- Midterm → week 8 (Friday)
- HW2 : corrections → Q-4: $\|A\|_2^2 \leq \|A\|_1 \cdot \|A\|_\infty$

Q-5: Suppose $1 - \|\Delta A\| \cdot \|A^{-1}\| > 0$

CS450: Numerical Analysis

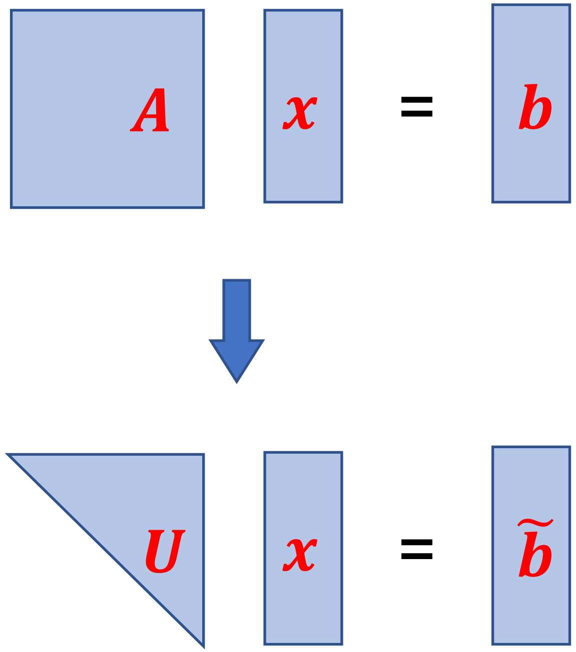
Howard Hao Yang

Assistant Professor, ZJU-UIUC Institute

27/10/2023

Solving Linear Systems

- For a square matrix $\boldsymbol{A}$, how to (systematically) solve for $\boldsymbol{Ax} = \boldsymbol{b}$?
  - Transform it into one whose solution is the same but easier to compute



Triangular Linear Systems

- What type of linear system is easy to solve?
  - Systems that form triangular matrices

$$
\begin{aligned}
a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= a_{1,n+1} \\
a_{22}x_2 + \cdots + a_{2n}x_n &= a_{2,n+1} \\
\ddots \qquad \vdots \qquad & \quad \vdots \\
a_{nn}x_n &= a_{n,n+1}
\end{aligned}
$$

  - Back-substitution

$$x_n = \frac{a_{n,n+1}}{a_{nn}},$$

$$x_i = \frac{a_{i,n+1} - \sum_{j=i+1}^{n} a_{ij}x_j}{a_{ii}}, \quad i = n-1, \ldots, 1$$

Elementary Elimination Matrices

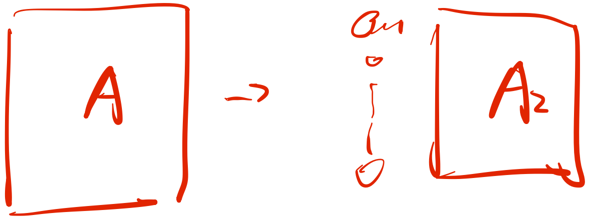
- More generally, we can annihilate all entries below k-th in a n-dimensional vector $\boldsymbol{a}$ by transformation

$$\boldsymbol{M}_k\,\boldsymbol{a} = \begin{bmatrix} 1 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 1 & 0 & \cdots & 0 \\ 0 & \cdots & -m_{k+1} & 1 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & -m_n & 0 & \cdots & 1 \end{bmatrix} \begin{bmatrix} a_1 \\ \vdots \\ a_k \\ a_{k+1} \\ \vdots \\ a_n \end{bmatrix} = \begin{bmatrix} a_1 \\ \vdots \\ a_k \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

