

# Linear Least Squares Problem

Consider an equation for a stretched beam:

$$Y = x_1 + x_2 T$$

Where  $x_1$  is the original length,  $T$  is the force applied and  $x_2$  is the inverse coefficient of stiffness.

Suppose that the following measurements were taken:

T	10	15	20
Y	11.60	11.85	12.25

Corresponding to the overcomplete system:

$$11.60 = x_1 + x_2 \cdot 10$$

$$11.85 = x_1 + x_2 \cdot 15$$

$$12.25 = x_1 + x_2 \cdot 20$$

- can not be satisfied exactly...

# Linear Least Squares Problem

## Problem:

Given  $A(m \times n)$ ,  $m \geq n$ ,  $b(m \times 1)$  find  $\mathbf{x}(n \times 1)$  to minimize

$$\|A\mathbf{x}-b\|_2.$$

- If  $m > n$ , we have more equations than the number of unknowns, there is generally no  $x$  satisfying  $Ax=b$  exactly.
- This is an overcomplete *system*.

# Linear Least Squares

There are three different algorithms for computing the least square minimum.

1. Normal Equations (Cheap, less Accurate).
2. QR decomposition.
3. SVD (expensive, more reliable).

The first algorithm is the fastest and the least accurate among the three. On the other hand SVD is the slowest and most accurate.

# Normal Equations 1

Minimize the squared Euclidean norm of the residual vector:

$$\|r\|_2^2 = r^T r$$

$$r = b - Ax$$

$$\|r\|_2^2 = r^T r = (b - Ax)^T (b - Ax)$$

$$= b^T b - 2x^T A^T b + x^T A^T A x$$

To minimize we take the derivative with respect to  $x$  and set it to zero:

$$-2A^T b + 2A^T A x = 0$$

Which reduces to an  $(n \times n)$  linear system commonly known as NORMAL EQUATIONS:

$$A^T A x = A^T b$$

# Normal Equations 2

$$11.60 = x_1 + x_2 \quad 10$$

$$11.85 = x_1 + x_2 \quad 15$$

$$12.25 = x_1 + x_2 \quad 20$$

$$\min \|Ax - b\|_2$$

$$\begin{matrix} & A & x & & b \\ \left[ \begin{array}{|c|c|} \hline 1 & 10 \\ \hline 1 & 15 \\ \hline 1 & 20 \\ \hline \end{array} \right] & \left[ \begin{array}{|c|} \hline x_1 \\ \hline x_2 \\ \hline \end{array} \right] & = & \left[ \begin{array}{|c|} \hline 11.60 \\ \hline 11.85 \\ \hline 12.25 \\ \hline \end{array} \right] \end{matrix}$$

# Normal Equations 3

We must solve the system

$$A^T A x = A^T b$$

For the following values

$$A = \begin{bmatrix} 1 & 10 \\ 1 & 15 \\ 1 & 20 \end{bmatrix}$$

$$A^T = \begin{bmatrix} 1 & 1 & 1 \\ 10 & 15 & 20 \end{bmatrix}$$

$$A^T \cdot A = \begin{bmatrix} 3 & 45 \\ 45 & 725 \end{bmatrix}$$

$$B = \begin{bmatrix} 11.60 \\ 11.85 \\ 12.25 \end{bmatrix}$$

$$x = (A^T \cdot A)^{-1} \cdot A^T \cdot b = \begin{bmatrix} 10.925 \\ 0.650 \end{bmatrix}$$

Math for CS

$(A^T A)^{-1} A^T$  is called a Pseudo-inverse

Lecture 4

6

# QR factorization 1

- A matrix  $Q$  is said to be orthogonal if its columns are orthonormal, i.e.  $Q^T \cdot Q = I$ .
- Orthogonal transformations preserve the Euclidean norm since

$$\|Qx\|_2^2 = (Qx)^T(Qx) = x^T Q^T Q x = x^T x = \|x\|_2^2$$

- Orthogonal matrices can transform vectors in various ways, such as rotation or reflections but they do not change the Euclidean length of the vector. Hence, they preserve the solution to a linear least squares problem.

