Schur factorization

Kyle Miller

31 March 2016

For the 3:30 discussion, I only showed that the A' matrix had the same eigenvalues as A (less λ_1), but I didn't actually show that the multiplicity itself carried over, which was a large mistake on my part. The good news is that there is an easier way which both shows A' has the same eigenvalues and shows they occur with the same multiplicity. This version will be given below. I am giving the whole proof in full to make sure no other mistakes remain.

Theorem 1. *Let A be an $n \times n$ matrix with n real eigenvalues (with multiplicity). Then A can be written as $A = URU^T$ with U orthogonal and R upper triangular $n \times n$ matrices.*

Proof. We prove this by induction on n.

- If $n = 1$, then $A = \begin{pmatrix} a \end{pmatrix}$ for some a, and $A = \begin{pmatrix} 1 \end{pmatrix} \begin{pmatrix} a \end{pmatrix} \begin{pmatrix} 1 \end{pmatrix}^T$.
- If $n > 1$ and Schur factorization works for matrices of size $(n-1) \times (n-1)$, then:
  - Let $\lambda_1, \dots, \lambda_n$ be the real eigenvalues (with multiplicity), which we know exist by hypothesis.
  - Let u_1 be an eigenvector of unit length with eigenvalue λ_1. There is one: take any eigenvector associated with λ_1 (i.e., any vector in the nontrivial $\mathrm{Nul}(A - \lambda_1 I_n)$) and normalize it.
  - Let $u_2, \dots, u_n \in \mathbb{R}^n$ be vectors so that $\{u_1, \dots, u_n\}$ is an orthonormal basis. One way to do this:
    * Create a basis $\{u_1, v_2, \dots, v_n\}$ for $\mathbb{R}^n$ by iteratively taking a vector v_{k+1} not in the span of $\{u_1, v_1, \dots, v_k\}$ so far, and add it to the set. The resulting set $\{u_1, v_1, \dots, v_{k+1}\}$ is independent by construction. This process must terminate with n vectors because $\mathbb{R}^n$ is n-dimensional.
    * Orthonormalize by Gram-Schmidt. Since u_1 is unit-length, u_1 stays the same in the resulting orthonormal basis.
  - Let $V = \begin{pmatrix} v_1 & \cdots & v_n \end{pmatrix}$, which is an orthogonal matrix.
  - The matrix of A relative to this basis has $\lambda_1 e_1$ as its first column.
    * The matrix for A relative to this basis is $V^{-1}AV$, which, since V is orthogonal, is $V^T AV$.

  * The first column of a matrix B is Be_1, since the standard matrix of a linear transformation T is $\begin{pmatrix} T(e_1) & \cdots & T(e_n) \end{pmatrix}$, and the standard matrix of a matrix is the matrix.
  * So, we calculate $V^T A V e_1$.
  * Since Ve_1 is the first column of V, we now have $V^T A u_1$.
  * Since u_1 is an eigenvector of A, we now have $V^T \lambda_1 v_1 = \lambda_1 V^T v_1$.
  * Since $V^T = V^{-1}$ and $v_1 = Ve_1$, then $V^T v_1 = e_1$, so we now have $\lambda_1 e_1$.
  * Thus, $V^T A V e_1 = \lambda_1 e_1$.
- Then $V^T A V$ is of the form

$$V^T A V = \left(\begin{array}{c|ccc} \lambda_1 & * & \cdots & * \\ \hline 0 & & & \\ \vdots & & A' & \\ 0 & & & \end{array}\right)$$

  where A' is some $(n-1) \times (n-1)$ matrix.
- The eigenvalues of A', with multiplicity, are $\lambda_2, \ldots, \lambda_n$.
  * The characteristic polynomial of A is $|A - \lambda I_n|$.
  * Since $V^T V = 1$, $|V^T||V| = 1$, so the characteristic polynomial equals $|V^T||A - \lambda I_n||V| = |V^T(A - \lambda I_n)V| = |V^T A V - \lambda I_n|$.
  * Using the form we calculated for $V^T A V$, this becomes

$$\left|\begin{array}{c|ccc} \lambda_1 - \lambda & * & \cdots & * \\ \hline 0 & & & \\ \vdots & & A' - \lambda I_{n-1} & \\ 0 & & & \end{array}\right|$$

