Lecture 1

Matrix-Vector Multiplication

NLA Reading Group Spring'13

by İsmail Arı

$b = Ax$ ➔ *b* is a linear combination of the columns of *A*

$b = Ax$ ➔ $b_i = \sum_{j=1}^{n} a_{ij} x_j, \qquad i = 1, \ldots, m.$

The map $x \mapsto Ax$ is *linear* ➔ for any $x, y \in \mathbb{C}^n$ and $\alpha \in \mathbb{C}$

$$\begin{aligned} A(x+y) &= Ax + Ay, \\ A(\alpha x) &= \alpha A x. \end{aligned}$$

A Matrix Times a Vector

Let us re-write the matrix-vector multiplication

$$b = Ax = \sum_{j=1}^{n} x_j a_j$$

$$\begin{bmatrix} \\ b \\ \\ \end{bmatrix} = \begin{bmatrix} & & & \\ a_1 & a_2 & \cdots & a_n \\ & & & \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = x_1 \begin{bmatrix} \\ a_1 \\ \\ \end{bmatrix} + x_2 \begin{bmatrix} \\ a_2 \\ \\ \end{bmatrix} + \cdots + x_n \begin{bmatrix} \\ a_n \\ \\ \end{bmatrix}$$

"As mathematicians, we are used to viewing the formula $Ax = b$ as a statement that

A acts on x to produce b

The new formula, by contrast, suggests the interpretation that

x acts on A to produce b

Example: Vandermonde Matrix

The map from vectors of coefficients of polynomials p of degree $< n$ to vectors $(p(x_1), p(x_2), \dots, p(x_m))$ of sampled polynomial values is linear.

$$A = \begin{bmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^{n-1} \\ 1 & x_2 & x_2^2 & \cdots & x_2^{n-1} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_m & x_m^2 & \cdots & x_m^{n-1} \end{bmatrix}$$

$$c = \begin{bmatrix} c_0 \\ c_1 \\ c_2 \\ \vdots \\ c_{n-1} \end{bmatrix}, \qquad p(x) = c_0 + c_1 x + c_2 x^2 + \cdots + c_{n-1} x^{n-1}$$

The product Ac gives the sampled polynomial values:

$$(Ac)_i \;=\; c_0 + c_1 x_i + c_2 x_i^2 + \cdots + c_{n-1} x_i^{n-1} \;=\; p(x_i)$$

Do not see Ac as m distinct scalar summations. Instead, see A as a matrix of columns, each giving sampled values of a monomial*,

$$A = \left[\begin{array}{c|c|c|c|c} 1 & x & x^2 & \cdots & x^{n-1} \end{array} \right]$$

Thus, Ac is a single vector summation that at once gives a linear combination of these monomials,

$$Ac = c_0 + c_1 x + c_2 x^2 + \cdots + c_{n-1} x^{n-1} = p(x)$$

*In mathematics, a monomial is roughly speaking, a polynomial which has only one term.

A Matrix Times a Vector

$B = AC$ ➔ each column of B is a linear combination of the columns of A

$$b_j = Ac_j = \sum_{k=1}^{m} c_{kj} a_k$$

$$\begin{bmatrix} \\ b_1 \Bigg| b_2 \Bigg| \cdots \Bigg| b_n \\ \\ \end{bmatrix} = \begin{bmatrix} \\ a_1 \Bigg| a_2 \Bigg| \cdots \Bigg| a_m \\ \\ \end{bmatrix} \begin{bmatrix} c_1 \Bigg| c_2 \Bigg| \cdots \Bigg| c_n \end{bmatrix}$$

Thus b_j is a linear combinations of the columns a_k with coefficients c_{kj}

Example: Outer Product

$$\begin{bmatrix} \\ u \\ \\ \end{bmatrix} \begin{bmatrix} v_1 & v_2 & \cdots & v_n \end{bmatrix} = \left[\begin{array}{c|c|c|c} & & & \\ v_1 u & v_2 u & \cdots & v_n u \\ & & & \end{array}\right] = \begin{bmatrix} v_1 u_1 & \cdots & v_n u_1 \\ \vdots & & \vdots \\ v_1 u_m & \cdots & v_n u_m \end{bmatrix}$$

