**Rob 501 Handout: Grizzle**

**The SVD and Numerical Rank of a Matrix**

**Motivation:** In abstract linear algebra, a set of vectors is either linearly independent or not. There is nothing in between. For example, the set of vectors

$$\left\{ v^1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, v^2 = \begin{bmatrix} 0.999 \\ 1 \end{bmatrix} \right\}$$

is linearly independent. In this case, you look at it and say, yes, BUT, the vectors are "almost" dependent because when I take the determinant

$$\det \begin{bmatrix} 1 & 0.999 \\ 1 & 1 \end{bmatrix} = 0.001,$$

I get something pretty small, so I am OK with calling them dependent. Well, what about the set

$$\left\{ v^1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, v^2 = \begin{bmatrix} 10^4 \\ 1 \end{bmatrix} \right\}?$$

When you form the matrix and check the determinant, you get

$$\det \begin{bmatrix} 1 & 10^4 \\ 0 & 1 \end{bmatrix} = 1,$$

which seems pretty far from zero. So are these vectors "adequately" linearly independent?

**Maybe not!** Let's note that

$$\begin{bmatrix} 1 & 10^4 \\ 0 & 1 \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 10^{-4} & 0 \end{bmatrix} = \begin{bmatrix} 1 & 10^4 \\ 10^{-4} & 1 \end{bmatrix},$$

is clearly singular! Hence, we can add a very small perturbation to our vectors and make them dependent! This cannot be good! :(

**Question:** How to quantify the statement,"the rank is *nearly* 1" or more generally, how to quantify that a set of vectors is *nearly* linearly dependent?

**Answer:** The Singular Value Decomposition (**SVD**).

**Properties of the Singular Value Decomposition or SVD**
**(Based on a handout of Prof. Freudenberg)**

A good reference on numerical linear algebra is

G. H. Golub and C. F. van Loan, *Matrix Computations*, The Johns Hopkins University Press, 1983.

**Background required:** inner products, orthogonal vectors, norms.

**Remark:** In lecture, we will develop the SVD for real matrices. In practice, you may have a need to deal with matrices that have complex entries, so the handout also does things for inner products in $\mathbb{C}^n$. The generalization of a real *symmetric* matrix is called a *Hermitian* matrix. And the generalization of a real *orthogonal* matrix is called a *unitary matrix*. These are developed below.

**Real matrices:** The statement and proof of the SVD Theorem in the case of real matrices is given at the end of the handout.

**Hermitian:** Consider $x \in \mathbb{C}^n$. Then we define the vector "$x$ Hermitian" by $x^H := \bar{x}^\top$.That is, $x^H$ is the complex conjugate transpose of $x$. Similarly, for a matrix $A \in \mathbb{C}^{m\times n}$, we define $A^H \in \mathbb{C}^{n\times m}$ by $\bar{A}^\top$. We say that a square matrix $A \in \mathbb{C}^{n\times n}$ is a *Hermitian matrix* if $A = A^H$.

**Important things to note:**

- Similar to $A^\top A$ for real matrices, when $A$ is complex, $A^H A$ has e-values that are real and non-negative. The proof is similar to things we have done in lecture; if you care to see it, you can find it online.
- In MATLAB, $A' = A^H$. Yikes! It is not the ordinary transpose? No, it is the complex conjugate transpose. If you want the ordinary transpose, use `transpose(A)`.

```
>> A=[j,0;0,-j]

A =

   0 + 1.0000i        0
   0                  0 - 1.0000i

>> A'

ans =

   0 - 1.0000i        0
   0                  0 + 1.0000i
```

**Inner product on $\mathbb{C}^n$:** Given $x, y \in \mathbb{C}^n$. Let the elements $x$ and $y$ be noted

$$x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \quad \text{and} \quad y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}.$$

Then the Euclidean inner product is defined as

$$\begin{aligned} < x, y >: &= x^H y && (1) \\ &= \bar{x}_1 y_1 + \bar{x}_2 y_2 + \cdots + \bar{x}_n y_n && (2) \end{aligned}$$

We note that this puts the linearity on the right side of the "bracket", but as we have noted in HW, both definitions are common.

