CHAPTER 8

Skip Sections 8.1, 8.4, 8.6, some of 8.7, 8.8

Eigenvalues and Eigenvectors

Recall the vaccination problem from CP3:

$$\vec{X}_{k+1} = M_p \vec{X}_k$$

$$M_p = \begin{pmatrix} 1-\frac{1}{200}-p & 0 & \frac{1}{10000} \\ \frac{1}{200} & 1-\frac{1}{1000} & 0 \\ p & \frac{1}{1000} & 1-\frac{1}{10000} \end{pmatrix}$$

Suppose

$$M_p = U \Lambda U^{-1} \qquad \Lambda = \begin{pmatrix} \lambda_1 & & \\ & \lambda_2 & \\ & & \lambda_3 \end{pmatrix}.$$

Then

$$\vec{X}_k = M^k \vec{X}_0 = (U\Lambda U^{-1})(U\Lambda U^{-1}) \cdots (U\Lambda U^{-1})\vec{X}_0$$
$$= U \Lambda^k U^{-1} \vec{X}_0$$
$$= U \begin{pmatrix} \lambda_1^k & & \\ & \lambda_2^k & \\ & & \lambda_3^k \end{pmatrix} U^{-1} \vec{X}_0$$

$p=0$

$$\lambda_1 = 1, \quad \lambda_2 = .9989\cdots, \quad \lambda_3 = .995\ldots$$

$$\lim_{k\to\infty} \vec{X}_k = U \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} U^{-1} \vec{X}_0 \quad !$$

Recall that the iteration

$$\vec{x}_{k+1} = M\vec{x}_k + \vec{b} \quad , \quad M = I - A.$$

converges if $M^k \to 0$ as $k \to \infty$.

Suppose

$$M = U \Lambda U^{-1} \quad \Lambda = \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix}$$

where $|\lambda_i| < 1$ for $i = 1, 2, \ldots, n$.

Then $M^k = U \Lambda^k U^{-1} \to 0$!

The values λ_j are called eigenvalues and the columns of U are called eigenvectors.

<u>Definition</u> Let A be an $n \times n$ matrix. A scalar λ is called an eigenvalue of A if there is a non-zero vector $\vec{v}$, called an eigenvector such that

$$A\vec{v} = \lambda \vec{v}.$$

<u>Note:</u> This equation can be rewritten as

$$A\vec{v} = \lambda I \vec{v}$$

OR

$$(A-\lambda I)\vec{v}=\vec{0} \iff A-\lambda I \text{ is singular}$$

<u>Theorem</u> A scalar λ is an eigenvalue of an $n\times n$ matrix A if and only if $(A-\lambda I)$ is singular. The corresponding eigenvectors are solutions to the eigenvalue equation $(A-\lambda I)\vec{v}=\vec{0}$.

<u>Corollary</u> A scalar λ is an eigenvalue of the matrix A if and only if λ is a solution of the characteristic equation

$$\det(A-\lambda I)=0.$$

<u>Ex</u> Find the eigenvalues and eigenvectors of

$$A=\begin{pmatrix}3 & 1\\ 1 & 3\end{pmatrix}.$$

$$\det(A-\lambda I)=\det\begin{pmatrix}3-\lambda & 1\\ 1 & 3-\lambda\end{pmatrix} = (3-\lambda)^2-1$$

<u>Recall</u> $\det\begin{pmatrix}a & b\\ c & d\end{pmatrix} = ad-bc$

$$\det(A-\lambda I)=9+\lambda^2-6\lambda-1=\lambda^2-6\lambda+8$$

$$\lambda=\frac{6\pm\sqrt{36-32}}{2}=3\pm1=2,4$$

First eigenvector:

$$\left(\begin{array}{cc|c} 1 & 1 & 0 \\ 1 & 1 & 0 \end{array}\right) \longrightarrow \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix}$$

$$\uparrow$$

$$(A - I)\vec{v} = 0 \qquad \vec{v}_1 = \begin{pmatrix} -1 \\ 1 \end{pmatrix}$$

Second eigenvector:

$$\left(\begin{array}{cc|c} -1 & 1 & 0 \\ 1 & -1 & 0 \end{array}\right), \qquad \vec{v}_2 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

