Singular Value Decomposition (SVD)

- **Reading Assignments**

M. Petrou and P. Bosdogianni, *Image Processing: The Fundamentals*, John Wiley, 2000 (pp. 37-44 - examples of SVD, hard copy).

E. Trucco and A. Verri, *Introductory Techniques for 3D Computer Vision*, Prentice Hall (appendix 6, hard copy)

K. Kastleman, *Digital Image Processing*, Prentice Hall, (Appendix 3: Mathematical Background, hard copy).

F. Ham and I. Kostanic. *Principles of Neurocomputing for Science and Engineering*, Prentice Hall, (Appendix A: Mathematical Foundation for Neurocomputing, hard copy).

Singular Value Decomposition (SVD)

• Definition

- Any real mxn matrix A can be decomposed uniquely as

$$A = UDV^T$$

U is mxn and orthogonal (its columns are eigenvectors of AA^T)
$(AA^T = UDV^T VDU^T = UD^2U^T)$

V is nxn and orthogonal (its columns are eigenvectors of $A^T A$)
$(A^T A = VDU^T UDV^T = VD^2V^T)$

D is nxn diagonal (non-negative real values called *singular* values)

$D = diag(\sigma_1, \sigma_2, \ldots, \sigma_n)$ ordered so that $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n$
(if σ is a singular value of A, it's square is an eigenvalue of $A^T A$)

- If $U = (u_1\ u_2 \cdots u_n)$ and $V = (v_1\ v_2 \cdots v_n)$, then

$$A = \sum_{i=1}^{n} \sigma_i u_i v_i^T$$

(actually, the sum goes from 1 to r where r is the rank of A)

• An example

$$A = \begin{bmatrix} 1 & 2 & 1 \\ 2 & 3 & 2 \\ 1 & 2 & 1 \end{bmatrix}, \quad \text{then } AA^T = A^T A = \begin{bmatrix} 6 & 10 & 6 \\ 10 & 17 & 10 \\ 6 & 10 & 6 \end{bmatrix}$$

The eigenvalues of AA^T, $A^T A$ are:

$$\begin{bmatrix} \sigma_1 \\ \sigma_2 \\ \sigma_3 \end{bmatrix} = \begin{bmatrix} 28.86 \\ 0.14 \\ 0 \end{bmatrix}$$

The eigenvectors of AA^T, $A^T A$ are:

$$u_1 = v_1 = \begin{bmatrix} 0.454 \\ 0.766 \\ 0.454 \end{bmatrix}, u_2 = v_2 = \begin{bmatrix} 0.542 \\ -0.643 \\ 0.542 \end{bmatrix}, u_3 = v_3 = \begin{bmatrix} -0.707 \\ 0 \\ -0.707 \end{bmatrix}$$

The expansion of A is

$$A = \sum_{i=1}^{2} \sigma_i u_i v_i^T$$

Important: note that the second eigenvalue is much smaller than the first; if we neglect it from the above summation, we can represent A by introducing relatively small errors only:

$$A = \begin{bmatrix} 1.11 & 1.87 & 1.11 \\ 1.87 & 3.15 & 1.87 \\ 1.11 & 1.87 & 1.11 \end{bmatrix}$$

• Computing the rank using SVD

- The rank of a matrix is equal to the number of non-zero singular values.

• Computing the inverse of a matrix using SVD

- A square matrix A is nonsingular *iff* $\sigma_i \neq 0$ for all i

- If A is a nxn nonsingular matrix, then its inverse is given by

$$A^{-1} = VD^{-1}U^T$$

$$\text{where } D^{-1} = diag(\frac{1}{\sigma_1}, \frac{1}{\sigma_2}, \ldots, \frac{1}{\sigma_n})$$

- If A is singular or ill-conditioned, then we can use SVD to approximate its inverse by the following matrix:

$$A^{-1} = (UDV^T)^{-1} \approx VD_0^{-1}U^T$$

$$D_0^{-1} = \begin{cases} 1/\sigma_i & \text{if } \sigma_i > t \\ 0 & otherwise \end{cases}$$

(where t is a small threshold)

• The condition of a matrix

- Consider the system of linear equations

$$Ax = b$$

if small changes in b can lead to relatively large changes in the solution x, then we call A *ill-conditioned*.

- The ratio given below is related to the *condition* of A and measures the degree of singularity of A (the larger this value is, the closer A is to being singular)

$$\sigma_1/\sigma_n$$

(largest over smallest singular values)

• Least Squares Solutions of mxn Systems

- Consider the *over-determined* system of linear equations

$$Ax = b, \quad (A \text{ is } m\text{x}n \text{ with } m>n)$$

- Let r be the residual vector for some x:

$$r = Ax - b$$

- The vector x^* which yields the smallest possible residual is called a *least-squares* solution (it is an approximate solution).

$$||r|| = ||Ax^* - b|| \leq ||Ax - b|| \text{ for all } x \in R^n$$

- Although a least-squares solution always exist, it <u>might not</u> be unique !

- The least-squares solution x with the smallest norm $||x||$ <u>is unique</u> and it is given by:

$$A^T Ax = A^T b \text{ or } x = (A^T A)^{-1} A^T b = A^+ b$$

<u>Example:</u>

$$\begin{bmatrix} -11 & 2 \\ 2 & 3 \\ 2 & -1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 7 \\ 5 \end{bmatrix}$$

$$x = A^{+}b = \begin{bmatrix} -.148 & .180 & .246 \\ .164 & .189 & -.107 \end{bmatrix} \begin{bmatrix} 0 \\ 7 \\ 5 \end{bmatrix} = \begin{bmatrix} 2.492 \\ 0.787 \end{bmatrix}$$

• Computing A^{+} using SVD

- If $A^T A$ is ill-conditioned or singular, we can use SVD to obtain a least squares solution as follows:

$$x = A^{+}b \approx VD_0^{-1}U^T b$$

$$D_0^{-1} = \begin{cases} 1/\sigma_i & \text{if } \sigma_i > t \\ 0 & otherwise \end{cases}$$

(where t is a small threshold)

• Least Squares Solutions of nxn Systems

- If A is ill-conditioned or singular, SVD can give us a workable solution in this case too:

$$x = A^{-1}b \approx VD_0^{-1}U^T b$$

• Homogeneous Systems

- Suppose b=0, then the linear system is called homogeneous:

$$Ax = 0$$

(assume A is mxn and $A = UDV^T$)

- The minimum-norm solution in this case is x=0 (trivial solution).

- For homogeneous linear systems, the meaning of a least-squares solution is modified by imposing the constraint:

$$||x|| = 1$$

- This is a "constrained" optimization problem:

$$\min_{||x||=1} ||Ax||$$

- The minimum-norm solution for homogeneous systems is not always unique.

Special case: $rank(A) = n - 1$ ($m \geq n - 1$, $\sigma_n = 0$)

solution is $x = av_n$ (a is a constant)

General case: $rank(A) = n - k$ ($m \geq n - k$, $\sigma_{n-k+1} = \cdots = \sigma_n = 0$)

solution is $x = a_1 v_{n-k+1} + a_2 v_{n-k-1} + \cdots + a_k v_n$ (a_is is a constant)

with $a_1^2 + a_2^2 + \cdots + a_k^2 = 1$