  * Expanding along the first column, this gives $(\lambda_1 - \lambda)|A' - \lambda I_{n-1}|$.
  * Then $|A - \lambda I_n| = (\lambda_1 - \lambda)|A' - \lambda I_{n-1}|$, so we have related the characteristic polynomials of A and A'.
  * Thus, since λ_1 is a root of the characteristic polynomial for A, the rest of the roots $\lambda_2, \ldots, \lambda_n$ must be roots of the characteristic polynomial for A' with the same multiplicities.
- Then, since we are assuming Schur factorization works for $(n-1) \times (n-1)$ matrices, and since A' is such with $n-1$ real eigenvalues, with multiplicity, then $A' = W' R' (W')^T$ for some orthogonal W' and upper triangular R' $(n-1) \times (n-1)$ matrices.
- The matrix

$$\left(\begin{array}{ccc} 0 & \cdots & 0 \\ \hline & W' & \end{array}\right)$$

  obtained by inserting a 0 before each column of W' is still an orthogonal matrix, since the columns are still orthogonal and unit length.

- There are $n-1$ orthonormal vectors in this matrix, which we label by $w_2, \dots, w_n$. Then $\{e_1, w_2, \dots, w_n\}$ is an orthonormal basis of $\mathbb{R}^n$, since $e_1 \cdot w_i = 0$ for $2 \le i \le n$. Let W be the orthogonal matrix with this basis as columns.
- Claim:

$$W^T \left(\begin{array}{c|ccc} \lambda_1 & * & \cdots & * \\ \hline 0 & & & \\ \vdots & & A' & \\ 0 & & & \end{array}\right) W$$

is upper triangular. Let this matrix be R.

  * The first column of R is $\lambda_1 e_1$. We compute this via Re_1.
    · Since $We_1 = e_1$, then we need to multiply the inner matrix by e_1, which is $\lambda_1 e_1$.
    · Then, multiplying by W^T, we have $\lambda_1 W^T e_1 = \lambda_1 e_1$.
    · Thus, the first column of R is $\lambda_1 e_1$, which so far satisfies R being upper triangular.
  * For the remaining columns R has R' in place of A', with the $*$ entries replaced by some other scalars. Let $2 \le i \le n$.
    · Since $We_i = w_i$, we need to multiply the inner matrix by w_i, which, since the first component of w_i is 0, is (with w_i' being the vector consisting of the entries of w_i after the 0)

$$\left(\begin{array}{c} \text{stuff} \cdot w_i' \\ \hline \\ A' w_i' \\ \\ \end{array}\right) = \left(\begin{array}{c} \text{stuff} \cdot w_i' \\ \hline \\ W' R' (W')^T w_i' \\ \\ \end{array}\right)$$

    where $W'R'(W')^T w_i' = W'R'e_{i-1}$. Then, the above vector is $(\text{stuff} \cdot w_i')e_1 + Wr$, where r is $R'e_{i-1}$ (column $i-1$ of R') with a zero entry inserted at the beginning.
    · When this vector is multiplied by $W^T = W^{-1}$, we then have $W^{-1}((\text{stuff} \cdot w_i')e_1 + Wr) = (\text{stuff} \cdot w_i')W^{-1}e_1 + W^{-1}Wr = (\text{stuff} \cdot w_i')e_1 + r$.
    · Thus, the column is a column of R' with some scalar inserted before the first entry.
  * Together, these imply R is of the form

$$\left(\begin{array}{c|ccc} \lambda_1 & * & \cdots & * \\ \hline 0 & & & \\ \vdots & & R' & \\ 0 & & & \end{array}\right),$$

  and this is upper triangular since R' is.
- So, $W^T V^T A V W = R$. Let $U = VW$, so then $A = URU^T$. The product of orthogonal matrices is orthogonal, so U is orthogonal, and R is upper triangular. Therefore, this is a Schur factorization for A.

- Therefore, the factorization can be done for all $n \geq 1$.

$\square$