Example: Cumulative Sum

$$B = AR$$

$$\begin{bmatrix} \Big| & & \Big| \\ b_1 & \cdots & b_n \\ \Big| & & \Big| \end{bmatrix} = \begin{bmatrix} \Big| & & \Big| \\ a_1 & \cdots & a_n \\ \Big| & & \Big| \end{bmatrix} \begin{bmatrix} 1 & \cdots & 1 \\ & \ddots & \vdots \\ & & 1 \end{bmatrix}$$

$$b_j = Ar_j = \sum_{k=1}^{j} a_k$$

The matrix *R* is a discrete analogue of an indefinite integral operator

Range

range(A) is the space spanned by the columns of A

Nullspace

null(A) is the set of vectors that satisfy $Ax = 0$, where 0 is the 0-vector in $\mathbb{C}^m$

Rank

The column/row rank of a matrix is the dimension of its column/row space.
Column rank always equals row rank. So, we call this as *rank* of the matrix.

A matrix A of size m-by-n with $m \geq n$ has full rank iff it maps no two distinct vectors to the same vector.

Inverse

A nonsingular or invertible matrix is a square matrix of full rank.

$$\left[\begin{array}{c|c|c} e_1 & \cdots & e_m \end{array}\right] = I = AZ$$

I is the m-by-m identity. The matrix Z is the inverse of A.

$$AA^{-1} = A^{-1}A = I$$

For an m-by-m matrix A, the following conditions are equivalent:

(a) *A has an inverse A^{-1},*
(b) $\operatorname{rank}(A) = m$,
(c) $\operatorname{range}(A) = \mathbb{C}^m$,
(d) $\operatorname{null}(A) = \{0\}$,
(e) *0 is not an eigenvalue of A,*
(f) *0 is not a singular value of A,*
(g) $\det(A) \neq 0$.

We mention the determinant, though a convenient notion theoretically, rarely finds a useful role on numerical algorithms.

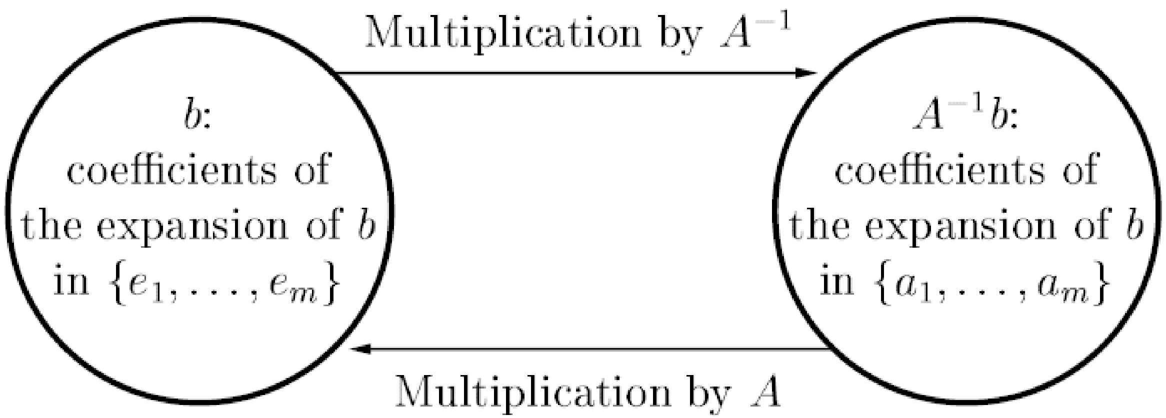
A Matrix Times a Vector

$$x = A^{-1}b$$

Do not think x as the result of applying A^{-1} to b. Instead, think it as the unique vector that satisfies the equation $Ax = b$.