where $m_i = \frac{a_i}{a_k}$, $i = k+1, \dots, n$

- Divisor a_k, called pivot, must be nonzero

Gaussian Elimination



- To reduce a general linear system of equations $\boldsymbol{Ax} = \boldsymbol{b}$ into upper triangular form, do the following
  - Construct $\boldsymbol{M}_1$ based on the first column $\boldsymbol{a}_1$ of $\boldsymbol{A}$, multiply both sides of the system, i.e., $\boldsymbol{M}_1\boldsymbol{Ax} = \boldsymbol{M}_1\boldsymbol{b}$, which annihilates first column of $\boldsymbol{A}$ below first row
  - Construct $\boldsymbol{M}_2$ based on the second column $\tilde{\boldsymbol{a}}_1$ of $\boldsymbol{M}_1\boldsymbol{A}$, multiply both sides of the system, i.e., $\boldsymbol{M}_2\boldsymbol{M}_1\boldsymbol{Ax} = \boldsymbol{M}_2\boldsymbol{M}_1\boldsymbol{b}$, which annihilates second column of $\boldsymbol{A}$ below second row
  - Continue until $\boldsymbol{A}$ becomes upper triangular
- Finally, solve the system by back-substitution

Solving the Linear Least Square

- How to find $\boldsymbol{x}^* = \mathrm{argmin}_{\boldsymbol{x}} ||\boldsymbol{Ax} - \boldsymbol{b}||_2$ where $\boldsymbol{A} \in R^{m \times n}$? $(m > n)$
  - Since $m > n$, the minimizer generally does not attain a zero residual $\boldsymbol{Ax} - \boldsymbol{b}$
  - The square of residual error can be written as

$$Err(\boldsymbol{x}) = (\boldsymbol{Ax} - \boldsymbol{b})^T(\boldsymbol{Ax} - \boldsymbol{b})$$

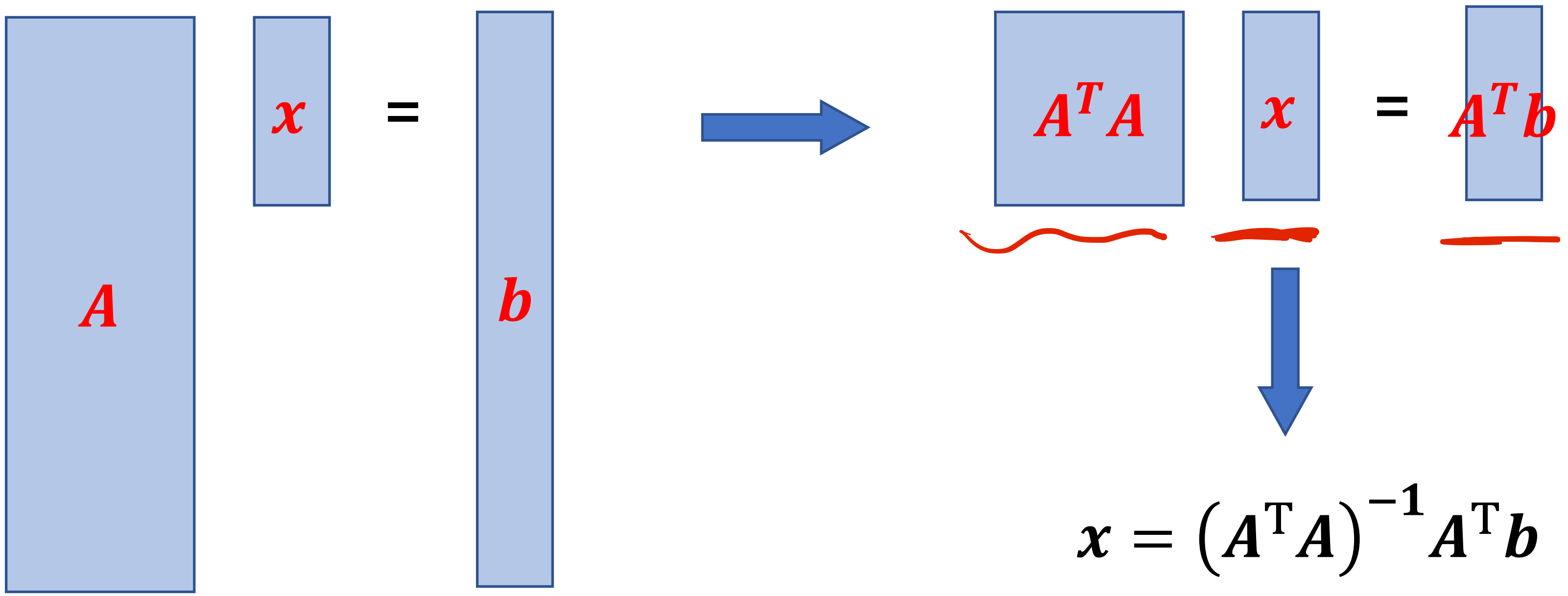
  - Let $\frac{\partial Err(\boldsymbol{x})}{\partial \boldsymbol{x}} = 2\boldsymbol{A}^T(\boldsymbol{Ax} - \boldsymbol{b}) = \boldsymbol{0}$, which is equivalent to $\left(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}\right)\boldsymbol{x} = \boldsymbol{A}^{\mathrm{T}}\boldsymbol{b}$
  - If $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$ is nonsingular, we obtain the solution as

