\ -s_2\gamma_2 \end{bmatrix}.$$

The norm of the residual is $|s_2\gamma_2|$.

If the residual $|s_2\gamma_2|$ is small enough,

$$R_2 = \begin{bmatrix} h_{11}^{(1)} & h_{21}^{(2)} \\ 0 & h_{22}^{(2)} \end{bmatrix}, \qquad g_2 = \begin{bmatrix} \gamma_1 \\ c_2\gamma_2 \end{bmatrix}$$

(remove last row). Then

$$y_2 = R_2^{-1} g_2, \qquad x_2 = x_0 + V_2 y_2.$$

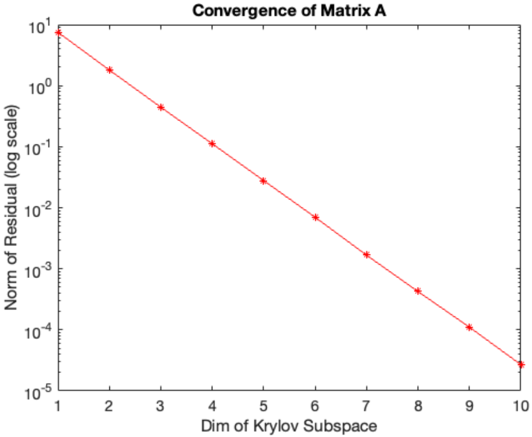
... and so on

Givens Rotations Applied to GMRES

1. Compute the next vector in V_k from Arnoldi, and h the last column $\widetilde{H}_k$.
2. Apply the previous rotations to h.
3. Let $\widetilde{R}_k = \left[\begin{array}{c|c} \widetilde{R}_{k-1} & | \\ \hline 0 & \begin{array}{c} h \\ | \end{array} \end{array}\right]$, $\quad \widetilde{g}_k = \left[\begin{array}{c} \widetilde{g}_{k-1} \\ \hline 0 \end{array}\right]$.
4. Apply Ω_k to $\widetilde{R}_k$ and to $\widetilde{g}_k$ so the $(k+1, k)$ entry in $\widetilde{R}_k$ is annihilated.
5. Test the residual $|\widetilde{g}_k(k)|$, i.e. the last entry
6. If satisfied, let $R_k = \widetilde{R}_k(1:k, 1:k)$ and $g_k = \widetilde{g}_k(1:k)$ be $k \times k$ and $k \times 1$ respectively. Else, go to 1.
7. Let $y_k = R_k^{-1} g_k$ and $x_k = x_0 + V_k y_k$.

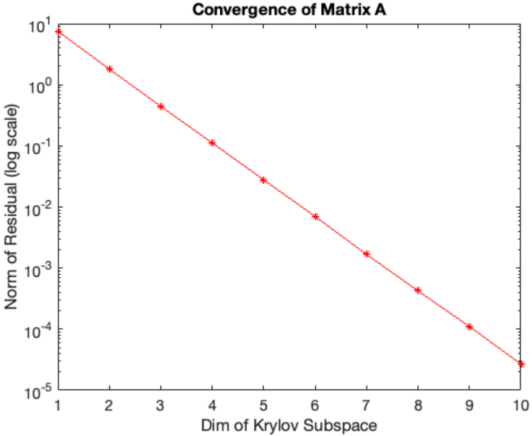
Example 1 Continuation

What happens if the residual is too large in the m-th subspace?



Example 1 Continuation

What happens if the residual is too large in the m-th subspace?



Answer: Restart the algorithm!

Restarted GMRES

$[x_m, r_m, \|r_m\|] = \text{GMRES}(A, b, x_0, m, \text{tol}, \text{num-restarts})$

repeat

  Compute $r_0 = b - Ax_0$, $\beta = \|r_0\|_2$, $v_1 = r_0/\beta$

  Run the Arnoldi Procedure starting with v_1

  $x_0 = x_m$

until $\|b - Ax_m\|_2 <$ tol or "num-restarts" is met

Compute $y_m = \underset{y \in \mathbb{R}^m}{\text{argmin}} \left\| \beta e_1 - \widetilde{H}_m y \right\|_2$ and $x_m = x_0 + V_m y_m$

Restarted GMRES

$[x_m, r_m, \|r_m\|] = \text{GMRES}(A, b, x_0, m, \text{tol}, \text{ num-restarts })$

1: **repeat**
2: Compute $r_0 = b - Ax_0$, $\beta = \|r_0\|_2$, $v_1 = r_0/\beta$
3: Run the Arnoldi Procedure starting with v_1
4: $x_0 = x_m$
5: **until** $\|b - Ax_m\|_2 < \text{tol}$ or "num-restarts" is met
6: Compute $y_m = \underset{y \in \mathbb{R}^m}{\operatorname{argmin}} \left\| \beta e_1 - \widetilde{H}_m y \right\|_2$ and $x_m = x_0 + V_m y_m$

Now we can control two parameters — the size of the Krylov subspace to build to, m, and the number of restarts.