Multiplication by A^{-1} is a change of basis operation.

$A^{-1}b$ is the vector of coefficients of the expansion of b in the basis of columns of A.



Lecture 2

Orthogonal Vectors and Matrices

NLA Reading Group Spring’13

by İsmail Arı

Adjoint

The complex conjugate of a scalar z, written $\bar{z}$ or z^*, is obtained by negating its imaginary part.

The hermitian conjugate or adjoint of an m-by-n matrix A, written A^*, is the n-by-m matrix whose i, j entry is the complex conjugate of the j, i entry of A.

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \end{bmatrix} \quad \Longrightarrow \quad A^* = \begin{bmatrix} \overline{a}_{11} & \overline{a}_{21} & \overline{a}_{31} \\ \overline{a}_{12} & \overline{a}_{22} & \overline{a}_{32} \end{bmatrix}$$

If $A = A^*$, A is hermitian.

For real A, adjoint is known as transpose and shown as A^T.

If $A = A^T$, then A is symmetric.

Inner Product

$$x^* y = \sum_{i=1}^{m} \overline{x}_i y_i$$

Euclidean length of x

$$\|x\| = \sqrt{x^* x} = \left(\sum_{i=1}^{m} |x_i|^2 \right)^{1/2}$$

The inner product is bilinear, i.e. linear in each vector separately:

$$\begin{aligned} (x_1 + x_2)^* y &= x_1^* y + x_2^* y, \\ x^* (y_1 + y_2) &= x^* y_1 + x^* y_2, \\ (\alpha x)^* (\beta y) &= \overline{\alpha} \beta x^* y. \end{aligned}$$

Orthogonal Vectors

A pair of vectors x and y are orthogonal if $x^*y = 0$.

Two sets of vectors X and Y are orthogonal if every $x \in X$ is orthogonal to $y \in Y$.

A set of nonzero vectors S is orthogonal if its elements are pairwise orthogonal.

A set of nonzero vectors S is orthonormal if it is orthogonal, in addition, every $x \in S$ has $\|x\| = 1$.

The vectors in an orthogonal set S are linearly independent.

Sketch of the proof:

- ❑ Assume than they were not independent and propose a nonzero vector by linear combination of the members of S
- ❑ Observe that its length should be larger than 0
- ❑ Use the bilinearity of inner products and the orthogonality of S to contradict the assumption

$\Rightarrow$ If an orthogonal set $S \subseteq \mathbb{C}^m$ contains m vectors, then it is a basis for $\mathbb{C}^m$.

Components of a Vector

Inner products can be used to decompose arbitrary vectors into orthogonal components.

Assume

$\{q_1, q_2, \dots, q_n\}$: an orthonormal set
v: an arbitrary vector

Utilizing the scalars $q_j^* v$ as coordinates in an expansion, we find that

$$r = v - (q_1^* v)q_1 - (q_2^* v)q_2 - \cdots - (q_n^* v)q_n$$

is orthogonal to $\{q_1, q_2, \dots, q_n\}$

Thus we see that v can be decomposed into $n + 1$ orthogonal components:

$$v = r + \sum_{i=1}^{n} (q_i^* v)q_i = r + \sum_{i=1}^{n} (q_i q_i^*)v$$

If $\{q_i\}$ is a basis for $\mathbb{C}^m$, then n must be equal to m

$$v = \underbrace{\sum_{i=1}^{m} (q_i^* v) q_i}_{1} = \underbrace{\sum_{i=1}^{m} (q_i q_i^*) v}_{2}$$

1 We view v as a sum of coefficients $q_j^* v$ times vectors q_i.

2 We view v as a sum of orthogonal projections of v onto the various directions of q_i. The ith projection operation is achieved by the very special rank-one matrix $q_i q_i^*$.

