Jim Lambers
CME 335
Spring Quarter 2010-11
Lecture 6 Notes

The SVD Algorithm

Let A be an $m \times n$ matrix. The *Singular Value Decomposition* (SVD) of A,

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

where U is $m \times m$ and orthogonal, V is $n \times n$ and orthogonal, and Σ is an $m \times n$ diagonal matrix with nonnegative diagonal entries

$$\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_p, \quad p = \min\{m, n\},$$

known as the *singular values* of A, is an extremely useful decomposition that yields much information about A, including its range, null space, rank, and 2-norm condition number. We now discuss a practical algorithm for computing the SVD of A, due to Golub and Kahan.

Let U and V have column partitions

$$U = \begin{bmatrix} \mathbf{u}_1 & \cdots & \mathbf{u}_m \end{bmatrix}, \quad V = \begin{bmatrix} \mathbf{v}_1 & \cdots & \mathbf{v}_n \end{bmatrix}.$$

From the relations

$$A\mathbf{v}_j = \sigma_j \mathbf{u}_j, \quad A^T\mathbf{u}_j = \sigma_j \mathbf{v}_j, \quad j = 1, \ldots, p,$$

it follows that

$$A^T A\mathbf{v}_j = \sigma_j^2 \mathbf{v}_j.$$

That is, the squares of the singular values are the eigenvalues of A^TA, which is a symmetric matrix.

It follows that one approach to computing the SVD of A is to apply the symmetric QR algorithm to A^TA to obtain a decomposition $A^TA = V\Sigma^T\Sigma V^T$. Then, the relations $A\mathbf{v}_j = \sigma_j\mathbf{u}_j$, $j = 1, \ldots, p$, can be used in conjunction with the QR factorization with column pivoting to obtain U. However, this approach is not the most practical, because of the expense and loss of information incurred from computing A^TA.

Instead, we can *implicitly* apply the symmetric QR algorithm to A^TA. As the first step of the symmetric QR algorithm is to use Householder reflections to reduce the matrix to tridiagonal form, we can use Householder reflections to instead reduce A to *upper bidiagonal form*

$$U_1^T A V_1 = B = \begin{bmatrix} d_1 & f_1 & & & \\ & d_2 & f_2 & & \\ & & \ddots & \ddots & \\ & & & d_{n-1} & f_{n-1} \\ & & & & d_n \end{bmatrix}.$$

It follows that $T = B^T B$ is symmetric and tridiagonal.

We could then apply the symmetric QR algorithm directly to T, but, again, to avoid the loss of information from computing T explicitly, we implicitly apply the QR algorithm to T by performing the following steps during each iteration:

1. Determine the first Givens row rotation G_1^T that *would* be applied to $T - \mu I$, where μ is the Wilkinson shift from the symmetric QR algorithm. This requires only computing the first column of T, which has only two nonzero entries $t_{11} = d_1^2$ and $t_{21} = d_1 f_1$.

2. Apply G_1 as a *column* rotation to columns 1 and 2 of B to obtain $B_1 = BG_1$. This introduces an unwanted nonzero in the $(2, 1)$ entry.

3. Apply a Givens row rotation H_1 to rows 1 and 2 to zero the $(2, 1)$ entry of B_1, which yields $B_2 = H_1^T BG_1$. Then, B_2 has an unwanted nonzero in the $(1, 3)$ entry.

4. Apply a Givens column rotation G_2 to columns 2 and 3 of B_2, which yields $B_3 = H_1^T BG_1G_2$. This introduces an unwanted zero in the $(3, 2)$ entry.

5. Continue applying alternating row and column rotations to "chase" the unwanted nonzero entry down the diagonal of B, until finally B is restored to upper bidiagonal form.

By the Implicit Q Theorem, since G_1 is the Givens rotation that would be applied to the first column of T, the column rotations that help restore upper bidiagonal form are essentially equal to those that would be applied to T if the symmetric QR algorithm was being applied to T directly. Therefore, the symmetric QR algorithm is being correctly applied, implicitly, to B.

To detect decoupling, we note that if any superdiagonal entry f_i is small enough to be "declared" equal to zero, then decoupling has been achieved, because the ith subdiagonal entry of T is equal to $d_i f_i$, and therefore the ith subdiagonal entry of T must be zero as well. If a diagonal entry d_i becomes zero, then decoupling can be achieved as follows:

- If $d_i = 0$, for $i < n$, then Givens row rotations applied to rows i and k, for $k = i + 1, \ldots, n$, can be used to zero the entire ith row. The SVD algorithm can then be applied separately to $B_{1:i,1:i}$ and $B_{i+1:n,i+1:n}$.

- If $d_n = 0$, then Givens column rotations applied to columns i and n, for $i = n-1, n-2, \ldots, 1$, can be used to zero the entire nth column. The SVD algorithm can then be applied to $B_{1:n-1,1:n-1}$.

In summary, if *any* diagonal or superdiagonal entry of B becomes zero, then the tridiagonal matrix $T = B^T B$ is no longer unreduced and deflation is possible.

Eventually, sufficient decoupling is achieved so that B is reduced to a diagonal matrix Σ. All Householder reflections that have pre-multiplied A, and all row rotations that have been applied to B, can be accumulated to obtain U, and all Householder reflections that have post-multiplied A, and all column rotations that have been applied to B, can be accumulated to obtain V.