# QR factorization 2

Any matrix  $\mathbf{A}_{(m \cdot n)}$  can be represented as

$$\mathbf{A} = \mathbf{Q} \cdot \mathbf{R}$$

, where  $\mathbf{Q}_{(m \cdot n)}$  is orthonormal and  $\mathbf{R}_{(n \cdot n)}$  is upper triangular:

$$\left[ \vec{a}_1 \mid \dots \mid \vec{a}_n \right] = \left[ \vec{q}_1 \mid \dots \mid \vec{q}_n \right] \cdot \begin{bmatrix} r_{11} & r_{12} & \dots & r_{1n} \\ \mathbf{0} & r_{11} & \dots & \dots \\ \mathbf{0} & \mathbf{0} & \dots & \dots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & r_{nn} \end{bmatrix}$$

# QR factorization 2

- Given  $A$ , let its **QR** decomposition be given as  $A=Q \cdot R$ , where  $Q$  is an  $(m \times n)$  orthonormal matrix and  $R$  is upper triangular.
- QR factorization transform the linear least square problem into a triangular least squares.

$$Q \cdot R \cdot x = b$$

$$R \cdot x = Q^T \cdot b$$

$$x = R^{-1} \cdot Q^T \cdot b$$

Matlab Code:

```
>> [ Q, R ] = qr(A)
Q =
   -0.5774    0.7071    0.4082
   -0.5774     0     -0.8165
   -0.5774   -0.7071    0.4082

R =
   -1.7321   -25.9808
         0    -7.0711
         0         0

>> x = R \ (Q' * y)
x =
   10.9250
    0.0650
```

# Singular Value Decomposition

- Normal equations and **QR** decomposition only work for fully-ranked matrices (i.e.  $\text{rank}(A) = n$ ). If  $A$  is rank-deficient, that there are infinite number of solutions to the least squares problems and we can use algorithms based on SVD's.

- Given the SVD:  $A = U\Sigma V^T$

$U_{(m \times m)}$ ,  $V_{(n \times n)}$  are orthogonal

$\Sigma$  is an  $(m \times n)$  diagonal matrix (singular values of  $A$ )

The minimal solution corresponds to:

$$x = V\Sigma^{-1}U^Tb$$

# Singular Value Decomposition

Matlab Code:

```
>> [U,S,U] = svd(A)
U =
    0.3723   -0.8335    0.4082
    0.5572   -0.1510   -0.8165
    0.7422    0.5314    0.4082

S =
   26.9777         0
         0    0.4540
         0         0

U =
    0.0620   -0.9981
    0.9981    0.0620

>> x = (U * S * U') \ y
x =
   10.9250
    0.0650
```

# Linear algebra review - SVD

Fact: for every  $\mathbf{A}_{m \times n}$ ,  $\text{rank}(\mathbf{A}) = p$ ,  
there exist  $\mathbf{U}_{m \times p}$ ,  $\mathbf{V}_{n \times p}$ ,  $\mathbf{\Sigma}_{p \times p}$  such that

$$\mathbf{U}^T \mathbf{U} = \mathbf{I}$$

$$\mathbf{V}^T \mathbf{V} = \mathbf{I}$$

$$\mathbf{\Sigma} = \text{diag}(\sigma_1, \sigma_2, \dots, \sigma_p), \quad \sigma_i > 0$$

$$\mathbf{A} = \mathbf{U} \mathbf{\Sigma} \mathbf{V}^T = \sum_{i=1}^p \sigma_i (\mathbf{u}_i \cdot \mathbf{v}_i^T)$$

$\Rightarrow$

$$\mathbf{A}^T \mathbf{A} = \mathbf{V} \mathbf{\Sigma}^2 \mathbf{V}^T$$

$$\mathbf{A} \mathbf{A}^T = \mathbf{U} \mathbf{\Sigma}^2 \mathbf{U}^T$$

# Approximation by a low-rank matrix

Fact II: let  $\mathbf{A}_{m \times n}$  have rank  $p$ .

let  $\mathbf{A} = \mathbf{U} \mathbf{\Sigma} \mathbf{V}^T = \sum_{i=1}^p \sigma_i (\mathbf{u}_i \cdot \mathbf{v}_i^T)$  be the SVD of  $\mathbf{A}$ ,  $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_p$

Then  $\tilde{\mathbf{A}} = \sum_{i=1}^r \sigma_i (\mathbf{u}_i \cdot \mathbf{v}_i^T) = \mathbf{U} \tilde{\mathbf{\Sigma}} \mathbf{V}^T$ ,  $\tilde{\mathbf{\Sigma}} = \text{diag}(\sigma_1, \sigma_2, \dots, \sigma_r)$ ,  $r < p$

