Numerical Matrix Analysis

Lecture Notes #3 — Orthogonal Vectors, Matrices and Norms

Peter Blomgren,
⟨blomgren.peter@gmail.com⟩

Department of Mathematics and Statistics
Dynamical Systems Group
Computational Sciences Research Center
San Diego State University
San Diego, CA 92182-7720

http://terminus.sdsu.edu/

Spring 2020

Outline

1. Introduction
   - Recap
2. Fundamental Concepts
   - Adjoint / Hermitian
   - Inner Products, Matrix Properties, Orthogonality
   - Unitary Matrices, Vector Norms, Matrix Norms
3. Next...
   - Looking Ahead

Previously...

A quick review / crash course in basic linear algebra:

- Vectors: Transpose, Addition & Subtraction
- Matrix-Vector Product
- The Vandermonde Matrix ... and Linear Least Squares Problems
- Matrix-Matrix Product
- The Transpose of a Matrix (A^T)
- The Range and Nullspace of a Matrix A
- The Rank of a Matrix $A_{m\times n}$
- The Inverse of a Matrix A

Now... ...More Fundamental Concepts

The **Adjoint** a.k.a **Hermitian** (Transpose, or Conjugate) of a matrix $A \in \mathbb{C}^{m \times n}$...

For a scalar $z \in \mathbb{C}$, $z = a + bi$, the **complex conjugate** $\bar{z}$, or z^* is obtained by negating the imaginary part, *i.e.* $z^* = a - bi$.

Note that if $z \in \mathbb{R}$, then $z^* = z$.

For a matrix $A \in \mathbb{C}^{m \times n}$, the Hermitian Conjugate $A^* \in \mathbb{C}^{n \times m}$ is the matrix

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \\ a_{41} & a_{42} \end{bmatrix} \Rightarrow \mathbf{A}^* = \begin{bmatrix} a_{11}^* & a_{21}^* & a_{31}^* & a_{41}^* \\ a_{12}^* & a_{22}^* & a_{32}^* & a_{42}^* \end{bmatrix}$$

The Hermitian Conjugate

If $A = A^*$, the matrix A is said to be **Hermitian**.

Note that a **Hermitian matrix must be square.**

In the case that A is real-valued, *i.e.* $A \in \mathbb{R}^{m \times n}$, then $A = A^* = A^T$ (the Hermitian conjugate equals the **transpose**).

If $A = A^T$, the matrix A is said to be **Symmetric**.

Our book (TREFETHEN-BAU) tends to state results and theorems in terms of complex vectors and matrices, and hence use the Hermitian conjugate, *i.e.* $\vec{x}^*$ is a row-vector.

If this is disturbing to you, just imagine that all quantities are real, and that $^* \equiv {}^T$.

The advantage of this approach is that we are able to state the most general results.

The Inner Product of Two Vectors — a.k.a the **dot product**

The **inner product**, denoted $\langle \vec{x}, \vec{y} \rangle$, of two column vectors $\vec{x},\ \vec{y} \in \mathbb{C}^m$ is defined

$$\langle \vec{x}, \vec{y} \rangle = \vec{x}^* \vec{y} = \sum_{i=1}^{m} x_i^* y_i$$

note that the inner product is a scalar quantity.

The **Euclidean length**, $\|\vec{x}\|$, of $\vec{x} \in \mathbb{C}^m$ is defined

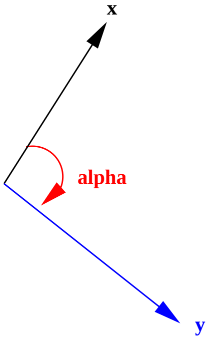
$$\|\vec{x}\| = \sqrt{\langle \vec{x}, \vec{x} \rangle} = \sqrt{\vec{x}^* \vec{x}} = \sqrt{\sum_{i=1}^{m} |x_i|^2}$$

Inner Product: Geometrical Interpretation

The inner product can also be written

$$\langle \vec{x}, \vec{y} \rangle = \vec{x}^* \vec{y} = \|\vec{x}\| \cdot \|\vec{y}\| \cdot \cos(\alpha)$$

where α is the angle between $\vec{x}$ and $\vec{y}$



Inner Product: Properties — Bi-Linearity

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

$$\begin{array}{lrcl} (1) & (\vec{x}_1 + \vec{x}_2)^* \, \vec{y} & = & \vec{x}_1^* \, \vec{y} + \vec{x}_2^* \, \vec{y} \\ (2) & \vec{x}^* \, (\vec{y}_1 + \vec{y}_2) & = & \vec{x}^* \, \vec{y}_1 + \vec{x}^* \, \vec{y}_2 \\ (3) & (\alpha\vec{x})^*(\beta\vec{y}) & = & \alpha^* \beta \, \vec{x}^* \, \vec{y} \end{array}$$

where $\vec{x}$, $\vec{x}_1$, $\vec{x}_2$, $\vec{y}$, $\vec{y}_1$, $\vec{y}_2 \in \mathbb{C}^m$, and α, $\beta \in \mathbb{C}$.

