Univ. of Washington

Lecture : Vector Products and Matrix Multiplication

Winter 2021

Lecturer: Dan Calderone

Inner products

General notation: $\langle y, x\rangle$

Specific inner products:

- Vectors in $\mathbb{R}^n$: $\langle y, x\rangle = y \cdot x = y^T x = \sum_{i=1}^n y_i x_i$
- Vectors in $\mathbb{C}^n$: $\langle y, x\rangle = y^* x = \sum_{i=1}^n y_i^* x_i$
- Integrable functions on $f : [0,1] \to \mathbb{C}^n$: $\langle f, g\rangle = \int_{[0,1]} f^*(t)g(t)\, dt$

One of the fundamental uses of an inner product is to compute the *2-norm* or *length* of a vector by taking an inner product of vector with itself. $|x|_2 = \sqrt{\langle x, x\rangle}$. More generally, inner products tell you how much two vectors *line up with each other*. Along these lines, we have the identity

$$\sqrt{\langle x, x\rangle} = y^T x = |y||x|\cos(\theta) \tag{1}$$

where θ is the angle between x and y. A way to see this directly is to apply the law of cosines to $|x - y|^2$

$$(x-y)^T(x-y) = x^T x + y^T y - 2x^T y = |x|^2 + |y|^2 - 2|x||y|\cos(\theta) \tag{2}$$

When $y^T x = 0$, $cos(\theta) = 0$ and the angle between the two vectors is either 90^o and -90^o and the vectors are *perpendicular* or *orthogonal*. If y is a *unit vector*, ie. $|y| = 1$, then $y^T x = |x|cos(\theta)$, ie. $y^T x$ is the amount of x in the direction of y. If we then multiply this quantity by the unit vector y again, we get the component of x in the y-direction or the *projection of x onto y*, $\text{proj}_y x$. If y is not a unit vector, we can use the unit vector $y/|y|$. This leads to the general formula for a 1-dimensional projection matrix

$$\text{proj}_y x = \frac{1}{|y|^2} y y^T x = y(y^T y)^{-1} y^T x \tag{3}$$

More generally, if we want to project x onto a large subspace spanned by the columns of Y, we can compute

$$\text{proj}_Y x = Y(Y^T Y)^{-1} Y^T x \tag{4}$$

Outer Products

The *outer product* of x and y is given by

$$xy^T = \begin{bmatrix} x_1y_1 & \cdots & x_1y_n \\ \vdots & & \vdots \\ x_ny_1 & \cdots & x_ny_n \end{bmatrix} \tag{5}$$

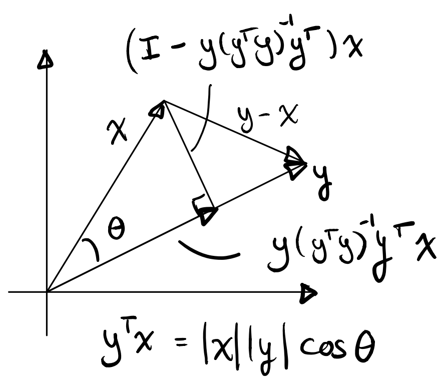
Outer products are clearly rank-1 and are sometimes called *dyads*. Note that a 1-dimensional projection matrix is the outer product of a unit vector with itself.

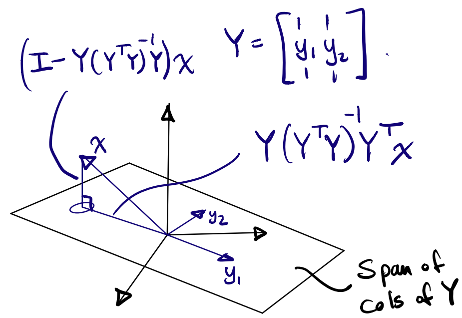
Matrix Inner Products

Let $X, Y \in \mathbb{R}^{nxm}$. The inner product of two matrices is

$$\sum_i \sum_j X_{ij} Y_{ij} = \text{Tr}(Y^T X) \tag{6}$$

where the trace operator $\text{Tr}(\cdot)$ is the sum of the diagonal elements. The Frobenius-norm of a matrix is equivalent to the vector two norm $|X|_F = \sqrt{\text{Tr}(X^T X)}$.





Norms

Properties of Norms

For a vector space $\mathcal{V}$ over a field $\mathcal{F}$, a **norm** is a nonnegative-valued function $\|\cdot\| : \mathcal{V} \to \mathbb{R}$.