$$\boldsymbol{x} = \left(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}\right)^{-\mathbf{1}}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{b}$$

  - This is known as the normal equation

Solving the Linear Least Square (cont'd)

- An algebraic view of the minimizer
  - Typically, the matrix $\boldsymbol{A}$ is a tall matrix, and the solution procedure of linear regression is somewhat like the following



$$\boldsymbol{x} = \left(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}\right)^{-1}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{b}$$

Solving the Normal Equation

- How to solve $\boldsymbol{A}^T\boldsymbol{A}\boldsymbol{x} = \boldsymbol{A}^T\boldsymbol{b}$?
- Remember the LU factorization? Since $\boldsymbol{A}^T\boldsymbol{A}$ is a square matrix, we can do LU factorization to it and then use back substitution?

$\kappa(A^TA) = (\kappa(A))^2$

Solving the Normal Equation

$\arg\min_x \|Ax-b\|_2^2$

$N(A) = N(A^TA)$

- How to solve $A^TAx = A^Tb$?
- Remember the LU factorization? Since A^TA is a square matrix, we can do LU factorization to it and then use back substitution?
- However, consider the following

$A^TA = \begin{bmatrix} 1 & \varepsilon & 0 \\ 1 & 0 & \varepsilon \end{bmatrix} \begin{bmatrix} 1 & 1 \\ \varepsilon & 0 \\ 0 & \varepsilon \end{bmatrix} = \begin{bmatrix} 1+\varepsilon^2 & 1 \\ 1 & 1+\varepsilon^2 \end{bmatrix}$

$\to 1$

$\text{rank}(A) = 2$

$$A = \begin{bmatrix} 1 & 1 \\ \epsilon & 0 \\ 0 & \epsilon \end{bmatrix}, \quad \text{where } 0 < \epsilon < \sqrt{\epsilon_{\text{mach}}}$$

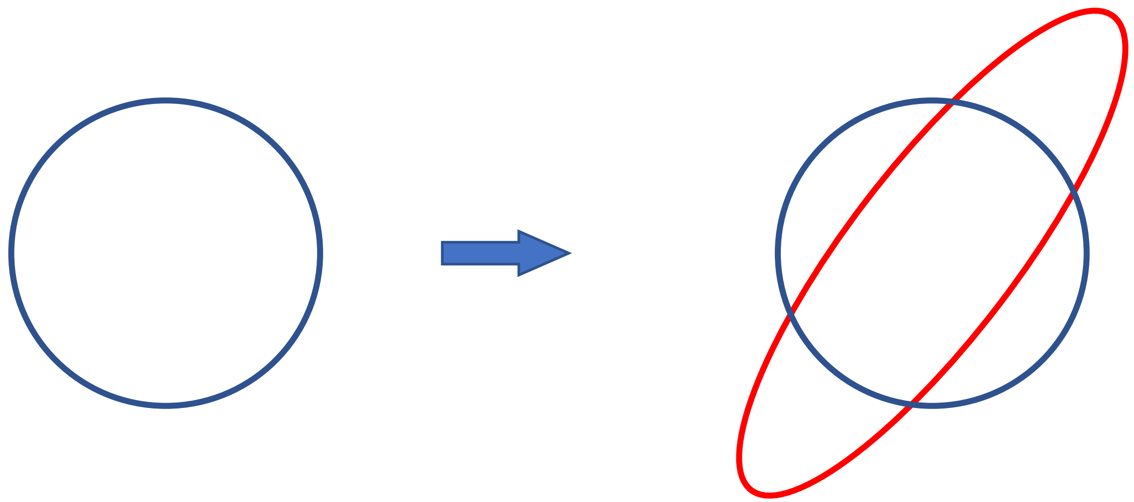
$\text{rank}(A^TA) = 1$

$$\Rightarrow A^TA = \begin{bmatrix} 1+\epsilon^2 & 1 \\ 1 & 1+\epsilon^2 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$$

- We want something more robust…

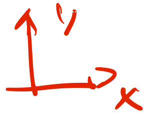
Recall: What does Matrix do (in general)?