Unitary Matrices

If $Q^* = Q^{-1}$, Q is unitary.

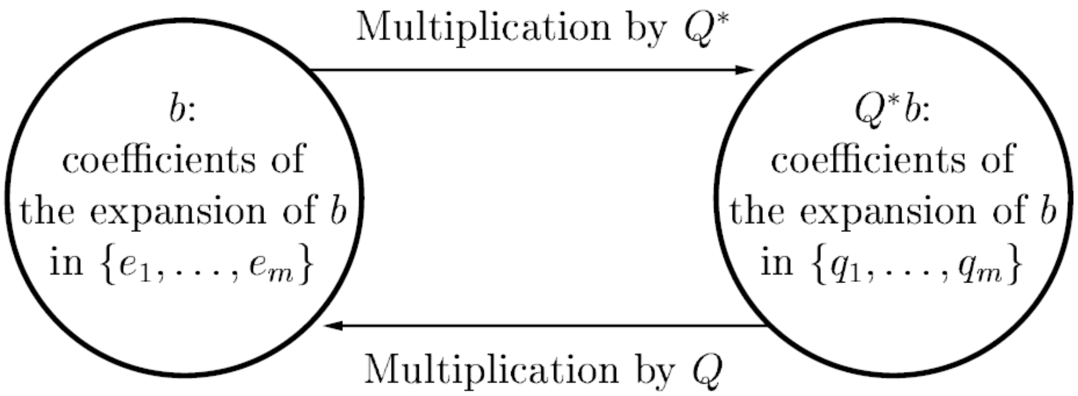
$$Q^*Q = I$$

$$\begin{bmatrix} \quad q_1^* \quad \\ \hline q_2^* \\ \hline \vdots \\ \hline q_m^* \end{bmatrix} \left[\begin{array}{c|c|c|c} q_1 & q_2 & \cdots & q_m \end{array} \right] = \begin{bmatrix} 1 & & & \\ & 1 & & \\ & & \ddots & \\ & & & 1 \end{bmatrix}$$

$$q_i^* q_j = \delta_{ij}$$

Multiplication by a Unitary Matrix

Q^*b is the vector of coefficients of the expansion of b in the basis of columns of A.



Multiplication by a unitary matrix or its adjoint preserve geometric structure in the Euclidean sense, because inner products are preserved.

$$(Qx)^*(Qy) = x^*y$$

The invariance of inner products means that angles between vectors are preserved, and so are their lengths:

$$||Qx|| = ||x||$$

In the real case, multiplication by an orthonormal matrix Q corresponds to a rigid rotation (if $\det Q = 1$) or reflection (if $\det Q = -1$) of the vector space.

Lecture 3
Norms

NLA Reading Group Spring’13

by İsmail Arı

Vector norms

The essential notions of size and distance in a vector space are captured by norms.

A *norm* is a function $||\cdot|| : \mathbb{C}^m \to \mathbb{R}$

In order to conform a reasonable notion of length, a norm must satisfy

$$(1)\quad ||x|| \geq 0, \text{ and } ||x|| = 0 \text{ only if } x = 0,$$
$$(2)\quad ||x+y|| \leq ||x|| + ||y||,$$
$$(3)\quad ||\alpha x|| = |\alpha|\, ||x||.$$

for all vectors x and y and for all scalars $\alpha \in \mathbb{C}$.

p-norms

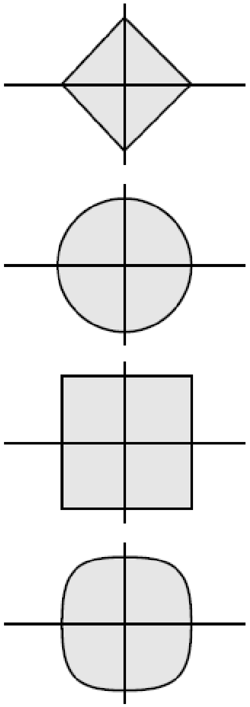
The closed unit ball $\{x \in \mathbb{C}^m\colon \|x\| \leq 1\}$ corresponding to each norm is illustrated to the right for the case $m = 2$.