Example 1 with RESTARTED GMRES

We consider the same matrix

$$A = 2I + \frac{1}{2\sqrt{n}}\text{randn}(n), \qquad n = 1000, \qquad m = 5, \qquad \text{tol.} = 10^{-10}$$

and run GMRES with up to 10 restarts.

Example 1 with RESTARTED GMRES

We consider the same matrix

$$A = 2I + \frac{1}{2\sqrt{n}}\text{randn}(n), \qquad n = 1000, \qquad m = 5, \qquad \text{tol.} = 10^{-10}$$

and run GMRES with up to 10 restarts.

Timing: 0.014687 seconds

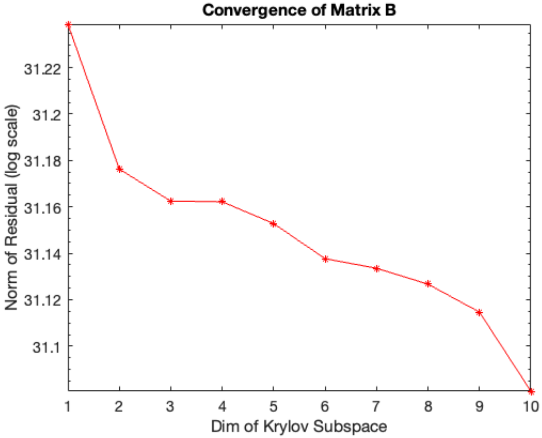
Recall:
No restarts was 0.018848

Example 2

$$B = \frac{1}{2\sqrt{n}}\text{randn}(n), \qquad n = 1000, \qquad m = 10, \qquad \text{tol.} = 10^{-10}.$$

Example 2

$$B = \frac{1}{2\sqrt{n}}\mathrm{randn}(n), \qquad n = 1000, \qquad m = 10, \qquad \text{tol.} = 10^{-10}.$$



Timing: 0.014687 seconds

Example 2 – Restarts

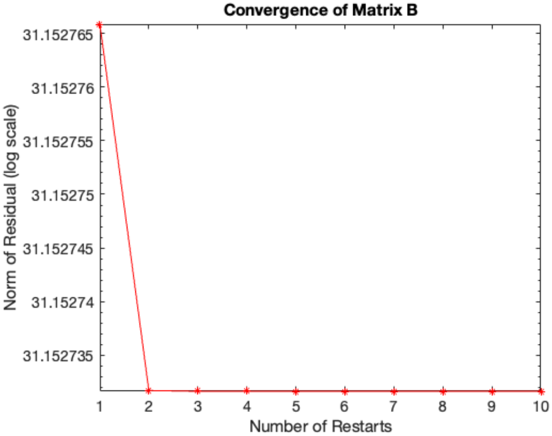
$$B = \frac{1}{2\sqrt{n}}\text{randn}(n), \qquad n = 1000, \qquad m = 5, \qquad \text{tol.} = 10^{-10}$$

(up to 10 restarts)

Example 2 – Restarts

$$B = \frac{1}{2\sqrt{n}}\text{randn}(n), \qquad n = 1000, \qquad m = 5, \qquad \text{tol.} = 10^{-10}$$

(up to 10 restarts)