- What does matrix multiplication do to a linear (sub)space?
  - Rotation & stretching
- What type of matrices do not distort an objective?
  - Orthogonal matrices



Orthogonal Matrices

$\begin{cases} x = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \\ y = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \end{cases}$



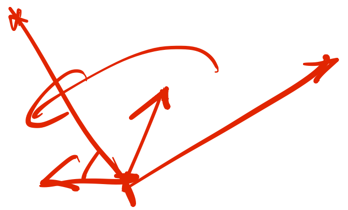
$z_i^T z_j = \begin{cases} 1, & i = j \\ 0, & i \neq j \end{cases}$

- Two vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ are said to be orthogonal if $\boldsymbol{x}^T\boldsymbol{y} = 0$
- A square matrix $\boldsymbol{Q}$ is said to be an orthogonal matrix if $\boldsymbol{Q}^T\boldsymbol{Q} = \boldsymbol{Q}\boldsymbol{Q}^T = \boldsymbol{I}$
- Nice property: if $\boldsymbol{Q}$ is an orthogonal matrix, then $\boldsymbol{Q}^{-1} = \boldsymbol{Q}^T$
- Multiplication of vector by orthogonal matrix preserves Euclidean norm

Qv

$$(\|\boldsymbol{Q}\boldsymbol{v}\|_2)^2 = (\boldsymbol{Q}^T\boldsymbol{v})^T\boldsymbol{Q}^T\boldsymbol{v} = \boldsymbol{v}^T\boldsymbol{Q}\boldsymbol{Q}^T\boldsymbol{v} = (\|\boldsymbol{v}\|_2)^2$$

- If $\boldsymbol{Q}$ is an orthogonal matrix, what would be $(\boldsymbol{Q}^T\boldsymbol{x})^T\boldsymbol{Q}^T\boldsymbol{y}$? $= x^TQQ^Ty = x^Ty$



Orthogonal Matrices (Cont'd)

- Examples and interpretation

$\mathbb{R}^2$

① $Q = \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix} \rightarrow Q^TQ = \begin{bmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{bmatrix} \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix}$

$= \begin{bmatrix} \cos^2\theta + \sin^2\theta = 1 & \underbrace{-\cos(\theta)\sin(\theta) + \sin(\theta)\cos(\theta)}_{=0} \\ 0 & 1 \end{bmatrix}$

rotate θ

② $Q = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \rightarrow Q^TQ = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$

$Q \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} x \\ -y \end{bmatrix} \rightarrow$ reflect

Wisdom from the Past

- When solving a linear system, we decompose the square matrix $\boldsymbol{A}$ into an LU form—we want something similar when $\boldsymbol{A}$ becomes a rectangular matrix
- For a rectangular matrix $\boldsymbol{A} \in R^{m\times n}$ with $m > n$, it can be decomposed as

$$\boldsymbol{A} = \underline{\boldsymbol{Q}} \begin{bmatrix} \boldsymbol{R} \\ \boldsymbol{O} \end{bmatrix}$$

where $\boldsymbol{Q}$ is an $m\times m$ orthogonal matrix and $\boldsymbol{R}$ is an upper triangular matrix

- This is known as the QR factorization

QR Factorization for LSQ

$\min_x \| Ax - b \|_2^2$

ZJU-UIUC INSTITUTE

- If we have a QR factorization of matrix $\boldsymbol{A} \in R^{m \times n}$, i.e.,

$$Q^T \boldsymbol{A} = \cancel{Q^T} \boldsymbol{Q} \begin{bmatrix} \boldsymbol{R} \\ \boldsymbol{O} \end{bmatrix}$$

$\min_x \| Q^T (Ax - b) \|_2^2$

$\min_x \left\| \begin{bmatrix} R \\ O \end{bmatrix} x - \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} \right\|_2^2$