$$\|x\|_1 = \sum_{i=1}^{m} |x_i|,$$

$$\|x\|_2 = \left(\sum_{i=1}^{m} |x_i|^2 \right)^{1/2} = \sqrt{x^* x},$$

$$\|x\|_\infty = \max_{1 \leq i \leq m} |x_i|,$$

$$\|x\|_p = \left(\sum_{i=1}^{m} |x_i|^p \right)^{1/p} \quad (1 \leq p < \infty).$$



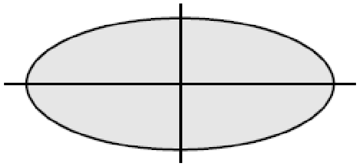
Weighted p-norms

Introduce the diagonal matrix W whose ith diagonal entry is the weight $w_i \neq 0$.

$$\|x\|_W = \|Wx\|$$

Example: a weighted 2-norm

$$\|x\|_W = \left(\sum_{i=1}^{m} |w_i x_i|^2 \right)^{1/2}.$$



The most important norms in this book are the unweighted 2-norm and its induced matrix form.

Matrix Norms Induced by Vector Norms

An $m \times n$ matrix can be viewed as a vector in an mn-dimensional space: each of the mn entries of the matrix is an independent coordinate.

$\Rightarrow$ Any mn-dimensional norm can be used for measuring the "size" of such a matrix.

However, certain special matrix norms are more useful than the vector norms.

These are the induced matrix norms, defined in terms of the behavior of a matrix as an operator between its normed domain and range spaces.

Given vector norms $\|\cdot\|_{(n)}$ and $\|\cdot\|_{(m)}$ on the domain and range of $A \in \mathbb{C}^{m\times n}$, respectively, the induced matrix norm $\|A\|_{(m,n)}$ is the smallest number C for which