Euclidean vector norm: As in class, the vector norm associated with this inner product is given by

$$||x||_2 := \sqrt{< x, x >} \tag{3}$$

$$= \sqrt{\sum_{i=1}^{n} |x_i|^2} \tag{4}$$

We often omit the subscript "2" when we are discussing the Euclidean norm (or "2-norm") exclusively.

Euclidean matrix norm: Given $A \in \mathbb{C}^{m\times n}$. Then the *matrix norm induced by the Euclidean vector norm* is given by:

$$||A||_2 := \max_{x^H x=1} ||Ax|| \tag{5}$$

$$= \max_{x^H x=1} \sqrt{x^H A^H A x} \tag{6}$$

$$= \sqrt{\max_{x^H x=1} x^H A^H A x} \tag{7}$$

$$= \sqrt{\lambda_{\max}(A^H A)} \tag{8}$$

where $\lambda_{\max}(A^H A)$ denotes the largest eigenvalue of the matrix $A^H A$. **(As noted above, all the eigenvalues of a matrix having the form $A^H A$ are real and non-negative.)(Also, recall HW 2)**

Orthogonality: Two vectors $x, y \in \mathbb{C}^n$ are *orthogonal* if $< x, y >= 0$.

Orthonormal Set: A collection of vectors $\{x_1, x_2, \cdots, x_m\} \in \mathbb{C}^n$ is said to be an *orthonormal set* if

$$< x_i, x_j >= \begin{cases} 0, & i \neq j \\ 1, & i = j \end{cases} \qquad \text{(Hence } ||x_i|| = 1, \forall\, i.\text{)}$$

**Unitary Matrix:** A matrix $U \in \mathbb{C}^{n\times n}$ is *unitary* if $U^H U = U U^H = I_n$.

**Fact:** If $U$ is a unitary matrix, then the columns of $U$ form an orthonormal basis (ONB) for $\mathbb{C}^n$.

Proof of Fact: Denote the columns of $U$ as $U = \begin{bmatrix} u_1 & u_2 & \cdots & u_n \end{bmatrix}$. Then

$$U^H U = \begin{bmatrix} u_1^H \\ u_2^H \\ \vdots \\ u_n^H \end{bmatrix} \begin{bmatrix} u_1 & u_2 & \cdots & u_n \end{bmatrix} = \begin{bmatrix} u_1^H u_1 & u_1^H u_2 & \cdots & u_1^H u_n \\ u_2^H u_1 & u_2^H u_2 & \cdots & u_2^H u_n \\ \vdots & \vdots & \ddots & \vdots \\ u_n^H u_1 & u_n^H u_2 & \cdots & u_n^H u_n \end{bmatrix} = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}$$

**For real matrices, unitary is the same thing as orthogonal.**

**Example:**

```
U =

    0.1259    0.9920
    0.9920   -0.1259

>>U*U'

ans =

    1.0000    0.0000
    0.0000    1.0000
```

**Unitary matrices are effectively rotation matrices:** they do not change the length of a vector, nor the angle between two vectors. Indeed,

1) From $U^H U = UU^H = I_n$, it follows that $U^{-1} = U^H$

2) Let's compute the inner product of $Ux$ and $Uy$:

$$< Ux, Uy > := (Ux)^H Uy = x^H U^H Uy = x^H y =: < x, y >$$

3) It follows that
(a) norm of $Ux$ equals the norm of $x$:

$$||Ux||^2 := < Ux, Ux > = < x, x > =: ||x||^2$$

(b) angle between $x$ and $y$ is the same as the angle between $Ux$ and $Uy$:

$$\cos(\angle(x, y)) := \frac{< x, y >}{||x||\ ||y||} = \frac{< Ux, Uy >}{||Ux||\ ||Uy||} =: \cos(\angle(Ux, Uy))$$

4) All of the e-values of $U$ have magnitude 1. Indeed, suppose that $\lambda$ is an e-value with e-vector $v$: $Uv = \lambda v$

Applying norms to both sides of the above yields: $||Uv|| = ||\lambda v||$

But, by item (3) above and properties of norms:

$$||Uv|| = ||v|| \text{ and } ||\lambda v|| = |\lambda|\ ||v||$$

which, with the above, implies $|\lambda| = 1$.