- Then, the equation $\boldsymbol{Ax} \cong \boldsymbol{b}$ reduces to

$R x = c_1$

$$Q^T A \boldsymbol{x} = \begin{bmatrix} \boldsymbol{R} \\ \boldsymbol{O} \end{bmatrix} \boldsymbol{x} \cong \begin{bmatrix} \boldsymbol{c}_1 \\ \boldsymbol{c}_2 \end{bmatrix} = \boldsymbol{Q}^T \boldsymbol{b}$$

$\min_x \left\| \begin{bmatrix} Rx - c_1 \\ 0 \cdot x - c_2 \end{bmatrix} \right\|_2^2$

$\min_x \| Rx - c_1 \|_2^2 + \| c_2 \|_2^2$

- The following holds

$$\|\boldsymbol{Ax} - \boldsymbol{b}\|_2^2 = \|\boldsymbol{Q}^T(\boldsymbol{Ax} - \boldsymbol{b})\|_2^2$$
$$= \|\boldsymbol{Rx} - \boldsymbol{c}_1\|_2^2 + \|\boldsymbol{c}_2\|_2^2$$

QR Factorization for LSQ (Cont'd)

- If we have a QR factorization of matrix $\boldsymbol{A}$, i.e., $\boldsymbol{A} = \boldsymbol{Q}\boldsymbol{R}$, the following holds

$$\|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}\|_2^2 = \|\boldsymbol{R}\boldsymbol{x} - \boldsymbol{c}_1\|_2^2 + \|\boldsymbol{c}_2\|_2^2$$

- The second term is the residual error, which we cannot do anything on it
- The first term, we can choose $\boldsymbol{x}$ to minimize it, and the solution is given by solving $\boldsymbol{R}\boldsymbol{x} - \boldsymbol{c}_1 = \boldsymbol{0}$, which can be achieved via back substitution
- In this way, we avoid the cross-product matrix, saving us from round-off error issues
- The remaining question now is... how to perform QR factorization?

QR Factorization for Rectangular Matrices

$m>n$

- To decompose matrix $\boldsymbol{A} \in R^{m\times n}$ into the following

Q orthogonal

$$\boldsymbol{A} = \boldsymbol{Q}\begin{bmatrix}\boldsymbol{R}\\ \boldsymbol{O}\end{bmatrix} = A = \begin{bmatrix} | & | & & | \\ a_1 & a_2 & \cdots & a_n \\ | & | & & | \end{bmatrix}$$

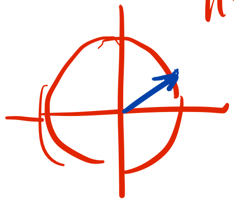
- Can we leverage the elementary elimination matrices?

$M_1 \to a_1$ $\quad M_1 A = \begin{bmatrix} a_{11} & \\ 0 & \\ \vdots & \\ 0 & \end{bmatrix} \to M_2,\quad M_2 M_1 A = \begin{bmatrix} a_{11} & a_{12} & \\ 0 & a_{22} & \\ 0 & 0 & \\ \vdots & \vdots & \\ 0 & 0 & \end{bmatrix}$

$(M_{n-1}\cdots M_1)A = \begin{bmatrix} \nabla \\ 0 \end{bmatrix}$

$A = (M_{n-1}\cdots M_1)^{-1}\begin{bmatrix} R \\ 0 \end{bmatrix}$

Householder Transformations



$\|x\|_2=1$

- (Definition) Given a unit-length vector $\boxed{\boldsymbol{x}} \in R^n$, i.e., $\boldsymbol{x}^T\boldsymbol{x} = 1$. The following $n \times n$ matrix is called a Householder transformation

$$\boldsymbol{H} = \boldsymbol{I} - 2\boldsymbol{x}\boldsymbol{x}^T$$

- The Householder transformation is symmetric and orthogonal
- Therefore $\boldsymbol{H}^{-1} = \boldsymbol{H}$

H^T

$$H^T = (I - 2xx^T)^T = I - 2(x^T)^T x^T = I - 2xx^T$$

$$H^T H = (I - 2xx^T)^T (I - 2xx^T)$$

$$= I - 4xx^T + 4x\underbrace{x^T x}_{=1}x^T = I$$

Householder Transformations (cont'd)