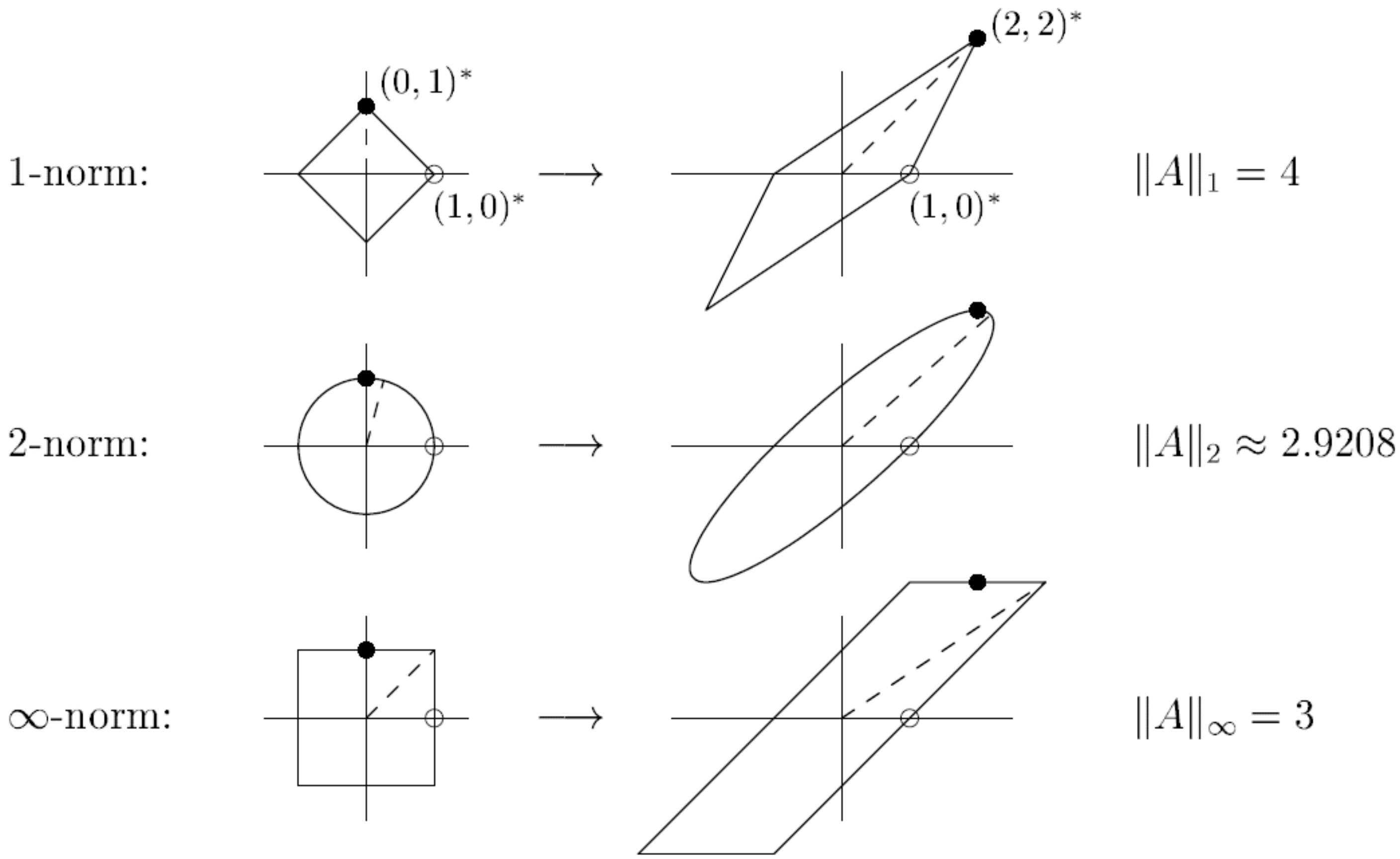
$$\|Ax\|_{(m)} \leq C\|x\|_{(n)}$$

In other words, it is the maximum factor by which A can stretch a vector x.

$$\|A\|_{(m,n)} = \sup_{\substack{x\in\mathbb{C}^n \\ x\neq 0}} \frac{\|Ax\|_{(m)}}{\|x\|_{(n)}} = \sup_{\substack{x\in\mathbb{C}^n \\ \|x\|_{(n)}=1}} \|Ax\|_{(m)}$$

A Toy Example

$$A = \begin{bmatrix} 1 & 2 \\ 0 & 2 \end{bmatrix}$$



The p-norm of a Diagonal Matrix

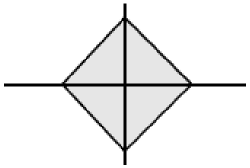
$$D = \begin{bmatrix} d_1 & & & \\ & d_2 & & \\ & & \ddots & \\ & & & d_m \end{bmatrix}$$

$$\|D\|_p = \max_{1 \leq i \leq m} |d_i|$$

The 1-norm of a Matrix

For any $m \times n$ matrix A, $\|A\|_1$ is equal to the maximum column sum of A.

$$A = \left[\begin{array}{c|c|c} a_1 & \cdots & a_n \end{array} \right]$$

Consider x be in  $\{x \in \mathbb{C}^n : \sum_{j=1}^{n} |x_j| \leq 1\}$

$$\|Ax\|_1 = \left\| \sum_{j=1}^{n} x_j a_j \right\|_1 \leq \sum_{j=1}^{n} |x_j| \, \|a_j\|_1 \leq \max_{1 \leq j \leq n} \|a_j\|_1$$

By choosing $x = e_j$, where j maximizes $\|a_j\|_1$, we attain:

$$\|A\|_1 = \max_{1 \leq j \leq n} \|a_j\|_1$$

The ∞-norm of a Matrix

For any $m \times n$ matrix A, $||A||_\infty$ is equal to the maximum row sum of A.