## Singular Value Decomposition Theorem

**Complex matrices:** Consider $A \in \mathbb{C}^{m\times n}$. Then there exist unitary matrices

$$\begin{aligned} U &= \begin{bmatrix} u_1 & u_2 & \cdots & u_m \end{bmatrix} \\ V &= \begin{bmatrix} v_1 & v_2 & \cdots & v_n \end{bmatrix} \end{aligned}$$

such that

$$A = \begin{cases} U \begin{bmatrix} \Sigma \\ \\ 0 \end{bmatrix} V^H, & m \geq n \\ U \begin{bmatrix} \Sigma & 0 \end{bmatrix} V^H, & m \leq n \end{cases}$$

where

$$\Sigma = \begin{bmatrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_p \end{bmatrix}, \; p = \min(m,n)$$

and

$$\sigma_1 \geq \sigma_2 \geq \cdots \sigma_p \geq 0$$

Terminology: We refer to $\sigma_i$ as the $i$'th singular value, to $u_i$ as the $i$'th left singular vector, and to $v_i$ as the $i$'th right singular vector.

**Real Matrices:** Consider $A \in \mathbb{R}^{m\times n}$. Then there exist orthogonal matrices

$$\begin{aligned} U &= \begin{bmatrix} u_1 & u_2 & \cdots & u_m \end{bmatrix} \\ V &= \begin{bmatrix} v_1 & v_2 & \cdots & v_n \end{bmatrix} \end{aligned}$$

such that

$$A = \begin{cases} U \begin{bmatrix} \Sigma \\ \\ 0 \end{bmatrix} V^\top, & m \geq n \\ U \begin{bmatrix} \Sigma & 0 \end{bmatrix} V^\top, & m \leq n \end{cases}$$

where

$$\Sigma = \begin{bmatrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_p \end{bmatrix}, \ p = \min(m,n)$$

and

$$\sigma_1 \geq \sigma_2 \geq \cdots \sigma_p \geq 0$$

Terminology: We refer to $\sigma_i$ as the $i$'th singular value, to $u_i$ as the $i$'th left singular vector, and to $v_i$ as the $i$'th right singular vector.

**Theorem:** rank(A) = number of nonzero singular values.

**Idea:** numerical rank(A) = # of nonzero singular values larger than a threshold?

**Matlab:** help SVD

```
SVD  Singular value decomposition.
 [U,S,V] = SVD(X) produces a diagonal matrix S, of the same
 dimension as X and with non-negative diagonal elements in
 decreasing order, and unitary matrices U and V so that
 X = U*S*V'.

  By itself, SVD(X) returns a vector containing diag(S).
```

**Example 1:**

```
>> A=[1,2,3;2,1,3]

A =

     1     2     3
     2     1     3

>> [U,S,V]=svd(A)

U =

    0.7071    0.7071
    0.7071   -0.7071


S =

    5.1962         0         0
         0    1.0000         0


V =
    0.4082   -0.7071    0.5774
    0.4082    0.7071    0.5774
    0.8165   -0.0000   -0.5774

>> U'*U

ans =

    1.0000    0.0000
    0.0000    1.0000
```

```
>> A-U*S*V'

ans =

   1.0e-14 *

   -0.0222   -0.0444    0.0444
   -0.0888    0.1110    0.0444
```

**Example 2:**

```
>> A=[1,0.98;1,1]

A =

    1.0000    0.9800
    1.0000    1.0000

>> [U,S,V]=svd(A)

U =

    0.7036   -0.7106
    0.7106    0.7036


S =

    1.9900         0
         0    0.0100
```

```
V =

    0.7106   -0.7036
    0.7036    0.7106


>> U'*U

ans =

    1.0000    0.0000
    0.0000    1.0000

>> A-U*S*V'

ans =

   1.0e-15 *

         0    0.5551
         0    0.4441
```

> **Fact:** The <u>numerical rank</u> of $A$ is the number of singular values that are larger than a given threshold. Often the threshold is chosen as a percentage of the largest singular value.

