MAT225 Section Summary: 7.4

The Singular Value Decomposition (SVD):

Fundamental Theorem of Linear Algebra

Summary

That's right: The Fundamental Theorem of Linear Algebra. The SVD ties it all together. Rather than focus on the technicalities, I want to focus on the "bang". If you can understand the Singular Value Decomposition, then you understand this course. If you are weak in any part, then you will not really understand this theorem. Understanding this section is the best preparation for the final exam.

The Singular Value Decomposition: Let A be an m x n matrix with **rank** r. There exists

1. an m x n matrix
$$\Sigma = \begin{bmatrix} D_{r \text{ x } r} & 0 \\ 0 & 0 \end{bmatrix}_{m \text{ x } n}$$
for which D is diagonal, with positive entries (the **singular values**) $\sigma_2 \geq \sigma_2 \geq \ldots \geq \sigma_r > 0$, and

2. orthogonal matrices $U_{m \text{ x } m}$ and $V_{n \text{ x } n}$

such that
$$A = U\Sigma V^T$$

The 0 matrices are included simply to pad D (if necessary) to make the dimensions right. Here is Σ with the dimensions indicated explicitly:

$$\Sigma = \begin{bmatrix} D_{r \text{ x } r} & 0_{r \text{ x } (n-r)} \\ 0_{(m-r) \text{ x } r} & 0_{(m-r) \text{ x } (n-r)} \end{bmatrix}_{m \text{ x } n}$$

For example, if A happens to be invertible, then there are no zero matrices, and $\Sigma = D$.

Now your first impulse might be to say "so what?" (But don't say it in my hearing!) Understanding this is the true key to understanding A either

1. as an image made up as rank-one subimages, or
2. as a linear transformation $\mathbf{x} \to A\mathbf{x}$ taking a ball to an ellipsoid

(my two favorite applications of matrices).

You might also want to know what this has to do with all the **symmetric matrices** we've been looking at: the connection is crucial. One way that the SVD arises is by considering the problem of solving a constrained optimization problem:

What is the maximum value of $\|A\mathbf{x}\|$ given that $\|\mathbf{x}\| = 1$?

Solving this is equivalent to solving the problem

What is the maximum value of $\|A\mathbf{x}\|^2$ given that $\|\mathbf{x}\| = 1$?

But this is equivalent to solving

What is the maximum value of $\mathbf{x}^T(A^TA)\mathbf{x}$ given that $\|\mathbf{x}\| = 1$?

So it's equivalent to solving a problem about **constrained optimization of quadratic forms**....

Now, to make my life easy, I'm going to think of $m \geq n$ (A is rectangular, and its "height" is greater than or equal to its "width").

The matrix A^TA is **positive semi-definite**: that is, it has positive (≥ 0) **eigenvalues**, which we can calculate as

$$A^TA_{n \text{ x } n} = V_{n \text{ x } n}\Lambda_{n \text{ x } n}V^T_{n \text{ x } n}$$

Now, here's a curious fact, which I'm going to gloss over entirely (in order to get to the "bang!"): AA^T (which is also positive semi-definite) can be **orthogonally diagonalized** as

$$AA^T = U_{m \text{ x } m}\begin{bmatrix} \Lambda & 0 \\ 0 & 0 \end{bmatrix}_{m \text{ x } m} U^T_{m \text{ x } m}$$

(Same Λ! - same eigenvalues). Again, the zero matrices are only included to (possibly) make the dimensions work out. If $r = m$, then they disappear. Now for the kicker:

1. the **singular values of** A, σ_1, σ_2, ..., σ_r are the positive square roots of the elements of Λ! I.e., $D = \sqrt{\Lambda}$.

2. The singular vectors are the eigenvectors of $A^T A$ and AA^T.

It's easy to check that the SVD formula recreates $A^T A$ and AA^T. Give it a try!

Now the best way to think about A from the standpoint of the image problem is to throw out the irrelevant stuff, and write

$$A = U_{m \text{ x } r} D V^T_{r \text{ x } n}$$

where we've thrown out the eigenvectors of $A^T A$ and the eigenvectors of AA^T corresponding to the zero weights of Σ. This might be called the **reduced** SVD of A. Hence we can write a decomposition analogous to the spectral decomposition of a symmetric matrix for A, as

$$A = \sigma_1 \mathbf{u}_1 \mathbf{v}_1^T + \sigma_2 \mathbf{u}_2 \mathbf{v}_2^T + \ldots + \sigma_r \mathbf{u}_r \mathbf{v}_r^T$$

This is the matrix A composed of a sum of rank-one outer product matrices, weighted from most important to least (by the size of σ_i). An example of this can be found at http://www.nku.edu/∼longa/classes/mat225/days/dad/Abe.jpg

You know that the rank of A and A^T is the same. The **row space** of one and the **column space** of the other are equivalent. The SVD makes that clear.

From the standpoint of the transformation $\mathbf{x} \rightarrow A\mathbf{x}$, taking a ball in $\mathbb{R}^n$to an ellipsoid in $\mathbb{R}^m$, the better way to think of A is as

$$A = U \Sigma V^T$$

where by $A\mathbf{x}$ we understand the succession of transformations $A\mathbf{x} = U \Sigma V^T \mathbf{x}$ as follows:

1. $$V^T \mathbf{x}$$

   (the expression of $\mathbf{x}$ in the **basis** of V, as **projections** onto the basis vectors). This is effectively a rotation/reflection of the unit (radius) ball in $\mathbb{R}^n$into position for easy scaling.

2.
$$\Sigma V^T \mathbf{x}$$
represents the scaling of the vector along each of the principal axes (the conversion of a ball into an ellipsoid), including possible squashing of some of the dimensions corresponding to the **null-space of** A. The resulting vectors are in $\mathbb{R}^m$.

3. Then
$$U \Sigma V^T \mathbf{x}$$
represents the rotation/reflection of the resulting ellipsoid in $\mathbb{R}^m$so that the result is oriented as it should be, since $A\mathbf{x}$ is not necessarily alligned well with the **standard basis**. This is the image of the vector $\mathbf{x}$ in the column space of A, as a linear combination of the column space basis U of A.

Here's a nice picture due to Cliff Long and Tom Hern that captures these three steps, at

http://www.nku.edu/∼longa/classes/mat225/days/dad/SVD.jpg

The rank of a matrix is equal to the number of non-zero singular values. The **condition number** (or at least the most common definition of it) is given as the ratio of the largest to smallest singular values (infinity if rank $r \neq n$, since it's as if the matrix has a singular value of 0).

It is all really too marvellous for words. We'll need to look at some pictures, and a few example problems.

Example: #15, p. 481

Note: in order to calculate the SVD of a matrix, you need only find D and *either* U or V – you don't need both. That's because if $A = UDV^T$, then if we know D and V, then $U = AVD^{-1}$

This leads to the idea of a **pseudo-inverse** of a matrix:
$$A^+ = VD^{-1}U^T$$

This is the closest thing to an inverse that a general matrix A has!

There's a nice picture due to Cliff Long and Tom Hern at

http://www.nku.edu/~longa/classes/mat225/days/dad/pseudo.jpg

Example: #9, p. 481