Linear algebra review

Tyler Chen

If you're feeling a bit rusty, these are the linear algebra highlights that you will need to get started with some of the pages on this site.

This is by no means a comprehensive introduction to linear algebra, but hopefully can provide a refresher on the topics necessary to understand the conjugate gradient algorithm. I do assume that you have seen linear algebra before, so if everything here looks foreign, I suggest taking a look at Khan Academy videos first.

Some notation

I'll generally use capital letters to denote matrices, and lower case letters to denote vectors.

When I am talking about the entries of a matrix (or vector), I will use brackets to indicate this. For instance, $[A]_{4,2}$ is the $4,2$ entry of the matrix A. If I want to take an etire row or column I will indicate this with a colon. So $[A]_{2,:}$ is the 2nd row of A (think of this as taking the $(2,i)$-entries for all i) while $[A]_{:,1}$ is the first column of A. If v is a vector, then I will often only write one index. Thus, $[v]_3$ denotes the 3-rd element of v regardless of if v is a row or column vector.

Some definitions

We will denote the *transpose* of a matrix by T, and the *conjugate transpose* (also known as *Hermitian transpose*) by H.

A matrix A is called *symmetric* if $A^\mathsf{T} = A$, and is called *Hermitian* if $A^\mathsf{H} = A$.

The identity matrix will be denoted I. Occasionally it may be denoted by I_k to emphasize that it is of size k.

A vector is called *normal* if it has norm one.

Two vectors are called *orthogonal* if their inner product is zero.

If two vectors are both normal, and are orthogonal to one another, they are called *orthonormal*.

A matrix U is called unitary if $U^{\mathsf{H}}U = UU^{\mathsf{H}} = I$. This is equivalent to all the columns being (pairwise) orthonormal.

A Hermitian (symmetric) matrix is called *positive definite* if $x^{\mathsf{H}}Ax > 0$ ($x^{\mathsf{T}}Ax > 0$) for all x. This is equivalent to having all positive eigenvalues.

An *eigenvalue* of a square matrix A is any scaler λ for which there exists a vector v so that $Av = \lambda v$. The vector v is called an *eigenvector*.

Different perspectives on matrix multiplication

Matrix vector products

Let's start with a matrix A of size $m \times n$ (m columns and n rows), and a vector v of size $n \times 1$ (n columns and 1 row).

Then the product Av is well defined, and the i-th entry of the product is given by,

$$[Av]_i = \sum_{j=1}^{n} [A]_{i,j}[v]_k$$

There are perhaps two dominant ways of thinking about this product. The first is that the i-th entry is the matrix product of the i-th row of A with v. That is,

$$[Av]_i = [A]_{i,:}v$$

Alternatively, and arguably more usefully, the product Av can be though of as the linear combination of the columns of A, where the coefficients are the entries of v. That is,

$$Av = \sum_{k=1}^{m} [v]_k [A]_{:,k}$$

For example, suppose we have vectors $q_1, q_2, \ldots, q_k \in \mathbb{R}^n$, and that Q is the $n \times k$ matrix whose columns are $\{q_1, q_2, \ldots, q_k\}$. Then saying x is in the span of $\{q_1, q_2, \ldots, q_k\}$ by deifnition means that there exists coefficients c_i such that,

$$x = c_1 q_1 + c_2 q_2 + \cdots + c_k q_k$$

This this exactly the same as saying there exists a vector $c \in \mathbb{R}^k$ such that,

$$x = Qc$$

Understanding this perspective on matrix vector products will be very useful in understanding the matrix form of the Arnolidi and Lanczos algorithms.

