Week 5: Wednesday, Feb 29

Of cabbages and kings

The past three weeks have covered quite a bit of ground. We've looked at linear systems and least squares problems, and we've discussed Gaussian elimination, QR decompositions, and singular value decompositions. Rather than doing an overly hurried introduction to iterative methods for solving linear systems, I'd like to go back and show the surprisingly versatile role that the SVD can play in thinking about all of these problems.

Geometry of the SVD

How should we understand the singular value decomposition? We've already described the basic algebraic picture:

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

where U and V are orthonormal matrices and Σ is diagonal. But what about the geometric picture?

Let's start by going back to something we glossed over earlier in the semester: the characterization of the matrix 2-norm. By definition, we have

$$\|A\|_2 = \max_{x \neq 0} \frac{\|Ax\|_2}{\|x\|_2}$$

This is equivalent to

$$\|A\|_2^2 = \max_{x \neq 0} \frac{\|Ax\|_2^2}{\|x\|_2^2} = \max_{x \neq 0} \frac{x^T A^T A x}{x^T x}.$$

The quotient $\phi(x) = (x^T A^T A x)/(x^T x)$ is differentiable, and the critical points satisfy

$$0 = \nabla \phi(x) = \frac{2}{x^T x} \left(A^T A x - \phi(x) x \right)$$

That is, the critical points of ϕ – including the value of x that maximizes ϕ – are eigenvectors of A. The corresponding eigenvalues are values of $\phi(x)$. Hence, the largest eigenvalue of $A^T A$ is $\sigma_1^2 = \|A\|_2^2$. The corresponding

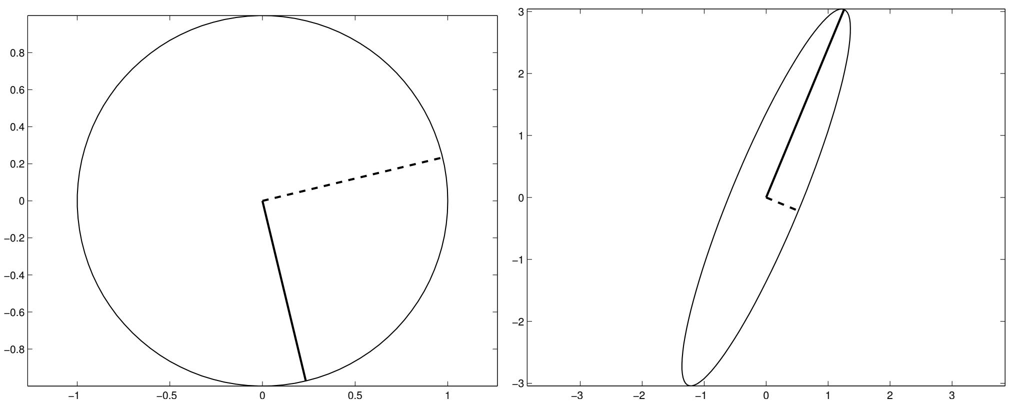

Figure 1: Graphical depiction of an SVD of $A \in \mathbb{R}^{2\times 2}$. The matrix A maps the unit circle (left) to an oval (right); the vectors v_1 (solid, left) and v_2 (dashed, left) are mapped to the major axis $\sigma_1 u_1$ (solid, right) and the minor axis $\sigma_2 u_2$ (dashed, right) for the oval.

eigenvector v_1 is the right singular vector corresponding to the eigenvalue σ_1^2; and $Av_1 = \sigma_1 u_1$ gives the first singular value.

What does this really say? It says that v_1 is the vector that is stretched the most by multiplication by A, and σ_1 is the amount of stretching. More generally, we can *completely* characterize A by an orthonormal basis of right singular vectors that are each transformed in the same special way: they get scaled, then rotated or reflected in a way that preserves lengths. Viewed differently, the matrix A maps vectors on the unit sphere into an ovoid shape, and the singular values are the lengths of the axes. In Figure 1, we show this for a particular example, the matrix

$$A = \begin{bmatrix} 0.8 & -1.1 \\ 0.5 & -3.0 \end{bmatrix}.$$

Conditioning and the distance to singularity

We have already seen that the condition number for linear equation solving is

$$\kappa(A) = \|A\|\|A^{-1}\|$$

When the norm in question is the operator two norm, we have that $\|A\| = \sigma_1$ and $\|A^{-1}\| = \sigma_n^{-1}$, so

$$\kappa(A) = \frac{\sigma_1}{\sigma_n}$$

That is, $\kappa(A)$ is the ratio between the largest and the smallest amounts by which a vector can be stretched through multiplication by A.

