CSC320: Linear Algebra Review

Marcus Brubaker
t4brubak@cdf.toronto.edu

7th February 2007

Note that these notes are very rough. Please consult a book or some other, more authoritative, source for more detailed explanations.

1 Vectors in Euclidean Spaces

A vector in an real n-dimensional space is simply an n-tuple. That is, an ordered set of n real numbers. We denote the set of all such vectors as $\mathbb{R}^n$.

- The Euclidean inner product (also called the dot product) of two vectors $u = [u_1, \ldots, u_n]^T$ and $v = [v_1, \ldots, v_n]^T$ is defined as

$$u \cdot v = \sum_{i=1}^{n} u_i v_i$$

- The length or magnitude of a vector u is

$$\|u\| = \sqrt{u \cdot u}$$

- The distance between two vectors u and v is

$$\begin{aligned} d(u, v) &= \|u - v\| \\ &= \sqrt{\sum_{i=1}^{n} (u_i - v_i)^2} \end{aligned}$$

- The inner product is closely related to the relative orientations of u and v. Specifically,

$$u \cdot v = \|u\| \, \|v\| \cos \theta$$

  where θ is the angle between u and v. This useful relationship also gives us some intuition about the inner product.

  - If u and v are orthogonal (perpendicular) to one another ($\theta = \frac{\pi}{2}$), then $u \cdot v = 0$.

  - If u and v are pointing in similar directions ($\theta < \frac{\pi}{2}$), $u \cdot v > 0$ and $u \cdot v = \|u\| \|v\|$ if they are pointing in exactly the same direction.
  - If u and v are pointing in opposite directions ($\theta > \frac{\pi}{2}$), $u \cdot v < 0$ and $u \cdot v = -\|u\| \|v\|$ if they are pointing in exactly opposite directions.

- We can write the inner product in terms of matrix multiplication as

$$u \cdot v = u^T v$$

2 Matrices

Let A be an $n \times m$ matrix. One important quantity that describes this matrix is $\text{rank}(A)$, defined as the maximal number of linearly independent rows (or, equivalently, the maximal number of linearly independent columns). Note that $\text{rank}(A) \leq \min(m, n)$.

We are often interested in solving a linear equation of the form $Ax = b$, where x is a $m \times p$ matrix of unknowns, and b is a $m \times p$ matrix. Due to the nature of matrix multiplication we can view x and b as a collection of column vectors that can be solved for independently, so let us assume that $p = 1$. Also note that when $b = 0$ there is always a trivial solution of $x = 0$.

It is not always possible to solve such systems exactly (in which case the linear system is said to be "inconsistent") and sometimes there may be infinitely many solutions. We can break the problem into different cases, based on the rank of A and the rank of the "augmented" matrix $[A\ b]$:

- **No solution.** $\boxed{\text{rank}(A) < \text{rank}([A\ b])}$

  This is a typical result for an over-determined system (a "tall" matrix, where $n > m$), however the lack of an exact solution may also be due to numerical issues in representing floating-point numbers. Under these conditions our new goal is to find an *approximate* solution $\hat{x}$ that is closest to satisfying all the equations in a least-squares sense (see Normal Equations section below).

- **Unique solution.** $\boxed{\text{rank}(A) = \text{rank}([A\ b]) = m}$

  If the inverse of A is well-defined (a "square" matrix, with $n = m$, that is also non-singular), then the unique solution is $x = A^{-1}b$, where A^{-1} is the inverse of A.

- **Infinite number of solutions.** $\boxed{\text{rank}(A) = \text{rank}([A\ b]) < m}$

  This is a typical result for an under-determined system (a "wide" matrix, where $n < m$). More specifically, if there exists some vector $v \neq 0$ for which $Av = 0$ (in the right-nullspace of A), then all vectors of the form $x + \alpha v$ will satisfy the linear system, where x is any solution to $Ax = b$ and α is an arbitrary scalar. In this case, we're often interested in some *particular* solution, for example, one whose norm $||x||$ is smallest.