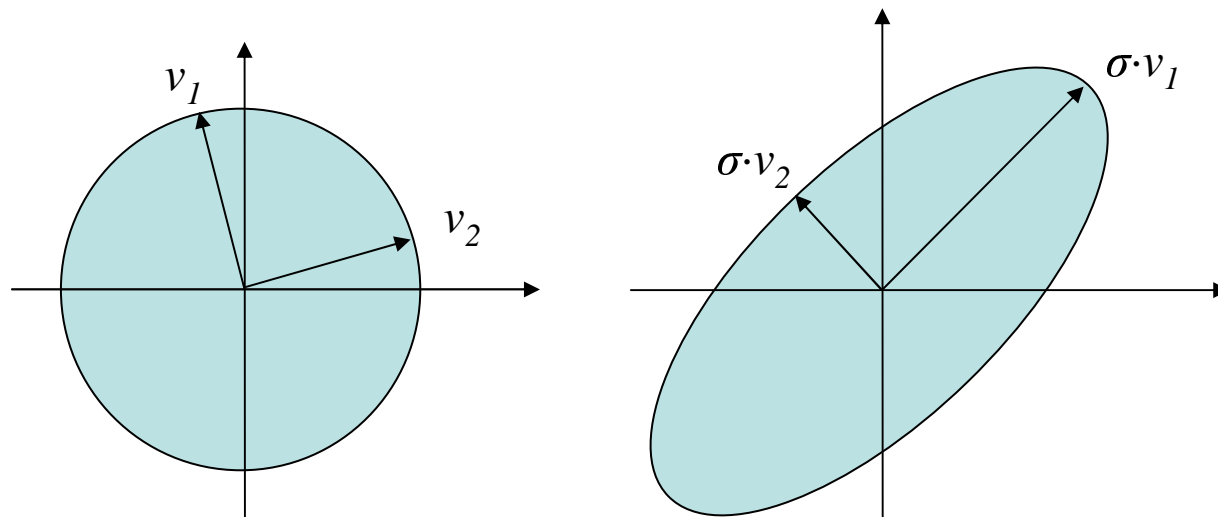
Is the best rank  $r$  approximation to  $\mathbf{A}$  in  $\|\cdot\|_2$ :

$$\min_{\mathbf{X}: \text{rank}(\mathbf{X})=r} \{\|\mathbf{A} - \mathbf{X}\|_2\} = \|\mathbf{A} - \tilde{\mathbf{A}}\|_2 = \|\mathbf{U}(\mathbf{\Sigma} - \tilde{\mathbf{\Sigma}})\mathbf{V}^T\|_2 = \sigma_{r+1}$$

# Geometric Interpretation of the SVD

$$\vec{b} = A_{(m \times n)} \cdot \vec{x}$$

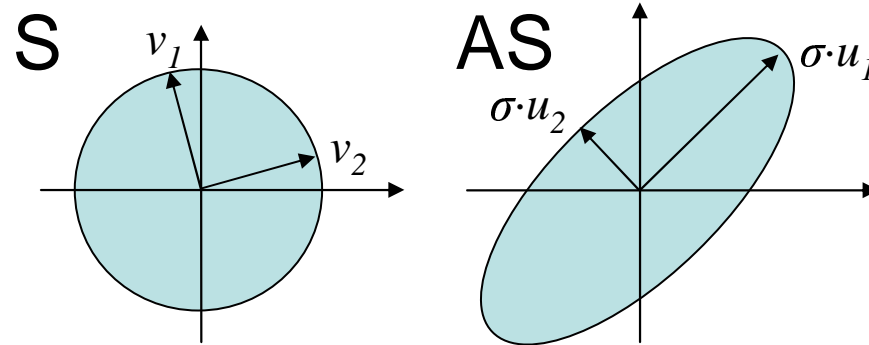
The image of the unit sphere under any  $m \times n$  matrix is a *hyperellipse*



# Left and Right singular vectors

We can define the properties of  $A$  in terms of the shape of  $AS$

$$AS = A_{(m \times n)} \cdot S$$



*Singular values of  $A$*  are the lengths of principal axes of  $AS$ , usually written in non-increasing order  $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_n$

$n$  *left singular vectors of  $A$*  are the unit vectors  $\{u_1, \dots, u_n\}$ , oriented in the directions of the principal semiaxes of  $AS$  numbered in correspondance with  $\{\sigma_i\}$

$n$  *right singular vectors of  $A$*  are the unit vectors  $\{v_1, \dots, v_n\}$ , of  $S$ , which are the preimages of the principal semiaxes of  $AS$ :  $Av_i = \sigma_i u_i$

