Numerical Linear Algebra Review

Stephen Bates

July 2, 2019

Runtimes cheat sheet

Let A be a $d_1 \times d_2$ matrix, let B be a $d_2 \times d_3$ matrix, and let v_1 and v_2 be vectors of dimension d_1 and d_2, respectively.

- Computing AB takes $O(d_1 d_2 d_3)$ operations.
- Computing the QR factorization of A takes $O(d_1^2 d_2)$ operations. This is only defined for $d_1 > d_2$.
- Computing the SVD of A takes $O(d_1 d_2 \min(d_1, d_2))$ operations.
- For A symmetric, computing the eigendecomposition takes $O(d_1^3)$ operations.
- For A square, computing A^{-1} takes $O(d_1^3)$ operations.
- For A square, solving $Ax = v_1$ for x takes $O(d_1^3)$ operations.

Matrix solves

A common task in numerical linear algebra is to find a solution x to the equation $Ax = v_1$, for a square matrix A. For example, to compute a least-squares fit, we compute $(X^T X)^{-1} X^T y$, which requires one matrix solve. Since A^{-1} can be computed in $O(d_1^3)$ operations, we know that solve $Ax = v_1$ can be down in $O(d_1^3)$ operations.

In practice, however, one does not solve $Ax = v_1$ for x by computing the inverse matrix A^{-1}, because there are faster and more numerically stable alternatives. These alternatives still require $O(d_1^3)$ operations, but the constants are more favorable. Two techniques are commonly used. The general-purpose solution is to first compute the QR factorization, and then with $A = QR$, solving $Rx = Q^T v_1$ can be done quickly, since solving upper triangular systems only takes $O(d_1^2)$ time. As an alternative, if A is PSD, such as in the linear regression case, we use the Cholesky factorization $A = LL^T$ where L is an lower triangular matrix. To solve $Ax = v_1$, we then solve two triangular systems $Lz = v_1$ for z and then $L^T x = z$ for x. The Cholesky decomposition is faster than the QR method, but only applies when A is PSD.

QR decomposition

The QR decomposition takes the following form:

$$A = QR,$$

where Q is $d_1 \times d_2$ with orthogonal columns, and R is upper triangular.

Cholesky decomposition

For a PSD matrix A, the Cholesky decomposition takes the following form:

$$A = LL^T$$

where L is a $d_1 \times d_1$ lower triangular matrix.

SVD and eigendecomposition

The SVD is the most important factorization for statistics, and it takes the following form:

$$A = UDV^T,$$

where U is $d_1 \times r$ with orthogonal columns, D is $r \times r$ diagonal with positive entries decreasing along the diagonal, V is $d_2 \times r$ with orthogonal columns, and r is the rank of A. The SVD exists for any matrix A. Computing the SVD is done by first computing the A^TA and then taking the eigendecomposition of this matrix, $A^TA = W\tilde{D}W^t$. Since $A^TA = VD^2V^T$ from the definition of the SVD, we have that $D^2 = \tilde{D}$ and $V = W$. U can then be obtained by matrix multiplications. Alternatively, we could compute the eigendecomposition of $AA^T = UD^2U^T$, which will be faster when $d_2 > d_1$.

The eigendecomposition is very closely related to the SVD. The real eigendecomposition exists if and only if A is symmetric, and if A is PSD then the eigendecomposition and the SVD are the same.