**Example:** $5 \times 5$ **matrix**

$$A = \begin{bmatrix} -32.57514 & -3.89996 & -6.30185 & -5.67305 & -26.21851 \\ -36.21632 & -11.13521 & -38.80726 & -16.86330 & -1.42786 \\ -5.07732 & -21.86599 & -38.27045 & -36.61390 & -33.95078 \\ -36.51955 & -38.28404 & -19.40680 & -31.67486 & -37.34390 \\ -25.28365 & -38.57919 & -31.99765 & -38.36343 & -27.13790 \end{bmatrix}$$

```
[U,Sigma,V]=svd(A);
```

$$\Sigma = \begin{bmatrix} 132.459 & 0.00000 & 0.00000 & 0.00000 & 0.00000 \\ 0.00000 & 37.70811 & 0.00000 & 0.00000 & 0.00000 \\ 0.00000 & 0.00000 & 33.41836 & 0.00000 & 0.00000 \\ 0.00000 & 0.00000 & 0.00000 & 19.34060 & 0.00000 \\ 0.00000 & 0.00000 & 0.00000 & 0.00000 & 0.79164 \end{bmatrix}$$

Note that the smallest singular value $\sigma_5 = 0.79164$ is less than 1% of the largest singular value $\sigma_1 = 132.459$. In many cases, one might say that the numerical rank of $A$ was 4 instead of 5.

**This notion of numerical rank can be formalized by asking the following question:** Suppose $\text{rank}(A) = r$. How far away is $A$ from a matrix of rank strictly less than $r$?

**Fact:** Suppose that $\text{rank}(A) = r$, so that $\sigma_r$ is the smallest non-zero singular value. Then

(i) if an $n \times m$ matrix $E$ satisfies $||E|| < \sigma_r$, then $\text{rank}(A + E) = r$.

(ii) there exists $E$ with $||E|| = \sigma_r$ and $\text{rank}(A + E) < r$.

**Corollary:** Suppose $A$ is square and invertible. Then $\sigma_r$ measures the distance from $A$ to the nearest singular matrix.

**Example: Using A above**

```
>> d=diag(Sigma);
>> d(end)=0;
>> D=diag(d);
>> B=U*D*V';
>> E=A-B;
```

$$E = \begin{bmatrix} -0.04169 & 0.12122 & 0.09818 & -0.21886 & 0.05458 \\ 0.02031 & -0.05906 & -0.04784 & 0.10663 & -0.02659 \\ 0.01966 & -0.05716 & -0.04629 & 0.10320 & -0.02574 \\ 0.07041 & -0.20476 & -0.16584 & 0.36968 & -0.09220 \\ -0.08160 & 0.23728 & 0.19218 & -0.42839 & 0.10684 \end{bmatrix}$$

```
>> max(sqrt(eig(E'*E)))

 0.7916

 >> [U,Sigma,V]=svd(A-E);
```

$$\Sigma = \begin{bmatrix} 132.45977 & 0.00000 & 0.00000 & 0.00000 & 0.00000 \\ 0.00000 & 37.70811 & 0.00000 & 0.00000 & 0.00000 \\ 0.00000 & 0.00000 & 33.41836 & 0.00000 & 0.00000 \\ 0.00000 & 0.00000 & 0.00000 & 19.34060 & 0.00000 \\ 0.00000 & 0.00000 & 0.00000 & 0.00000 & 0.00000 \end{bmatrix}$$

I added a matrix with norm 0.7916 and made the (exact) rank drop from 4 to 5! How cool is that? It really shows that the matrix was close to a singular matrix.