So, $A = \begin{pmatrix} 3 & 1 \\ 1 & 3 \end{pmatrix}$ has eigenvalue $\lambda = 2$ with eigenvector $\begin{pmatrix} -1 \\ 1 \end{pmatrix}$

$A = \begin{pmatrix} 3 & 1 \\ 1 & 3 \end{pmatrix}$ has eigenvalue $\lambda = 4$ with eigenvector $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$

3x3 determinants.

$$\det \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} = a_{11} \det \begin{pmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{pmatrix}$$

$$- a_{12} \det \begin{pmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{pmatrix} + a_{13} \det \begin{pmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{pmatrix}$$

This is using the top row. You can use any row or column if you keep the following sign table in mind

$$\begin{pmatrix} + & - & + \\ - & + & - \\ + & - & + \end{pmatrix}$$

<u>Second row</u>:

$$\det A = -a_{21} \det \begin{pmatrix} a_{12} & a_{13} \\ a_{32} & a_{33} \end{pmatrix} + a_{22} \det \begin{pmatrix} a_{11} & a_{13} \\ a_{31} & a_{33} \end{pmatrix}$$

$$- a_{23} \det \begin{pmatrix} a_{11} & a_{12} \\ a_{31} & a_{32} \end{pmatrix}$$

<u>Ex</u> Find the eigenvalues of

$$A = \begin{pmatrix} 0 & -1 & -1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{pmatrix}$$

$$\det(A - \lambda I) = \det \begin{pmatrix} -\lambda & -1 & -1 \\ 1 & 2-\lambda & 1 \\ 1 & 1 & 2-\lambda \end{pmatrix}$$

$$= (-\lambda)\left[(2-\lambda)^2 - 1\right] - (-1)\left[2 - \lambda - 1\right] + (-1)\left[1 - (2-\lambda)\right]$$

$$= (-\lambda)(2-\lambda)^2 + \cancel{\lambda} + 1 - \cancel{\lambda} + 1 - \lambda$$

$$= -\lambda^3 + 4\lambda^2 - 4\lambda - \lambda + 2 = -\lambda^3 + 4\lambda^2 - 5\lambda + 2$$

$$= -(\lambda - 1)^2(\lambda - 2) = 0 \qquad \lambda = 1, 2\ !$$

Other facts about eigenvalues (A is $n \times n$)

- A is invertible if and only if $\lambda = 0$ is not an eigenvalue

- A has at least one and at most n distinct eigenvalues.

- The eigenvalues of a triangular matrix are the entries on the diagonal.

Ex Find the eigenvalues of

$$A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}$$

A complex number is

$$z = x + iy, \quad x, y \in \mathbb{R}, \quad i = \sqrt{-1}$$

$$i^2 = -1$$

$$(x+iy)(\alpha + i\beta) = x\alpha + iy\alpha + ix\beta - y\beta$$
$$= (x\alpha - y\beta) + i(y\alpha + x\beta).$$

The complex plane $\mathbb{C}$ is the set of all numbers

$$\mathbb{C} = \{x + iy : x, y \in \mathbb{R}\}$$

$\mathbb{C}$ can be visualized as $\mathbb{R}^2$. Some definitions:

- For $z = x + iy$, $\bar{z} = x - iy$.
- For $z \in \mathbb{C}$. $|z| = \sqrt{\bar{z} z}$

$$|x + iy| = \sqrt{(x+iy)(x-iy)} = \sqrt{x^2 + y^2}$$

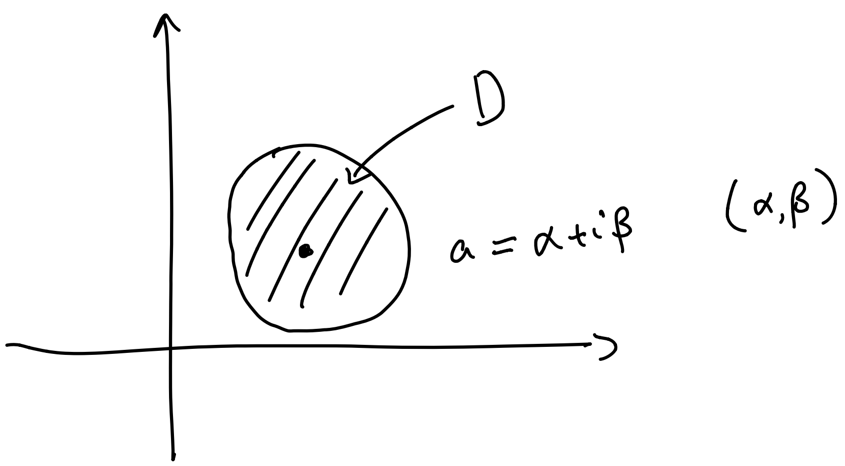
The Gershgorin Circle Theorem

Before we state the theorem, consider the set ($a \in \mathbb{C}$, $r \in \mathbb{R}$)

$$D = \{ z \in \mathbb{C} : |z - a| \leq r \}.$$

What is it? Set $z = x + iy$ $a = \alpha + i\beta$

$$|z-a|^2 = |((x-\alpha) + i(y-\beta))|^2$$
$$= (x-\alpha)^2 + (y-\beta)^2 \leq r^2$$



D is a disk centered at a with radius r.

<u>Definition</u> Let A be an nxn matrix, either real or complex. For each i, $1 \leq i \leq n$, define the Gershgorin disks

$$D_i = \{ z \in \mathbb{C} : |z - a_{ii}| \leq r_i \}$$

where $$r_i = \sum_{\substack{j=1 \\ j \neq i}}^{n} |a_{ij}|$$

Then define the Gershgorin domain

$$D_A = \bigcup_{i=1}^{n} D_i \subset \mathbb{C}$$

<u>Theorem</u> All the eigenvalues of A lie in its Gershgorin domain. Each Gershgorin disk contains at least one eigenvalue.

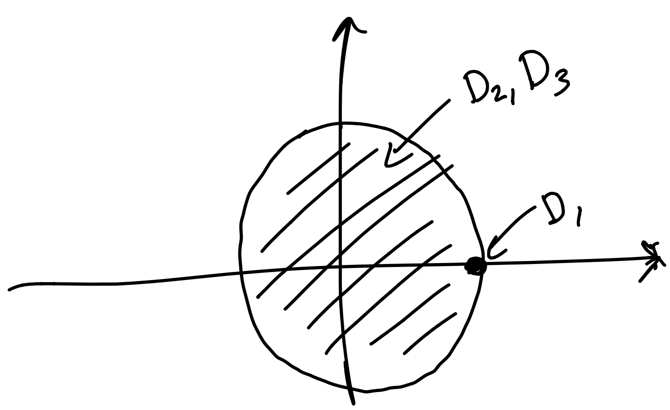
<u>Ex</u>

$$A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}$$

$$D_1 = \{ z \in \mathbb{C} : |z-1| \leq 0 \}$$

$$D_2 = \{ z \in \mathbb{C} : |z-0| \leq 1 \}$$

$$D_3 = \{ z \in \mathbb{C} : |z-0| \leq 1 \}$$



That doesn't tell us a whole lot.