There is another way to interpret this, too. If $A = U\Sigma V^T$ is a square matrix, then the smallest E (in the two-norm) such that $A - E$ is *exactly* singular is $A - \sigma_n u_n v_n^T$. Thus,

$$\kappa(A)^{-1} = \frac{\|E\|}{\|A\|}$$

is the *relative distance to singularity* for the matrix A. So a matrix is ill-conditioned exactly when a relatively small perturbation would make it exactly singular.

For least squares problems, we still write

$$\kappa(A) = \frac{\sigma_1}{\sigma_n},$$

and we can still interpret $\kappa(A)$ as the ratio of the largest to the smallest amount that multiplication by A can stretch a vector. We can also still interpret $\kappa(A)$ in terms of the distance to singularity – or, at least, the distance to rank deficiency. Of course, the actual sensitivity of least squares problems to perturbation depends on the angle between the right hand side vector b and the range of A, but the basic intuition that big condition numbers means problems can be very near singular – very nearly ill-posed – tells us the types of situations that can lead us into trouble.

Orthogonal Procrustes

The SVD can provide surprising insights in settings other than standard least squares and linear systems problems. Let's consider one interesting one that comes up when doing things like trying to align 3D models with each other.

Suppose we are given two sets of coordinates for m points in n-dimensional space, arranged into rows of $A \in \mathbb{R}^{m\times n}$ and $B \in \mathbb{R}^{m\times n}$. Let's also suppose the matrices are (approximately) related by a rigid motion that leaves the origin fixed. How can we recover the transformation? That is, we want an orthogonal matrix W that minimizes $\|AW - B\|_F^2$. This is sometimes called an *orthogonal Procrustes problem*, named in honor of the legendary Greek king Procrustes, who had a bed on which he would either stretch guests or cut off their legs in order to make them fit perfectly.

We can write $\|AW - B\|_F^2$ as

$$\|AW - B\|_F^2 = (\|A\|_F^2 + \|B\|_F^2) - \operatorname{tr}(W^T A^T B),$$

so minimizing the squared residual is equivalent to maximizing $\operatorname{tr}(W^T A^T B)$. Note that if $A^T B = U\Sigma V^T$, then

$$\operatorname{tr}(W^T A^T B) = \operatorname{tr}(W^T U\Sigma V^T) = \operatorname{tr}(VWU^T \Sigma) = \operatorname{tr}(Z\Sigma),$$

where $Z = VWU^T$ is orthogonal. Now, note that

$$\operatorname{tr}(Z\Sigma) = \operatorname{tr}(\Sigma Z) = \sum_i \sigma_i z_{ii}$$

is maximal over all orthogonal matrices when $z_{ii} = 1$ for each i. Therefore, the trace is maximized when $Z = I$, corresponding to $W = UV^T$.

Problems to Ponder

1. Suppose $A \in \mathbb{R}^{n\times n}$ is invertible and $A = U\Sigma V^T$ is given. How could we use this decomposition to solve $Ax = b$ in $O(n^2)$ additional work?
2. What are the singular values of A^{-1} in terms of the singular values of A?
3. Suppose $A = QR$. Show $\kappa_2(A) = \kappa_2(R)$.
4. Suppose that $A^T A = R^T R$, where R is an upper triangular Cholesky factor. Show that AR^{-1} is a matrix with orthonormal columns.
5. Show that if V and W are orthogonal matrices with appropriate dimensions, then $\|VAW\|_F = \|A\|_F$.
6. Show that if $X, Y \in \mathbb{R}^{m\times n}$ and $\operatorname{tr}(X^T Y) = 0$ then $\|X\|_F^2 + \|Y\|_F^2 = \|X + Y\|_F^2$.
7. Why do the diagonal entries of an orthogonal matrix have to lie between -1 and 1? Why must an orthogonal matrix with all ones on the diagonal be an identity matrix?