Matrix matrix products

Now, lets keep our matrix A of size $m \times n$, and add a matrix B of size $n \times p$. Then the product AB is well defined, and the i, j entry of the product is given by,

$$[AB]_{i,j} = \sum_{k=1}^{n} [A]_{i,k}[B]_{k,j}$$

Again we can view the i, j entry as the matrix product of the i-th row of A with the j-th column of B. That is,

$$[AB]_{i,j} = [A]_{i,:}[B]_{:,j}$$

On the other hand, we can view the j-th column of AB as the product of A with the j-th column of B. That is,

$$[AB]_{:,j} = AB_{:,j}$$

We can now use either of our perspectives on matrix vector products to view $AB_{:,j}$. This perspective is again useful for understanding the matrix forms of the Arnoldi and Lanczos algorithms.

Inner products and vector norms

Given two vectors x and y, the Euclidian inner product is defined as,

$$\langle x, y \rangle = x^{\mathsf{H}} y$$

This naturally defines the Euclidian norm (also called 2-norm) of a vector,

$$\|x\| = \|x\|_2 = \sqrt{\langle x, x \rangle}$$

A symmetric positive definite matrix A naturally induces the A-*inner product*, $\langle \cdot, \cdot \rangle_A$, defined by

$$\langle x, y \rangle_A = \langle x, Ay \rangle = \langle Ax, y \rangle$$

The associated norm, called the A-*norm* will is denoted $\| \cdot \|_A$ and is defined by,

$$\|x\|_A^2 = \langle x, x \rangle_A = \langle x, Ax \rangle = \|A^{1/2}x\|$$

Matrix norms

Usually the matrix norm 2-norm (also called operator norm, spectral norm, Euclidian norm) is defined by,

$$\|A\| = \sup_{v \neq 0} \frac{\|Av\|}{\|v\|}$$

It's always the case that the 2-norm of a matrix is the largest singular value of that matrix.

Since the singular values and eigenvalues of a positive definite matrix are the same, the 2-norm of a positive definite matrix is the largest eigenvalue.

The 2-norm is *submultiplicative*. That is, for any two matrices A and B,

$$\|AB\| \leq \|A\|\|B\|$$

The 2-norm is *unitarily invariant*. That is, if U is unitary then $\|UA\| = \|AU\| = \|A\|$.

Projections

The projection of x onto q is

$$\operatorname{proj}_q(x) = \frac{\langle x, q\rangle}{\langle q, q\rangle} q$$

If we *orthogonalize x against q*, we mean take the component of x orthogonal to q. That is,

$$x - \operatorname{proj}_q(x) = x - \frac{\langle x, q\rangle}{\langle q, q\rangle} q$$

In both cases, if q is normal, then $\langle q, q\rangle = 1$

A matrix is called a projection if $P^2 = P$. However, we will generally be more concerned with projecting onto a subspace. If Q has orthonormal columns, then QQ^{H} is a projector onto the span of the columns.

In particular, if $q_1, q_2, \dots, q_k$ are the columns of Q, then,

$$\begin{aligned} QQ^{\mathsf{H}}x &= q_1 q_1^{\mathsf{H}} x + q_2 q_2^{\mathsf{H}} x + \cdots + q_k q_k^{\mathsf{H}} x \\ &= \langle q_1, x\rangle q_1 + \langle q_2, x\rangle q_2 + \cdots + \langle q_k, x\rangle q_k \end{aligned}$$

This is just the sum of the projections of x onto each of $\{q_1, q_2, \dots, q_k\}$. Therefore, if we want to project onto a subspace V, it is generally helpful to have an orthonormal basis for this subspace.

The point in a subspace V nearest to a point x is the projection of x onto V (where projection is done with respect to the inner product and distance is measured with the induced norm).

Similarly, if we want to orthogonalize x against $q_1, q_2, \dots, q_k$ we simply remove the projection of x onto this space from x. That is,

$$x - QQ^*x = (I - QQ^*)x = x - \langle q_1, x\rangle q_1 - \langle q_2, x\rangle q_2 - \cdots - \langle q_k, x\rangle q_k$$

The resulting vector is orthogonal to each of $\{q_1, q_2, \dots, q_k\}$.