Ex

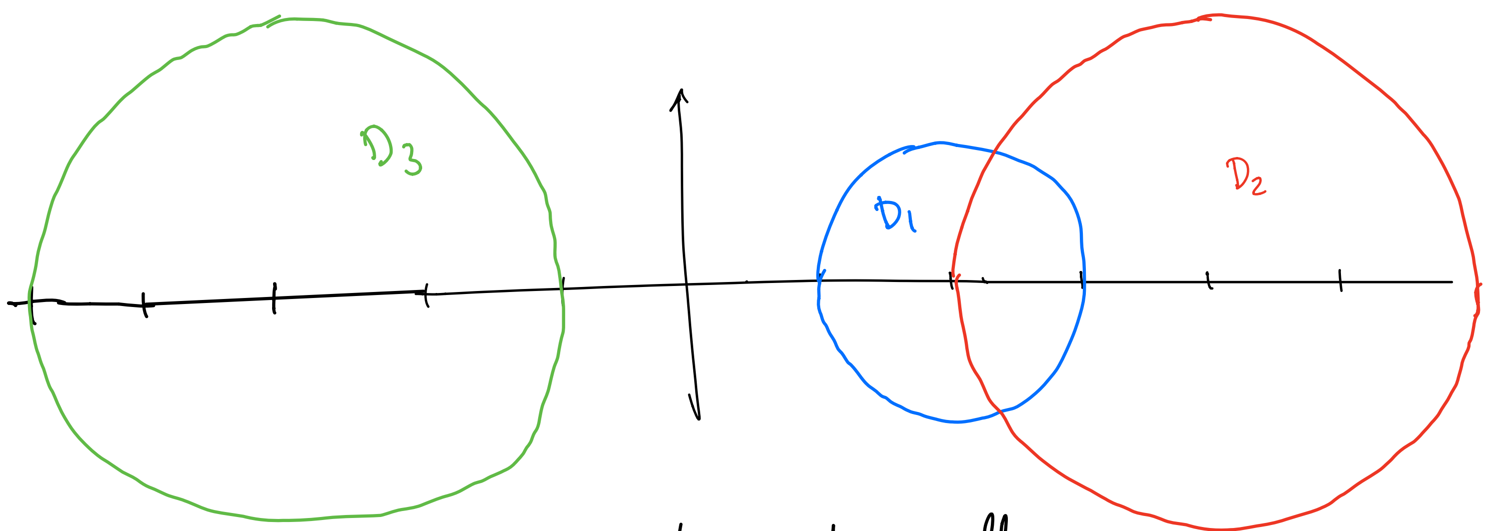
$$A = \begin{pmatrix} 2 & -1 & 0 \\ 1 & 4 & -1 \\ -1 & -1 & -3 \end{pmatrix}$$

$$D_1 = \{ z \in \mathbb{C} : |z-2| \leq 1 \}$$

$$D_2 = \{ z \in \mathbb{C} : |z-4| \leq 2 \}$$

$$D_3 = \{ z \in \mathbb{C} : |z+3| \leq 2 \}.$$

This tells us more



What does this tell us about diagonally dominant matrices?

Recall A, $n \times n$, is diagonally dominant if

$$|a_{ii}| > \sum_{\substack{j=1 \\ j \neq i}}^{n} |a_{ij}|$$

The absolute value of the center is larger than the radius of all the Gershgorin disks.

The distance from the center to the origin is larger than the radius.

The origin $z=0$ is not in any Gershgorin disk.

A is invertible.

<u>More on symmetric matrices</u>

Suppose $A = A^T$ where A is an $n\times n$ real matrix.

i) If λ is an eigenvalue of A, then λ is real

ii) If $\lambda_1 \neq \lambda_2$ are two eigenvalues of A with eigenvectors $\vec{v}_1, \vec{v}_2$. Then $\vec{v}_1 \cdot \vec{v}_2 = 0$.

i) Suppose $A\vec{v} = \lambda\vec{v}$. Then

conjugate of a vector.

$$\overline{A\vec{v}} = \overline{\lambda\vec{v}} = \bar{A}\,\underline{\vec{v}} = \bar{\lambda}\,\underline{\vec{v}}$$

$$= A\,\underline{\vec{v}} = \bar{\lambda}\,\underline{\vec{v}}$$

$\Rightarrow \bar{\lambda}$ is an eigenvalue for A with eigenvalue λ.