Pseudo-Inverse

A pseudo-inverse of a matrix A is a non-unique matrix which has properties similar to the inverse. The usefulness of a pseudo-inverse is that it can give reasonable solutions to matrices that wouldn't normally have them. The most common pseudo-inverse is the Moore-Penrose pseudo-inverse which is defined as:

$$A^* = \begin{cases} (A^TA)^{-1}A^T & \text{if } A^TA \text{ is invertible } (\text{rank}(A) = m) \\ A^T(AA^T)^{-1} & \text{if } AA^T \text{ is invertible } (\text{rank}(A) = n) \\ C^T(CC^T)^{-1}(B^TB)^{-1}B^T & \text{otherwise} \end{cases}$$

where $A = BC$, B is n by k, C is k by m and $k = \text{rank}(A)$. When used to find a solution to the system $Ax = b$, $\hat{x} = A^*b$ gives the shortest length least-squares solution. That is, it minimizes $\|\hat{x}\|$ and $\|A\hat{x} - b\|$.

Notice that if A is a square, invertible matrix then $A^* = A^{-1}$. Because of this the pseudo-inverse is also sometimes used when computing $A^{-1}b$ is unstable due to numerical considerations.

Normal Equations

If we have an inconsistent system of equations then there is no solution which will exactly satisfy each equation. It is then reasonable to ask if we can find an approximate solution which tries to every equation as well as possible in some sense.

Define $r = A\hat{x} - b$ to be the residual of a solution $\hat{x}$. Lets say we want the $\hat{x}$ which minimizes

$$\begin{aligned} \|r\|^2 &= (A\hat{x} - b) \cdot (A\hat{x} - b) \\ &= (A\hat{x} - b)^T(A\hat{x} - b) \\ &= (\hat{x}^TA^T - b^T)(A\hat{x} - b) \\ &= \hat{x}^TA^TA\hat{x} - \hat{x}^TA^Tb - bA\hat{x} + b^Tb \end{aligned}$$

We resort to calculus to solve this

$$\frac{\partial}{\partial \hat{x}} \|r\|^2 = 2A^TA\hat{x} - 2A^Tb$$

and by setting the derivative equal to zero and solving we get a new set of equations

$$A^TA\hat{x} = A^Tb$$

which is called the normal equations.

Notice the similarity to the psuedo-inverse. If $\text{rank}(A) = m$ then A^TA is invertible and we can solve for

$$\begin{aligned} \hat{x} &= (A^TA)^{-1}A^Tb \\ &= A^*b \end{aligned}$$

Orthogonal Matrices

We call a matrix A orthogonal when its rows/columns are not just linearly independant but also orthogonal and unit length. That is, for the ith and jth rows/columns r_i and r_j of A

$$r_i \cdot r_j = \begin{cases} 0 & i \neq j \\ 1 & i = j \end{cases}$$

This gives us a particularly nice property which is $A^TA = I = AA^T$. By the uniqueness of inverses this means that $A^T = A^{-1}$.

Singular Value Decomposition

Computing and inverting A^TA is often expensive and numerically unstable. There are other ways to solve the least-squares problem without computing A^TA directly. One method uses the singular value decomposition (SVD). Any matrix can be decomposed such that

$$A = USV^T$$

where U is a n by n orthogonal matrix, S is an n by m matrix whose diagonal entries are the singular values and V is a m by m orthogonal matrix. Strictly speaking this decomposition is not unique can be made unique by requiring that $S(1,1) \geq S(2,2) \geq S(3,3) \geq \cdots$.

To solve the normal equations substitute USV^T for A

$$\begin{aligned} \hat{x} &= (A^TA)^{-1}A^Tb \\ &= (VS^TU^TUSV^T)^{-1}VS^TU^Tb \\ &= (VS^TSV^T)^{-1}VS^TU^Tb \\ &= V(S^TS)^{-1}V^TVS^TU^Tb \\ &= V(S^TS)^{-1}S^TU^Tb \\ &= VS^{-1}U^Tb \end{aligned}$$

Notice that S is a non-square matrix. Since S is diagonal we abuse notation and define S^{-1} to be a m by n matrix with diagonal entries $S^{-1}(i,i) = \frac{1}{S(i,i)}$. Showing that $(S^TS)^{-1}S^T = S^{-1}$ is left as an exercise.

SVD can also be used to calculate the Moore-Penrose pseudo-inverse without worrying about the rank of the matrix

$$A^* = VS^{-1*}U^T$$

where S^{-1*} is a m by n matrix with diagonal entries

$$S^{-1*}(i,i) = \begin{cases} \frac{1}{S(i,i)} & S(i,i) \neq 0 \\ 0 & S(i,i) = 0 \end{cases}$$