For all $a \in \mathcal{F}$ and all $v, u \in \mathcal{V}$

Subadditivity/triangle inequality:	$\|u+v\| \leq \|u\| + \|v\|$
Absolute homogeneity:	$\|av\| = \|a\|\|v\|$
Nonnegativity:	$\|v\| \geq 0$
Zero vector:	if $\|v\| = 0$, then $v = 0$

For convenience from here on, we will use $|\cdot|$ for both absolute values and norms.

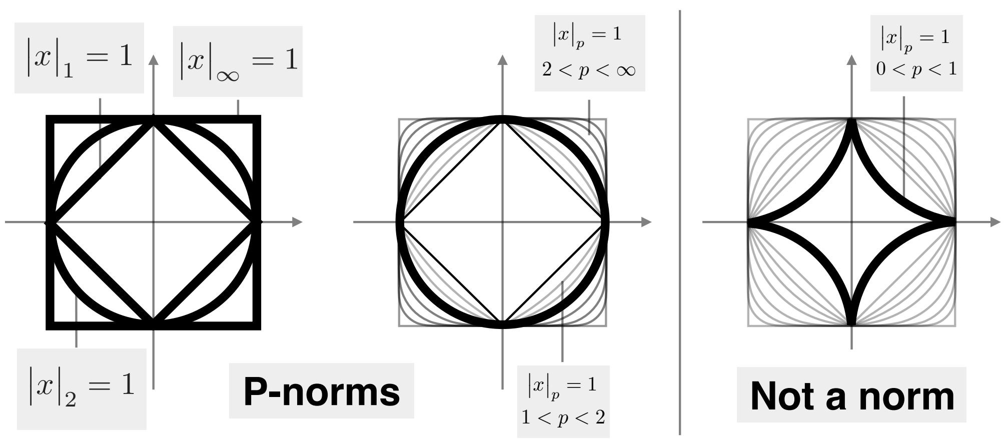
Vector Norms

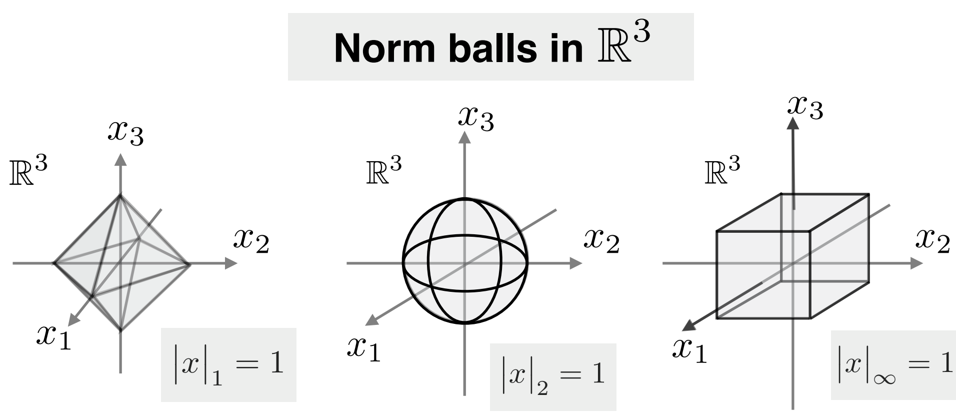
$$\text{p-norm:} \qquad |x|_p = \left(\sum_i |x_i|^p \right)^{\frac{1}{p}}, \qquad 1 \le p \le \infty$$

$$\text{2-norm:} \qquad |x|_2 = \left(\sum_i |x_i|^2 \right)^{\frac{1}{2}}$$

$$\text{1-norm:} \qquad |x|_1 = \left(\sum_i |x_i| \right)^{1}$$

$$\infty\text{-norm:} \qquad |x|_\infty = \lim_{p \to \infty} \left(\sum_i |x_i|^p \right)^{\frac{1}{p}} = \max_i |x_i|$$





Matrix Norms

Norms for matrices either think of the matrix as a reshaped vector (**element-wise norms**) or as an operator on vector spaces. Norms that treat matrices as operators are called **induced norms**.

Element-wise Matrix Norms

An element-wise matrix 2-norm is called the **Frobenius norm**, $\left|\cdot\right|_{\mathrm{F}}$. For $A \in \mathbb{R}^{m\times n}$

$$\left|A\right|_{\mathrm{F}} = \sum_{ij} \left|A_{ij}\right|^2 = \left(\mathrm{Tr}(A^*A)\right)^{\frac{1}{2}}$$

Note that considering the SVD of $A \in \mathbb{R}^{m\times n}$ (see later on)

$$A = U \begin{bmatrix} \Sigma & 0 \\ 0 & 0 \end{bmatrix} V^*, \qquad \Sigma = \begin{bmatrix} \sigma_1 & \cdots & 0 \\ \vdots & & \vdots \\ 0 & \cdots & \sigma_k \end{bmatrix}$$

and applying properties of traces (see later on), we get $|A|_{\mathrm{F}} = \left|\mathrm{diag}(\Sigma)\right|_2$, ie. the Frobenius norm is the 2-norm applied to a vector of the singular values.