$$\vec{v}^T A \underline{\vec{v}} = \vec{v}^T \bar{\lambda}\,\underline{\vec{v}} = \bar{\lambda}\,\vec{v}^T \underline{\vec{v}}$$

$$\parallel$$

$$\vec{v}^T A^T \underline{\vec{v}} = (A\vec{v})^T \underline{\vec{v}} = \lambda\,\vec{v}^T \underline{\vec{v}}$$

since $\vec{v}^T \overline{\underline{v}} \neq 0$, we conclude that

$$\lambda = \bar{\lambda} \iff \lambda \in \mathbb{R}.$$

ii) $A\vec{v}_1 = \lambda_1 \vec{v}_1$, $A\vec{v}_2 = \lambda_2 \vec{v}_2$

$$\lambda_1(\vec{v}_1^{\,T}\vec{v}_2) = (A\vec{v}_1)^T\vec{v}_2 = \vec{v}_1^{\,T} A^T \vec{v}_2$$

$$= \vec{v}_1^{\,T} A \vec{v}_2 = \lambda_2(\vec{v}_1^{\,T}\vec{v}_2)$$

$$\Rightarrow (\lambda_1 - \lambda_2)(\vec{v}_1^{\,T}\vec{v}_2) = 0$$

Since $\lambda_1 \neq \lambda_2$, we see that $\vec{v}_1^{\,T}\vec{v}_2 = 0$!

Spectral Factorization

Definition A matrix B is similar to A if there exists an invertible matrix V such that

$$AV = VB.$$

Theorem If A and B are similar, then they have the same eigenvalues.

Definition An $n \times n$ matrix A is diagonalizable if it is similar to a diagonal matrix:

$$AV = VD, \quad D = \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix}$$

Definition A spectral factorization of an $n\times n$ matrix A is given by

$$A = V D V^{-1}$$

where D is diagonal.

The columns of V are eigenvectors of A:

$$AV = VD, \quad A(\vec{v}_1\ \vec{v}_2) = (\vec{v}_1\ \vec{v}_2)\begin{pmatrix}\lambda_1 & 0 \\ 0 & \lambda_2\end{pmatrix},$$

$$(A\vec{v}_1\ \ A\vec{v}_2) = (\lambda_1\vec{v}_1\ \ \lambda_2\vec{v}_2)$$

$$\begin{cases} A\vec{v}_1 = \lambda_1\vec{v}_1 \\ A\vec{v}_2 = \lambda_2\vec{v}_2 \end{cases}$$

Not all matrices have a spectral factorization of this form.

$$A = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}.$$

Theorem Suppose $A = A^T$. Then A has a spectral factorization

$[D, V] = \text{eig}(A)$

$$A = V D V^{-1}.$$

Furthermore, V can be chosen so that $V^T = V^{-1}$.

Computing Eigenvalues

For deep mathematical reasons, the numerical computation of eigenvalues has to be iterative. We will focus on one algorithm. The confusingly named QR eigenvalue algorithm:

$$A_0 = A$$
$$[Q_k, R_k] = qr(A_k) \qquad k = 0, 1, 2, \dots$$
$$A_{k+1} = R_k Q_k.$$

$$A_k = Q_k R_k, \quad A_{k+1} = R_k Q_k$$
$$R_k = Q_k^T A_k, \quad A_{k+1} = Q_k^T A_k Q_k$$

$\Rightarrow$ A_k and A_{k+1} have the same eigenvalues.

So, if $A_k \to D$, diagonal, as $k \to \infty$, then the diagonal entries of D are the eigenvalues of A.

We will only apply this algorithm to symmetric matrices.