<u>Another Example:</u>

```
>> N=100;A=[1,N;0,1]; [U,S,V]=svd(A); A,S

A =

     1   100
     0     1


S =

  100.0100         0
         0    0.0100
```

**Final comments:**

(a) We have not had the time to do anything with the nullspace and range of an $m \times n$ matrix $A$; they are important subspaces.

**Nullspace:** $\mathbf{N}(A) := \{x \in \mathbb{C}^n \mid Ax = 0\}$

**Range:** $\mathbf{R}(A) := \{y \in \mathbb{C}^m \mid \exists x \in \mathbb{C}^n \text{ such that } y = Ax\}$

(b) **Fact:** Let $[U, \Sigma, V] = \texttt{svd}(A)$; Then the columns of $U$ are a basis for $\mathbf{R}(A)$ and the columns of $V$ are a basis for $\mathbf{N}(A)$.

(c) The SVD can also be used to compute an "effective" range and an "effective" nullspace of a matrix.

(d) Suppose that $\sigma_1 \geq ... \geq \sigma_r > \epsilon \geq \sigma_{r+1} \geq ...\sigma_n \geq 0$, so that $r$ is the "effective" or numerical rank of $A$. (Note the $\epsilon$ inserted between $\sigma_r$ and $\sigma_{r+1}$ to denote the break point.)

(e) Let $\mathbf{R}_{\text{eff}}(A)$ and $\mathbf{N}_{\text{eff}}(A)$ denote the effective range and effective nullspace of $A$, respectively. Then we can calculate bases for these subspaces by choosing appropriate singular vectors:

$$\mathbf{R}_{\text{eff}}(A) := \text{span}\{u_1, ..., u_r\} \text{ and } \mathbf{N}_{\text{eff}}(A) := \text{span}\{v_{r+1}, ..., v_n\}.$$

**Def.** An $m \times n$ matrix $\Sigma$ is rectangular diagonal if $\Sigma_{ij} = 0$ for $i \neq j$. The diagonal of $\Sigma$ is

$$\operatorname{diag}(\Sigma) = (\Sigma_{11},\ \Sigma_{22},\ \cdots,\ \Sigma_{kk})$$

where $k = \min(m,\ n)$.

**Examples** Consider rectangular matrices

$$\Sigma_1 = \begin{bmatrix} 3 & 0 & 0 & 0 \\ 0 & 4 & 0 & 0 \\ 0 & 0 & -1 & 0 \end{bmatrix} \quad \text{and} \quad \Sigma_2 = \begin{bmatrix} 1 & 0 \\ 0 & -6 \\ 0 & 0 \end{bmatrix}$$

Then,

$$\operatorname{diag}(\Sigma_1) = \begin{bmatrix} 3 & 4 & -1 \end{bmatrix} \quad \text{and} \quad \operatorname{diag}(\Sigma_2) = \begin{bmatrix} 1 & -6 \end{bmatrix}$$

**SVD Theorem:** Any $m \times n$ real matrix $A$ can be factored as

$$A = Q_1 \Sigma Q_2^\top$$

where

$$\begin{aligned} Q_1 &= m \times m \text{ orthogonal matrix} \\ Q_2 &= n \times n \text{ orthogonal matrix} \\ \Sigma &= m \times n \text{ rectangular diagonal matrix} \end{aligned}$$

and $\operatorname{diag}(\Sigma) = [\sigma_1,\ \sigma_2,\ \cdots,\ \sigma_k]$ satisfies $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_k \geq 0$ where $k = \min(m,\ n)$. Moreover, the columns of $Q_1$ are eigenvectors of $AA^\top$, the columns of $Q_2$ are eigenvectors of $A^\top A$, and the $(\sigma_i)^2$ are eigenvalues of both $AA^\top$ and $A^\top A$.

**Remark:** The entries of $\operatorname{diag}(\Sigma)$ are called singular values of $A$.

**Proof of the theorem:** $A^\top A$ is $n \times n$, real, and symmetric. Hence, there exist orthonormal eigenvectors $\{v^1, \cdots, v^n\}$ such that $A^\top A v^j = \lambda_j v^j$. Without loss of generality, we can assume that

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n \geq 0$$

If not, we simply re-order the $v^i$'s to make it so.