$$\begin{aligned} \left|A\right|_{\mathrm{F}} &= \left(\sum_{ij} \left|A_{ij}\right|^2\right)^{\frac{1}{2}} \\ &= \left(\mathrm{Tr}(A^*A)\right)^{\frac{1}{2}} \\ &= \left(\mathrm{Tr}\left(V \begin{bmatrix} \Sigma & 0 \\ 0 & 0 \end{bmatrix} U^*U \begin{bmatrix} \Sigma & 0 \\ 0 & 0 \end{bmatrix} V^*\right)\right)^{\frac{1}{2}} \\ &= \left(\mathrm{Tr}\left(\begin{bmatrix} \Sigma^2 & 0 \\ 0 & 0 \end{bmatrix} V^*V\right)\right)^{\frac{1}{2}} = \left(\sum_i \sigma_i^2\right)^{\frac{1}{2}} \end{aligned}$$

Induced Matrix Norms

Induced matrix norms intuitively measure how much a matrix increases (or decreases) the size of vectors it acts on. The induced p, q-norm of $A \in \mathbb{R}^{m\times n}$ gives the maximum q-norm of a vector $\left|Ax\right|_\beta$ where x is chosen from the unit ball of the p-norm.

$$\left|A\right|_{p,q} = \max_{|x|_p=1} \left|Ax\right|_q$$

or, equivalently.

$$\left|A\right|_{p,q} = \max_{x\neq 0} \frac{\left|Ax\right|_q}{\left|x\right|_p}$$

Sometimes we use $|\cdot|_p$ to refer to the induced p, p-norm. Some specific induced norm examples (again with SVD given above).

$$
\begin{aligned}
|A|_2 = |A|_{2,2} &= \max_{|x|_2=1} |Ax|_2 \\
&= \max_{|x|_2=1} (x^* A^* A x)^{\frac{1}{2}} \\
&= \max_{|x|_2=1} \left(x^* V \begin{bmatrix} \Sigma^2 & 0 \\ 0 & 0 \end{bmatrix} V^* x \right)^{\frac{1}{2}} = \sigma_{\text{max}}
\end{aligned}
$$

Block Matrix Multiplication

Consider a matrix $A \in \mathbb{R}^{m \times n}$ divided up into elements, columns, and rows

$$
A = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix} = \begin{bmatrix} | & \cdots & | \\ A_{:1} & & A_{:n} \\ | & \cdots & | \end{bmatrix} = \begin{bmatrix} - & A_{1:} & - \\ \vdots & & \vdots \\ - & A_{n:} & - \end{bmatrix} \tag{7}
$$

where we use the Matlab inspired notation $A_{:j}$ and $A_{i:}$ to represent the ith row and jth column of A respectively. We can define multiplying A by a vector x as

$$
Ax = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} a_{11}x_1 + \cdots + a_{1n}x_n \\ \vdots \\ a_{m1}x_1 + \cdots + a_{mn}x_n \end{bmatrix} \tag{8}
$$

$$
= \begin{bmatrix} | \\ A_{:1} \\ | \end{bmatrix} x_1 + \cdots + \begin{bmatrix} | \\ A_{:n} \\ | \end{bmatrix} x_n = \begin{bmatrix} [\,-A_{1:}-\,]x \\ \vdots \\ [\,-A_{m:}-\,]x \end{bmatrix} \tag{9}
$$

Note that we can interpret Ax as x selecting a particular linear combination of the columns of A. The *range of* A is the span of the columns of A, ie. the set of vectors $y \in \mathbb{R}^m$ that can be reached by selecting a suitable x, $y = Ax$. Alternatively, we can interpret Ax as taking the inner product between x with each of the rows of A. The *nullspace of* A is the set of vectors $x \in \mathbb{R}^n$ such that $Ax = 0$ or the set of vectors that are orthogonal to each of the rows of A.