Define the following norm on $m \times n$ matrices (!!!)

$$\|A\|_1 = \max_{1 \le i \le m} \sum_{j=1}^{n} |a_{ij}| \quad \boxed{\text{norm}(A, 1)}$$

Notation

$$\text{diag}(A) = \begin{pmatrix} a_{11} & & \\ & a_{22} & \\ & & \ddots \end{pmatrix}$$

Ex.

$$\left\| \begin{pmatrix} -1 & 2 \\ 1 & 1 \end{pmatrix} \right\| = \max(3, 2) = 3.$$

Theorem Suppose $A = A^T$. Let A_k, $k = 0, 1, 2, \ldots$ be the iterates of the QR eigenvalue algorithm. Then

$$\lim_{k \to \infty} A_k = D = \begin{pmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{pmatrix}.$$

Furthermore, if

$$\|\text{diag}(A_k) - A_k\|_1 < \varepsilon$$

Then by the Gershgorin Circle Theorem, every eigenvalue λ of A satisfies

$$|\lambda - \mu| < \varepsilon \text{ for some eigenvalue}$$

μ of A_k.

Singular Value Decomposition

Eigenvalues exist only for square matrices. All non-zero matrices have at least one singular value.

<u>Theorem</u> Suppose A is an $m \times n$ matrix. Then there exists an $m \times m$ matrix U, an $m \times n$ matrix Σ and an $n \times n$ matrix V such that

$$A = U \Sigma V^T$$

where U, V are orthogonal and Σ is diagonal with non-negative diagonal entries

$$\begin{bmatrix} \\ A \\ \\ \end{bmatrix} = \begin{bmatrix} & & \\ & U & \\ & & \end{bmatrix} \begin{bmatrix} \ddots & \\ \Sigma & \end{bmatrix} \begin{bmatrix} V \end{bmatrix}$$

largest

$$\Sigma = \begin{bmatrix} \sigma_1 & & & & & \\ & \sigma_2 & & & & \\ & & \ddots & & & \\ & & & \sigma_k & & \\ & & & & 0 & \\ & & & & & \ddots \; 0 \end{bmatrix}$$

smallest

The non-zero diagonal entries of Σ are called singular values.

<u>Theorem</u> The singular values of A are the square root of the non-zero eigenvalues of $A^T A$.

<u>Definition</u> The condition number for an $m \times n$ matrix A is

$$\text{cond}(A) = \frac{\sigma_1}{\sigma_{\min(m,n)}}$$

provided that A has $\min(m,n)$ singular values.

The condition number measures how many digits of accuracy you expect to loose when solving $A\vec{x} = \vec{b}$

$$\text{cond}(A) \approx 10^8$$

Then

computed

true

$$\frac{\|\tilde{x} - \vec{x}\|_2}{\|\vec{x}\|_2} \approx 2.2 \times 10^{-16} \cdot \text{cond}(A)$$

machine pecision

- The 2-norm of a matrix is its largest singular value

$$\|A\|_2 = \sigma_1$$

&

$$\|A\|_2 = 0$$

if it has no singular values.

- The rank of a matrix is equal to the number of singular values it has.

- Let $A = U \Sigma V^T$, $\Sigma = \begin{bmatrix} \sigma_1 & & & \\ & \ddots & & \\ & & \sigma_\ell & \\ & & & 0 \\ & & & & \ddots \end{bmatrix}$

The rank k approximation of A is given by

$$A_k = U \Sigma_k V^T, \quad \Sigma_k = \begin{bmatrix} \sigma_1 & & & \\ & \ddots & & \\ & & \sigma_k & \\ & & & 0 \\ & & & & \ddots \end{bmatrix}.$$

i.e. you delete the last $\ell - k$ singular values.

Note: $A_k = \begin{bmatrix} \vec{u}_1 \cdots \vec{u}_k \end{bmatrix} \begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_k \end{bmatrix} \begin{bmatrix} \vec{v}_1^T \\ \vdots \\ \vec{v}_k^T \end{bmatrix}$ ← less storage than A!