$$||A||_\infty = \max_{1 \leq i \leq m} ||a_i^*||_1$$

Cauchy-Schwartz and Hölder Inequalities

Let p and q satisfy $\frac{1}{p} + \frac{1}{q} = 1$, with $1 \leq p, q \leq \infty$. Then, the Hölder inequality states that, for any vectors x and y:

$$|x^*y| \leq ||x||_p ||y||_q$$

The Cauchy-Schwartz inequality is a special case $p = q = 2$:

$$|x^*y| \leq ||x||_2 ||y||_2$$

The 2-norm of a Row Vector

Consider $A = a^*$ where a is a column vector. For any x, we have:

$$\|Ax\|_2 = |a^*x| \leq \|a\|_2\|x\|_2$$

This bound is tight: observe that

$$\|Aa\|_2 = \|a\|_2^2$$

Therefore, we have

$$\|A\|_2 = \sup_{x \neq 0}\{\|Ax\|_2/\|x\|_2\} = \|a\|_2$$

The 2-norm of an Outer Product

Consider $A = uv^*$, where u is an m-vector and v is an n-vector. For any n-vector x, we can bound

$$||Ax||_2 = ||uv^*x||_2 = ||u||_2|v^*x| \leq ||u||_2||v||_2||x||_2$$

Therefore, we have

$$||A||_2 \leq ||u||_2||v||_2$$

This inequality is an equality for the case $x = v$.

Bounding $\|AB\|$ in an Induced Matrix Norm

$$\|ABx\|_{(\ell)} \leq \|A\|_{(\ell,m)}\|Bx\|_{(m)} \leq \|A\|_{(\ell,m)}\|B\|_{(m,n)}\|x\|_{(n)}$$

Therefore, the induced norm of AB must satisfy

$$\|AB\|_{(\ell,n)} \leq \|A\|_{(\ell,m)}\|B\|_{(m,n)}$$

General Matrix Norms

(1) $\|A\| \geq 0$, and $\|A\| = 0$ only if $A = 0$,

(2) $\|A + B\| \leq \|A\| + \|B\|$,

(3) $\|\alpha A\| = |\alpha| \, \|A\|$.

Frobenius Norm

The most important matrix norm which is not induced by a vector norm is the Hilbert-Schmidt or Frobenius norm, defined by

$$\|A\|_F = \left(\sum_{i=1}^{m} \sum_{j=1}^{n} |a_{ij}|^2 \right)^{1/2}$$

Observe that this s the same as the 2-norm of the matrix when viewed as an *mn*-dimensional vector.

Alternatively, we can write

$$\|A\|_F = \left(\sum_{j=1}^{n} \|a_j\|_2{}^2 \right)^{1/2}$$

$$\|A\|_F = \sqrt{\mathrm{tr}(A^* A)} = \sqrt{\mathrm{tr}(A A^*)}$$

Bounding Frobenius Norm

Let $C = AB$, then

$$\begin{aligned} \|AB\|_F^2 &= \sum_{i=1}^{n}\sum_{j=1}^{m} |c_{ij}|^2 \\ &\le \sum_{i=1}^{n}\sum_{j=1}^{m} \left(\|a_i\|_2 \, \|b_j\|_2\right)^2 \\ &= \sum_{i=1}^{n} \left(\|a_i\|_2\right)^2 \sum_{j=1}^{m} \left(\|b_j\|_2\right)^2 \quad = \quad \|A\|_F^2 \, \|B\|_F^2 . \end{aligned}$$

Invariance under Unitary Multiplication

The matrix 2-norm and Frobenius norm are invariant under multiplication by unitary matrices.

$$||QA||_2 = ||A||_2, \qquad ||QA||_F = ||A||_F.$$

This fact is still valid if Q is generalized to a rectangular matrix with orthonormal columns. Recall transformation used in PCA.