For $\lambda_j > 0$, say $1 \leq j \leq r$, we define

$$\sigma_j = \sqrt{\lambda_j}$$

and

$$q^j = \frac{1}{\sigma_j} A v^j \in \mathbb{R}^m$$

<u>Claim:</u> $\left(q^i\right)^\top q^j = \delta_{ij} = \begin{cases} 1 & i = j \\ 0 & 1 \neq j \end{cases}$ for $1 \leq i,\ j \leq r$.

<u>Proof of Claim:</u>

$$\begin{aligned} \left(q^i\right)^\top q^j &= \frac{1}{\sigma_i} \frac{1}{\sigma_j} \left(v^i\right)^\top A^\top A v^j \\ &= \frac{\lambda_j}{\sigma_i \sigma_j} \left(v^i\right)^\top v^j \\ &= \begin{cases} \frac{\lambda_j}{(\sigma_i)^2} & i = j \\ 0 & i \neq j \end{cases} \\ &= \begin{cases} 1 & i = j \\ 0 & 1 \neq j \end{cases} \end{aligned}$$

<u>End of proof of Claim.</u>

If $r < m$, we can extend the $q^i$'s to an orthonormal basis for $\mathbb{R}^m$. Define

$$Q_1 = [q^1 \mid q^2 \mid \cdots \mid q^m]$$
$$Q_2 = [v^1 \mid v^2 \mid \cdots \mid v^n]$$

and define $\Sigma = m \times n$ by

$$\Sigma_{ij} = \begin{cases} \sigma_i \delta_{ij} & 1 \leq i,\ j \leq r \\ 0 & \text{otherwise} \end{cases}$$

Then, $\Sigma$ is rectangular diagonal with

$$\operatorname{diag}(\Sigma) = [\sigma_1,\ \sigma_2,\ \cdots,\ \sigma_r,\ 0,\ \cdots,\ 0]$$

To complete the proof of the theorem, it is enough to show that $Q_1^\top A Q_2 = \Sigma$. We note that the $ij$ element of this matrix is

$$(Q_1^\top A Q_2)_{ij} = q_i^\top A v^j$$

If $j > r$, then $Av^j = 0$, and thus $q_i^\top A v^j = 0$, as required. If $i > r$, then $q^i$ was selected to be orthogonal to

$$\{q^1,\ \cdots,\ q^r\} = \{\frac{1}{\sigma_1} A v^1,\ \frac{1}{\sigma_2} A v^2,\ \cdots,\ \frac{1}{\sigma_r} A v^r\}$$

and thus $\left(q^i\right)^\top A v^j = 0$.

Hence we now consider $1 \leq i,\ j \leq r$ and compute that

$$\begin{aligned} \left(Q_1^\top A Q_2\right)_{ij} &= \frac{1}{\sigma_i} \left(v^i\right)^\top A^\top A v^j \\ &= \frac{\lambda_j}{\sigma_i} \left(v^i\right)^\top v^j \\ &= \sigma_i \delta_{ij} \end{aligned}$$

as required. **End of Proof**.

# Singular Value Decomposition (SVD)

<u>Def.</u> An $m \times n$ matrix $\Sigma$ is <u>rectangular diagonal</u> if $\Sigma_{ij} = 0$ for $i \neq j$.

<u>Examples</u>

$$\begin{bmatrix} 3 & 0 & 0 & 0 \\ 0 & 4 & 0 & 0 \\ 0 & 0 & -1 & 0 \end{bmatrix} \quad \text{or} \quad \begin{bmatrix} 1 & 0 \\ 0 & -6 \\ 0 & 0 \end{bmatrix}$$

The diagonal of $\Sigma$ is

$$\text{diag}(\Sigma) = (\Sigma_{11}, \Sigma_{22}, \ldots, \Sigma_{kk})$$

where $k = \min(m, n)$.