We now consider multiplying two matrices $A \in \mathbb{R}^{m \times n}$ and $B \in \mathbb{R}^{n \times k}$. Note that the inner dimensions must match.

$$
AB = \begin{bmatrix} a_{11}b_{11} + \cdots + a_{1n}b_{1n} & \cdots & a_{11}b_{1k} + \cdots + a_{1n}b_{nk} \\ \vdots & & \vdots \\ a_{m1}b_{11} + \cdots + a_{mn}b_{1n} & \cdots & a_{m1}b_{1k} + \cdots + a_{mn}b_{nk} \end{bmatrix} \tag{10}
$$

Note that this same formula works if you divide A and B into sub or *block matrices*.

$$A = \begin{bmatrix} A_{11} & \cdots & A_{1n} \\ \vdots & & \vdots \\ A_{m1} & \cdots & A_{mn} \end{bmatrix}, \qquad B = \begin{bmatrix} B_{11} & \cdots & B_{1k} \\ \vdots & & \vdots \\ B_{n1} & \cdots & B_{nk} \end{bmatrix} \tag{11}$$

$$AB = \begin{bmatrix} A_{11}B_{11} + \cdots + A_{1n}B_{1n} & \cdots & A_{11}B_{1k} + \cdots + A_{1n}B_{nk} \\ \vdots & & \vdots \\ A_{m1}B_{11} + \cdots + A_{mn}B_{1n} & \cdots & A_{m1}B_{1k} + \cdots + A_{mn}B_{np} \end{bmatrix} \tag{12}$$

Note that we can divide up A and B into any size sub-blocks as long as the inner dimensions of each appropriate A_{ij} and B_{jk} match. Two specific interesting cases are if we divide up A and B into columns or rows. Dividing A into rows and B into columns gives

$$AB = \begin{bmatrix} - & A_{1:} & - \\ \vdots & & \vdots \\ - & A_{n:} & - \end{bmatrix} \begin{bmatrix} | & \cdots & | \\ B_{:1} & & B_{:p} \\ | & \cdots & | \end{bmatrix} = \begin{bmatrix} A_{1:}B_{:1} & \cdots & A_{1:}B_{:p} \\ \vdots & & \vdots \\ A_{m:}B_{:1} & \cdots & A_{m:}B_{:p} \end{bmatrix} \tag{13}$$

Here we are taking *the inner products of each row of A with each column of B.* . We could also divide up A into columns and B into rows.

$$AB = \begin{bmatrix} | & \cdots & | \\ A_{:1} & & A_{:n} \\ | & \cdots & | \end{bmatrix} \begin{bmatrix} - & B_{1:} & - \\ \vdots & & \vdots \\ - & B_{n:} & - \end{bmatrix} = \begin{bmatrix} | \\ A_{:1} \\ | \end{bmatrix} \begin{bmatrix} - & B_{1:} & - \end{bmatrix} + \cdots + \begin{bmatrix} | \\ A_{:n} \\ | \end{bmatrix} \begin{bmatrix} - & B_{n:} & - \end{bmatrix} \tag{14}$$

Note that here, we have computed the *sum of the outer products of the matched columns of A and rows of B.*

We also note the following useful extension of this concept. Consider $A \in \mathbb{R}^{m \times n}$ $M \in \mathbb{R}^{n \times p}$, and $B \in \mathbb{R}^{p \times q}$. Using the inner product form above, we can compute

$$AMB = \begin{bmatrix} A_{1:}MB_{:1} & A_{1:}MB_{:q} \\ \vdots & \vdots \\ A_{m:}MB_{:1} & A_{m:}MB_{:q} \end{bmatrix} \tag{15}$$

It is worth noting that $[AMB]_{ij} = A_{i:}MB_{:j}$ Using the outer product form, we can compute

$$AMB = \sum_k \sum_l \begin{bmatrix} | \\ A_{:k} \\ | \end{bmatrix} M_{kl} \begin{bmatrix} - & B_{l:} & - \end{bmatrix} \tag{16}$$

Note that M_{kl} gives the scaling factor for the dyad $A_{:k}B_{l:}$. In (14), we have taken M to be the identity. Some other common and useful examples of block matrix multiplication are given by

$$AB = A\begin{bmatrix} B_1 & \cdots & B_k \end{bmatrix} = \begin{bmatrix} AB_1 & \cdots & AB_k \end{bmatrix} \tag{17}$$

Note in this example, if each B_j is a column, we can think of the matrix A as transforming each column separately.

$$AB = \begin{bmatrix} A_1 \\ \vdots \\ A_n \end{bmatrix} B = \begin{bmatrix} A_1 B \\ \vdots \\ A_n B \end{bmatrix} \tag{18}$$

$$AB = \begin{bmatrix} A_1 & \cdots & A_n \end{bmatrix} \begin{bmatrix} B_1 \\ \vdots \\ B_n \end{bmatrix} = A_1B_1 + \cdots + A_nB_n \tag{19}$$

$$AB = \begin{bmatrix} A_1 \\ \vdots \\ A_m \end{bmatrix} \begin{bmatrix} B_1 & \cdots & B_k \end{bmatrix} = \begin{bmatrix} A_1B_1 & \cdots & A_1B_k \\ \vdots & & \vdots \\ A_mB_1 & \cdots & A_mB_k \end{bmatrix} \tag{20}$$