Fact:

$$\|A - A_k\|_2 = \sigma_{k+1}$$

Just in case you are curious:

In our discussion of gradient descent, we made the mysterious choice:

$$h_k = \frac{(A^T\vec{b} - A^TA\vec{y}_k)^T(A^T\vec{b} - A^TA\vec{y}_k)}{(AA^T\vec{b} - AA^TA\vec{y}_k)^T(AA^T\vec{b} - AA^TA\vec{y}_k)}$$

Why? Suppose $A^TA\vec{x} = A^T\vec{b}$. Consider

$$\vec{y}_{k+1} - \vec{x} = \vec{y}_k - \vec{x} - h_k(A^TA\vec{y}_k - A^T\vec{b})$$

Choose h_k to make this small: Instead consider

$$A(\vec{y}_{k+1} - \vec{x}) = A(\vec{y}_k - \vec{x}) - h_k AA^TA(\vec{y}_k - \vec{x})$$

Take the 2-norm (squared)

$$\|A(\vec{y}_{k+1} - \vec{x})\|_2^2 = \|A\overbrace{(\vec{y}_k - \vec{x})}^{\vec{e}_k} - h_k AA^TA\overbrace{(\vec{y}_k - \vec{x})}^{\vec{e}_k}\|_2^2$$

$$= \vec{e}_k A^TA\vec{e}_k - 2h_k \vec{e}_k^{\,T} A^TAA^TA\vec{e}_k + h_k^2 \vec{e}_k^{\,T} A^TAA^TAA^TA\vec{e}_k$$

This is a quadratic polynomial in h_k. The minimum occurs for

$$h_k = \frac{\vec{e}_k^{\,T} A^TAA^TA\vec{e}_k}{\vec{e}_k^{\,T} A^TAA^TAA^TA\vec{e}_k}.$$

This looks complicated, but we can see that

$$A^TA\vec{e}_k = A^TA\vec{y}_k - A^T\vec{b} = \vec{r}_k$$

$$AA^TAe_k = A\vec{r}_k.$$

So

$$h_k = \frac{\vec{r}_k^{\,T}\vec{r}_k}{(A\vec{r}_k)^T A\vec{r}_k}$$

<u>Error analysis</u>:

$$\|A\vec{e}_{k+1}\|_2^2 = \|(A - h_k AA^TA)\vec{e}_k\|_2^2$$

$$= \|A\vec{e}_k\|_2^2 - 2\frac{\|\vec{r}_k\|_2^{\cancel{2}\,4}}{\|A\vec{r}_k\|_2^2}\cancel{\|A^TA\vec{e}_k\|_2^2} + \frac{\|\vec{r}_k\|_2^4}{\|A\vec{r}_k\|_2^{\cancel{4}\,2}}\cancel{\|AA^TA\vec{e}_k\|_2^2}$$

$$A^TA\vec{e}_k = \vec{r}_k$$

$$= \|A\vec{e}_k\|_2^2 - \frac{\|\vec{r}_k\|_2^4}{\|A\vec{r}_k\|_2^2}, \qquad \frac{\|\vec{r}_k\|_2^2}{\|A\vec{r}_k\|_2^2} = \frac{\vec{r}_k^{\,T}\vec{r}_k}{\vec{r}_k A^TA\vec{r}_k} \geq \frac{1}{\lambda_{max}(A^TA)}$$

$$\leq \|A\vec{e}_k\|_2^2 - \|\vec{r}_k\|_2^2\frac{1}{\lambda_{max}(A^TA)}$$

$$= \|A\vec{e}_k\|_2^2 \left(1 - \frac{1}{\lambda_{max}(A^TA)} \frac{\|\vec{r}_k\|_2^2}{\|A\vec{e}_k\|_2^2} \right)$$

$$\frac{\|\vec{r}_k\|_2^2}{\|A\vec{e}_k\|_2^2} = \frac{\|\vec{r}_k\|_2^2}{\|A(A^TA)^{-1}\vec{r}_k\|_2^2} = \frac{\vec{r}_k^T\vec{r}_k}{\vec{r}_k^T \cancel{(A^TA)^{-1}} \cancel{(A^TA)} (A^TA)^{-1} r_k}$$

$$\geq \lambda_{min}(A^TA).$$

$$\|A\vec{e}_{k+1}\|_2 \leq \sqrt{1 - \frac{1}{(K(A))^2}} \, \|A\vec{e}_k\|_2 .$$