<u>Example</u>

$\text{diag} = [3 \;\; 4 \;\; -1]$, or $[1, -6]$

<u>SVD Theorem</u> Any $m \times n$ real matrix $A$ can be factored as

$$A = Q_1 \Sigma Q_2^T$$

where $Q_1 = m \times m$ orthogonal matrix

$Q_2 = n \times n$ orthogonal matrix

$\Sigma = m \times n$ rectangular diagonal matrix, and

$\operatorname{diag} \Sigma = [\sigma_1, \sigma_2, \ldots, \sigma_k]$ satisfies

$$\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_k \geq 0$$

where $k = \min(m, n)$. Moreover, the ~~matrices~~ columns of $Q_1$ are e-vectors of $A \cdot A^T$, the columns of $Q_2$ are e-vectors of $A^T A$, and the $(\sigma_i)^2$ are e-values of both $A \cdot A^T$ and $A^T A$. $\boxtimes$

<u>Remark:</u> The entries of $\operatorname{diag}(\Sigma)$ are called <u>singular values</u> of $A$.

<u>Proof of the theorem</u>

$A^T A$ is $n \times n$ real and symmetric. Hence there exist orthonormal e-vectors $\{v^1, \ldots, v^n\}$ such that $A^T A v^j = \lambda_j v^j$. W.L.O.G., we can assume that

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n \geq 0$$

(if not, simply re-order the $v^i$'s).

For $\lambda_j > 0$, say $1 \leq j \leq r$, we define

$$\sigma_j = \sqrt{\lambda_j}$$

and

$$q^j = \frac{1}{\sigma_j} A v^j \in \mathbb{R}^m$$

Claim $(q^i)^T q^j = \delta_{ij} = \begin{cases} 1 & i=j \\ 0 & i \neq j \end{cases}$

for $1 \leq i,j \leq r$.

Pf.
$$(q^i)^T q^j = \frac{1}{\sigma_i} \frac{1}{\sigma_j} (v^i)^T A^T A v^j$$
$$= \frac{\lambda_j}{\sigma_i \sigma_j} (v^i)^T v^j$$
$$= \begin{cases} \frac{\lambda_i}{(\sigma_i)^2} & i=j \\ 0 & i \neq j \end{cases} = \begin{cases} 1 & i=j \\ 0 & i \neq j \end{cases}$$

$\square$

If $r < m$, we can extend the $q^i$'s to an orthonormal basis for $\mathbb{R}^m$.

Define

$$Q_1 = [q^1 | q^2 | \cdots | q^m]$$

$$Q_2 = [v^1 | v^2 | \cdots | v^n]$$

Define $\Sigma = m \times n$ by

$$\Sigma_{ij} = \begin{cases} \sigma_i \delta_{ij} & 1 \le i, j \le r \\ 0 & \text{otherwise.} \end{cases}$$

Then $\Sigma$ is rectangular diagonal with

$\operatorname{diag}(\Sigma) = [\sigma_1, \sigma_2, .., \sigma_r, 0 .., 0]$.

Proof of the Theorem

It is enough to show that

$$Q_1^T A Q_2 = \Sigma$$

$$(Q_1^T A Q_2)_{ij} = q_i^T A v^j$$

If $j > r$, then $Av^j = 0$, and thus

$$q_i^T A v^j = 0$$

If $i > r$, then $q^i$ selected to be orthogonal to $\{q^1, \dots, q^r\} = \{\frac{1}{\sigma_1} Av^1, \frac{1}{\sigma_2} Av^2, \dots, \frac{1}{\sigma_r} Av^r\}$, and thus $(q^i)^T A v^i = 0$.

Consider $1 \le i, j \le r$

$$(Q_1^T A Q_2)_{ij} = \frac{1}{\sigma_i} (v^i)^T A^T A v^j$$

$$= \frac{\lambda_i}{\sigma_i} v^{i\,T} v^j$$

$$= \sigma_i \, \delta_{ij}$$

as required. $\square$