The Singular Value Decomposition

From www.norsemathology.org

Contents

The Action of a Symmetric Matrix

First of all, it is useful to understand the action of a symmetric matrix in a linear transformation, $T : x \longrightarrow Ax$. We know that a symmetric matrix can be represented as

$$A = PDP^T$$

where P is an orthogonal matrix (that is, a matrix whose columns are orthogonal to each other, and of unit length). The columns of P are (orthogonal) eigenvectors of A, and D is the diagonal matrix of eigenvalues. Thus, when a vector x is multiplied by A,

$$Ax = PDP^Tx \qquad P^T\left(P\right) = \left(\underline{I}\right)$$

the geometric result can be visualized using the unit ball (of unit radius) in R^n. It is transformed into an ellipsoid, with the eigenvectors representing the axes of the ellipsoid, and the eigenvalues providing the scaling of each axis of the ellipsoid. We can also think of how this transformation comes about as a succession of three steps:

- Multiplication by P^T, by the transpose of the orthogonal matrix P, corresponds to a <u>rotation</u> of the space R^n. It rotates each of the eigenvectors p_i into the standard vector e_i.

- Multiplication by D represents a scaling of each of the now "standardized" eigenvectors: each is stretched by the appropriate factor, the eigenvalue;
- Multiplication of the result by P rotates the scaled eigenvectors back into their original positions.

In conclusion, you can think of the "action" of a symmetric matrix as a product of these three simple actions on R^n: a rotation, a scaling, and a rotation.

The SVD Theorem

It turns out that this aspect of symmetric matrices is true in general: <u>every</u> matrix can be thought of as the product of an orthogonal matrix, a diagonal matrix, and the transpose of an orthogonal matrix.

Every matrix A is related to two important symmetric matrices: $A^T A$ and $A A^T$. Since each of these two matrices is symmetric, each can be represented as a product

$$\begin{cases} A A^T = U D_1 U^T \\ A^T A = V D_2 V^T \end{cases}$$

Furthermore, the diagonal matrices D_1 and D_2 have positive entries on the diagonal. We can see that from the spectral decomposition of $A A^T$ and $A^T A$. For example,

$$A A^T = \lambda_1 u_1 u_1^T + \ldots + \lambda_m u_m u_m^T$$

Hence,

$$u_1^T A A^T u_1 = \lambda_1 u_1^T u_1 u_1^T u_1 + 0 = \lambda_1 (u_1 \cdot u_1)(u_1 \cdot u_1) = \lambda_1,$$

but since

$$u_1^T A A^T u_1 = A^T u_1 \cdot A^T u_1 = ||A^T u_1||^2 \geq 0 \text{, we know that } \lambda_1 \geq 0.$$

Theorem: Let A be an *mxn* matrix with real components. Then $A = U \Sigma V^T$ where U and V are as defined above, and Σ is the matrix with positive (≥ 0) entries such that $\Sigma \Sigma^T = D_1$ and $\Sigma^T \Sigma = D_2$.
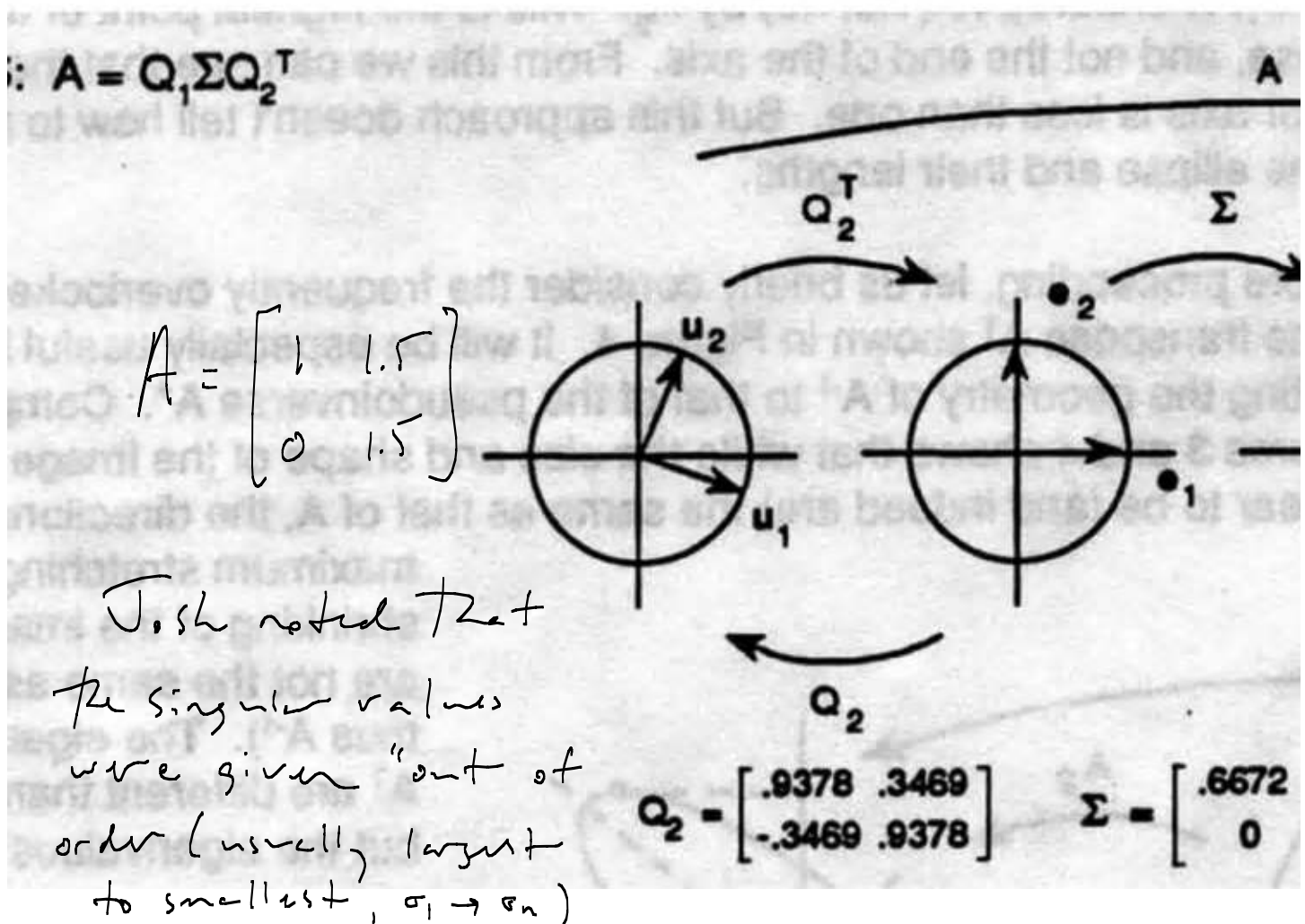
$$\begin{bmatrix} A \end{bmatrix} = \begin{bmatrix} U \end{bmatrix} \begin{bmatrix} \sigma_1 & & 0 \\ & \ddots & \\ 0 & & \sigma_n \\ \hline & 0 & \end{bmatrix} \begin{bmatrix} V^T \end{bmatrix} = \begin{bmatrix} U(:,1:n) \end{bmatrix} \begin{bmatrix} \sigma_1 & & 0 \\ & \ddots & \\ 0 & & \sigma_n \end{bmatrix} \begin{bmatrix} V^T \end{bmatrix}$$