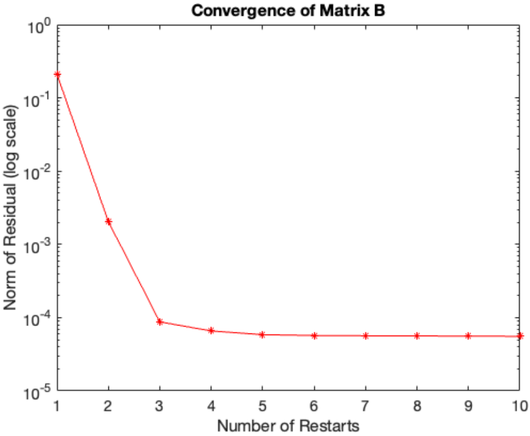


Timing: 0.012786 seconds

Recall:
No restarts was 0.014687

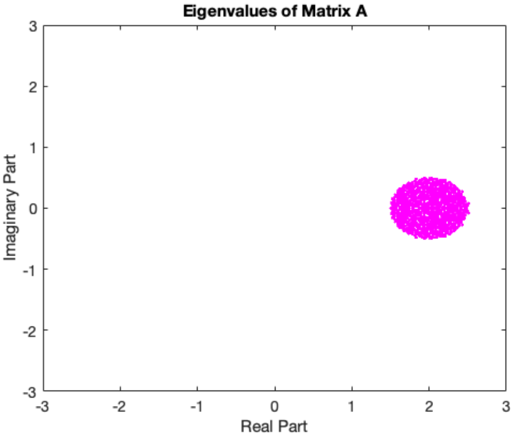
Example 2

We have to go all the way out to $m = 999$ for the following



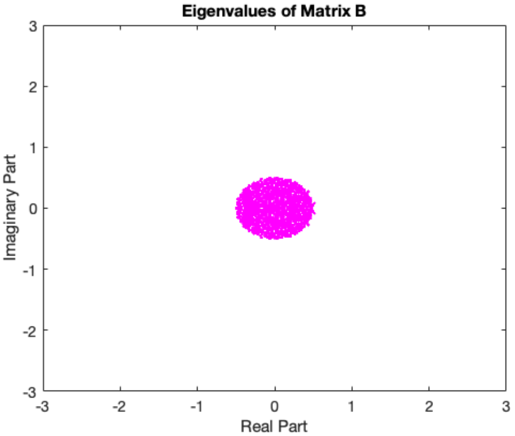
10.36 seconds!

Example 1



Eigenvalues of Matrix A
Imaginary Part
Real Part
3
2
1
0
-1
-2
-3
-3
-2
-1
0
1
2
3

Example 2


Eigenvalues of Matrix B
Imaginary Part
Real Part
3
2
1
0
-1
-2
-3
-3
-2
-1
0
1
2
3

Conclusion

- Introduced an iterative method for square linear system $Ax = b$.
- Developed machinery for efficiently working with this kind of method: Krylov Subspaces, Arnoldi Procedure, Givens Rotations.
- Discussed implementation challenges for GMRES.
- Up next:
  - Convergence: Chebyshev polynomials, min-max theorems, ...
  - Augmented GMRES: Append basis for $\mathcal{K}_m$ with other vectors ??
  - ...

Questions?

Theorem 4 (Closest Vector)

Let P be the orthogonal projector onto $\mathcal{K}$ along $\mathcal{K}^\perp$. Then for all $x \in \mathbb{R}^n$,

$$\min_{y\in\mathcal{K}} \|x - y\|_2 = \|x - Px\|_2.$$

Proof.

Since $x - Px \perp \mathcal{K}$ and $Px - y \in \mathcal{K}$, then

$$\begin{aligned} \|x - y\|_2^2 &= \|x - Px + (Px - y)\|_2^2 = \|x - Px\|_2^2 + \|Px - y\|_2^2 \\ &\geq \|x - Px\|_2^2 \end{aligned}$$

with equality when $y = Px$. □

Corollary 5

For $x \in \mathbb{R}^n$,

$$\min_{y\in\mathcal{K}} \|x - y\|_2 = \|x - y^*\|_2 \quad \Leftrightarrow \quad y^* \in \mathcal{K}, \quad x - y^* \perp \mathcal{K}.$$

Theorem 6 (Minimal Residual)

Let $x_m \in x_0 + \mathcal{K}_m$ be an approximate solution to $Ax = b$ with residual $r_m = b - Ax_m$. Then $\|r_m\|$ is minimized over $x_0 + \mathcal{K}_m$ if and only if $r_m \perp A\mathcal{K}_m(A, r_0)$.

$\boxed{\Rightarrow}$ Suppose the minimum is achieved. Then

$$
\begin{aligned}
r_m = b - Ax_m = \min_{x \in x_0 + \mathcal{K}_m} \|b - Ax\| &= \min_{y \in \mathcal{K}_m} \|b - A(x_0 + y)\| \\
&= \min_{y \in \mathcal{K}_m} \|r_0 - Ay\| \\
&= \min_{w \in A\mathcal{K}_m} \|r_0 - w\|
\end{aligned}
$$

This is achieved for $w = Pr_0$, where P is the orthogonal projector onto $A\mathcal{K}_m$. This means that $(I - P)r_0 \perp A\mathcal{K}_m$. But $(I - P)r_0 = r_0 - Pr_0 = r_0 - w$ and because w minimizes the above norms, it means that

$$
\min_{x \in x_0 + \mathcal{K}_m} \|b - Ax\| = \|r_0 - w\|.
$$

Then $b - Ax_m = r_0 - w$ and so $r_m = (I - P)r_0 \perp A\mathcal{K}_m$.

$\boxed{\Leftarrow}$ If $r_m \perp A\mathcal{K}_m$, then $b - Ax_m \perp A\mathcal{K}_m$. Since $Ax_m \in A\mathcal{K}_m$, so by Corollary 5 then

$$\min_{Ax \in A\mathcal{K}_m} \|b - Ax\| = \|b - Ax_m\| = \min_{x \in \mathcal{K}_m} \|b - Ax\|.$$