# Singular Value Decomposition

$$Av_i = \sigma_i u_i, \quad 1 \leq i \leq n$$

$$\begin{bmatrix} A \\ (m,n) \end{bmatrix} \begin{bmatrix} | & | & \dots & | \\ v_1 & v_2 & \dots & v_n \\ | & | & \dots & | \\ (n,n) \end{bmatrix} = \begin{bmatrix} | & | & \dots & | \\ u_1 & u_2 & \dots & u_n \\ | & | & \dots & | \\ (m,n) \end{bmatrix} \begin{bmatrix} \sigma_1 & & & \\ & \sigma_2 & & \\ & & \dots & \\ & & & \sigma_n \\ (n,n) \end{bmatrix}$$

$$AV = U\Sigma$$

$$A = U\Sigma V^* \quad - \text{Singular Value decomposition}$$

Matrices U,V are orthogonal and  $\Sigma$  is diagonal

# Matrices in the Diagonal Form

Every matrix is diagonal in appropriate basis:

$$\vec{\mathbf{b}} = \mathbf{A}_{(m \times n)} \cdot \vec{\mathbf{x}}$$

Any vector  $\mathbf{b}_{(m,1)}$  can be expanded in the basis of left singular vectors of A  $\{u_i\}$ ;

Any vector  $\mathbf{x}_{(n,1)}$  can be expanded in the basis of right singular vectors of A  $\{v_i\}$ ;

Their coordinates in these new expansions are:

$$\vec{\mathbf{b}}' = \mathbf{U}^* \cdot \vec{\mathbf{b}}; \quad \vec{\mathbf{x}}' = \mathbf{V}^* \cdot \vec{\mathbf{x}}$$

Then the relation  $\mathbf{b} = \mathbf{A}\mathbf{x}$  can be expressed in terms of  $\mathbf{b}'$  and  $\mathbf{x}'$ :

$$\vec{\mathbf{b}} = \mathbf{A}\vec{\mathbf{x}} \Leftrightarrow \mathbf{U}^* \cdot \vec{\mathbf{b}} = \mathbf{U}^* \mathbf{A}\vec{\mathbf{x}} = \mathbf{U}^* \mathbf{U} \mathbf{\Sigma} \mathbf{V}^* \vec{\mathbf{x}} \Leftrightarrow \vec{\mathbf{b}} = \mathbf{\Sigma} \vec{\mathbf{x}}'$$

# Rank of A

Let  $p = \min\{m, n\}$ , let  $r \leq p$  denote the number of nonzero singular values of A,

Then:

The rank of A equals to r, the number of nonzero singular values

Proof:

The rank of a diagonal matrix equals to the number of its nonzero entries, and in the decomposition  $A = U\Sigma V^*$ , U and V are of full rank

# Determinant of A

For  $A_{(m,m)}$ ,

$$|\mathbf{det}(A)| = \prod_{i=1}^m \sigma_i$$

Proof:

The determinant of a product of square matrices is the product of their determinants. The determinant of a Unitary matrix is 1 in absolute value, since:  $U^*U=I$ . Therefore,

$$|\mathbf{det}(A)| = |\mathbf{det}(U\Sigma V^*)| = |\mathbf{det}(U)| \|\mathbf{det}(\Sigma)\| |\mathbf{det}(V^*)| = |\mathbf{det}(\Sigma)| = \prod_{i=1}^m \sigma_i$$

# A in terms of singular vectors

For  $A_{(m,n)}$ , can be written as a sum of  $r$  rank-one matrices:

$$A = \sum_{j=1}^r \sigma_j u_j v_j^* \quad (1)$$

Proof:

If we write  $\Sigma$  as a sum of  $\Sigma_i$ , where  $\Sigma_i = \text{diag}(0, \dots, \sigma_i, \dots, 0)$ , then

(1)

Follows from

$$A = U \Sigma V^* \quad (2)$$

# Norm of the matrix

The  $L_2$  norm of the vector is defined as:

$$\|\vec{x}\|_2 = \sqrt{\vec{x}^T \vec{x}} = \sqrt{\sum_{i=1}^n |x_i|^2} \quad (1)$$

The  $L_2$  norm of the matrix is defined as:

$$\|A\|_2 = \sup \frac{\|A\vec{x}\|_2}{\|\vec{x}\|_2}$$

Therefore

$$\|A\|_2 = \max(\lambda_i)$$

, where  $\lambda_i$  are the eigenvalues  $A\vec{x}_i = \lambda_i \vec{x}_i$

# Matrix Approximation in SVD basis

For any  $v$  with  $0 \leq v \leq r$ , define

$$A_v = \sum_{j=1}^v \sigma_j u_j v_j^* \quad (1)$$

If  $v=p=\min\{m,n\}$ , define  $\sigma_{v+1}=0$ . Then

$$\|A - A_v\|_2 = \inf_{\substack{B \in \mathbb{C}^{m \times n} \\ \text{rank}(B) \leq v}} \|A - B\|_2 = \sigma_{v+1}$$