We can easily show that, with $A = U\Sigma V^T$, the formulas for AA^T and A^TA work out:

$$\begin{cases} AA^T = U\Sigma V^T V \Sigma^T U^T = U\Sigma\Sigma^T U^T = UD_1 U^T \\ A^T A = V\Sigma^T U^T U\Sigma V^T = V\Sigma^T\Sigma V^T = VD_2 V^T \end{cases}$$

Notice that the orthogonal diagonalization of a symmetric matrix <u>is</u> the singular value decomposition in that case.

What follows is an illustration of the SVD, as diagrammed by Cliff Long and Tom Hern in the case of R^2. In this example, the notation is $U = Q_1$ and $V = Q_2$.

: $A = Q_1\Sigma Q_2^T$

$A = \begin{bmatrix} 1 & 1.5 \\ 0 & 1.5 \end{bmatrix}$

Josh noted that the singular values were given "out of order (usually largest to smallest, $\sigma_1 \to \sigma_n$)

$Q_2 = \begin{bmatrix} .9378 & .3469 \\ -.3469 & .9378 \end{bmatrix}$ $\Sigma = \begin{bmatrix} .6672 \\ 0 \end{bmatrix}$

You see that the action of a general matrix A can also be viewed as a rotation, followed by a scaling, followed by a rotation. If A is not square, or not full rank, then there will be a null space, and some of the dimensions will be annihilated.

The SVD for image processing, image analysis, and

statistical analysis

If A is rank n, then A can be written as a spectral decomposition, too, in terms of the eigenvectors and the eigenvalues of $A^T A$ and AA^T. The eigenvectors of $A^T A$ and AA^T are called the **singular vectors** of A:

$$A = \sqrt{\lambda_1} u_1 v_1^T + \ldots + \sqrt{\lambda_n} u_n v_n^T$$

$$A = \sigma_1 u_1 v_1^T + \ldots + \sigma_n u_n v_n^T$$

and where, by convention, $\lambda_i \geq \lambda_j$ for $i \leq j$. The values $\sigma_i = \sqrt{\lambda_i}$ are called the **singular values** of the matrix. This is crucial for image processing: often an image is not full rank, so that there are few products; furthermore, it may be that an image contains noise, and the noise tends to be high frequency and of little total weight. It may be that it "hangs out" on the last few singular vector pairs. We simply drop those from the recomposition of the matrix A, and we will have **de-noised** the image.

On a similar note, the information contained in the last singular vector pairs may not be very important to the recomposition of the image, so that we can drop them without much loss of information. This represents **compression** of the image.

The SVD summarizes the basic spaces related to a matrix A

- The Col(A) is given by the columns of U corresponding to non-zero singular values.
- The Row(A) is given by the columns of V corresponding to non-zero singular values.
- The Nul(A) is given by the columns of V corresponding to zero singular values; and of course
- The Nul(A^T) is given by the columns of U corresponding to zero singular values.

Applications

- We can use the SVD to interpolate a matrix (http://www.nku.edu/~longa/classes/mat225/days/dad/svd3.doc) . See Long, A. and C. Long. *Surface Approximation and Interpolation Via Matrix SVD*. The College Mathematics Journal, Vol. 32, #1, January, 2001:20-25.
- When we can't invert a matrix, we can define a "pseudo-inverse" via the SVD (http://www.nku.edu/~longa/classes/mat225/days/dad/pseudo.jpg) . This is precisely what we do when we solve the linear regression/least squares problem in general: we can't solve $Ax = b$ because the system is inconsistent; but we can find a least-squares solution, by inverting (when possible). When it's not possible, we use a pseudo-inverse solution.

Retrieved from "http://norsemathology.org/wiki/index.php?title=The_Singular_Value_Decomposition"

- This page was last modified 15:22, 28 April 2008.