- Given a vector $\boldsymbol{x} \in R^n$ with $\boldsymbol{x}^T\boldsymbol{x} = 1$, the Householder transformation is

$$\boldsymbol{H} = \boldsymbol{I} - 2\boldsymbol{x}\boldsymbol{x}^T$$

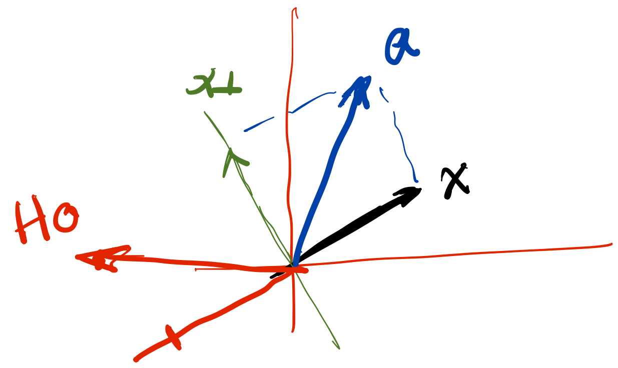
- Reflection property: for any vector $\boldsymbol{a} \in R^n$, $\boldsymbol{Ha}$ reflects $\boldsymbol{a}$ by the hyperplane perpendicular to $\boldsymbol{x}$

$$a = (a^T x)\, x + (a^T x_\perp) \cdot x_\perp$$

$$Ha = (I - 2xx^T)\, a$$

$$= (a^T x) \cdot x + (a^T x_\perp) \cdot x_\perp \quad - 2\, x \overbrace{x^T a}^{(a^T x)}$$

$$= -(a^T x) \cdot x + (a^T x_\perp) \cdot x_\perp$$

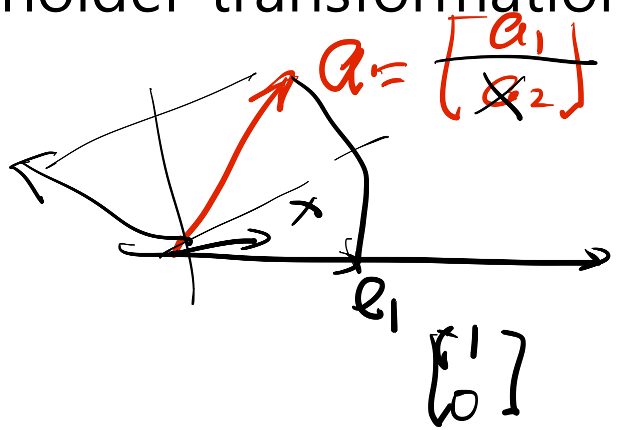


Householder Transformations (Cont'd)

$$\begin{bmatrix} a_1 \\ \hline 0 \\ 0 \\ 0 \end{bmatrix}$$

- Suppose we have $\boldsymbol{a} \in R^n$, and want to annihilate all elements below the first entry while preserving the norm
- One approach is to construct elimination matrix $\boldsymbol{M}_1$, but it is not orthogonal
- Can we leverage some ideas from the Householder transformation?

$$Ha = \quad e_1 \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$



Householder Transformations (Cont'd)

- Suppose we have $\boldsymbol{a} \in R^n$, and want to annihilate all elements below the first entry while preserving the norm
- Can we leverage some ideas from the Householder transformation?
- Problem: find vector $\boldsymbol{x} \in R^n$, such that $\boldsymbol{x}^T\boldsymbol{x} = 1$ and

$$\boldsymbol{H}\boldsymbol{a} = (\boldsymbol{I} - 2\boldsymbol{x}\boldsymbol{x}^T)\boldsymbol{a} = \alpha \boldsymbol{e}_1$$

where $\boldsymbol{e}_1 = (1,0,\cdots,0)^T$ and $\alpha = \|\boldsymbol{a}\|_2$

- Solution to this is

$$\boldsymbol{x} = \frac{\boldsymbol{a}}{\alpha} \pm \boldsymbol{e}_1$$

Example

- Do a Householder transform to the following vector

$$\boldsymbol{a} = \begin{bmatrix} 2 \\ 1 \\ 2 \end{bmatrix}$$

- We can calculate $\alpha = \|\boldsymbol{a}\|_2$ =3, hence