Associated Matrix Properties

For any two matrices A and B, of compatible dimensions, *i.e.* $A \in \mathbb{C}^{m \times n}$, and $B \in \mathbb{C}^{n \times k}$ the following holds

$$(AB)^* = B^* A^*$$

If the matrices A and B are square, and invertible, the following holds

$$(AB)^{-1} = B^{-1} A^{-1}$$

When necessary, we use the notation A^{-*} for $(A^*)^{-1} \equiv (A^{-1})^*$.

Orthogonal and Orthonormal Vectors

Two vectors are **orthogonal** if and only if $\langle \vec{x}, \vec{y} \rangle = \vec{x}^* \vec{y} = 0$,

$$0 = \frac{\vec{x}^* \vec{y}}{\|\vec{x}\| \cdot \|\vec{y}\|} = \cos(\alpha) \;\Leftrightarrow\; \alpha = \pi/2 + k \cdot \pi.$$

A **set** of **non-zero** vectors S is **orthogonal** if its elements are pairwise orthogonal, *i.e.*

$$\forall\, \vec{x}, \vec{y} \in S, \quad \vec{x} \neq \vec{y} \quad \Rightarrow \quad \vec{x}^* \vec{y} = 0$$

A **set** of vectors S is **orthonormal** if it is **orthogonal**, and $\forall \vec{x} \in S$, $\|\vec{x}\| = 1$.

Linear Independence of Orthogonal Set

Theorem (Linear Independence)

The vectors in an orthogonal set S are linearly independent.

Proof (Linear Independence of Orthogonal Vectors).

$\square$

Linear Independence of Orthogonal Set

Theorem (Linear Independence)

The vectors in an orthogonal set S are linearly independent.

Proof (Linear Independence of Orthogonal Vectors).

If the vectors in S are not independent, then $\exists \vec{v}_k \in S : \vec{v}_k \neq \vec{0}$, so that

$$\vec{v}_k = \sum_{i \neq k} c_i \vec{v}_i.$$

□

Linear Independence of Orthogonal Set

Theorem (Linear Independence)

The vectors in an orthogonal set S are linearly independent.

Proof (Linear Independence of Orthogonal Vectors).

If the vectors in S are not independent, then $\exists \vec{v}_k \in S : \vec{v}_k \neq \vec{0}$, so that

$$\vec{v}_k = \sum_{i \neq k} c_i \vec{v}_i.$$

Since $\vec{v}_k \neq 0$, $\langle \vec{v}_k, \vec{v}_k \rangle > 0$,

$\square$

Linear Independence of Orthogonal Set

Theorem (Linear Independence)

The vectors in an orthogonal set S are linearly independent.

Proof (Linear Independence of Orthogonal Vectors).

If the vectors in S are not independent, then $\exists \vec{v}_k \in S : \vec{v}_k \neq \vec{0}$, so that

$$\vec{v}_k = \sum_{i \neq k} c_i \vec{v}_i.$$

Since $\vec{v}_k \neq 0$, $\langle \vec{v}_k, \vec{v}_k \rangle > 0$, now we use the bi-linearity property of inner products, and the orthogonality of S:

$\square$

Linear Independence of Orthogonal Set

Theorem (Linear Independence)

The vectors in an orthogonal set S are linearly independent.

Proof (Linear Independence of Orthogonal Vectors).

If the vectors in S are not independent, then $\exists \vec{v}_k \in S : \vec{v}_k \neq \vec{0}$, so that

$$\vec{v}_k = \sum_{i \neq k} c_i \vec{v}_i.$$

Since $\vec{v}_k \neq 0$, $\langle \vec{v}_k, \vec{v}_k \rangle > 0$, now we use the bi-linearity property of inner products, and the orthogonality of S:

$$0 < \langle \vec{v}_k, \vec{v}_k \rangle = \left\langle \vec{v}_k, \sum_{i \neq k} c_i \vec{v}_i \right\rangle = \sum_{i \neq k} c_i \underbrace{\langle \vec{v}_k, \vec{v}_i \rangle}_{0,\, \forall i \neq k} = 0.$$

$\square$

Linear Independence of Orthogonal Set

Theorem (Linear Independence)

The vectors in an orthogonal set S are linearly independent.

Proof (Linear Independence of Orthogonal Vectors).

If the vectors in S are not independent, then $\exists \vec{v}_k \in S : \vec{v}_k \neq \vec{0}$, so that

$$\vec{v}_k = \sum_{i \neq k} c_i \vec{v}_i.$$

Since $\vec{v}_k \neq 0$, $\langle \vec{v}_k, \vec{v}_k \rangle > 0$, now we use the bi-linearity property of inner products, and the orthogonality of S:

$$0 < \langle \vec{v}_k, \vec{v}_k \rangle = \left\langle \vec{v}_k, \sum_{i \neq k} c_i \vec{v}_i \right\rangle = \sum_{i \neq k} c_i \underbrace{\langle \vec{v}_k, \vec{v}_i \rangle}_{0,\, \forall i \neq k} = 0.$$

This contradicts the assumption that the vectors are linearly dependent, hence proving the theorem. $\square$

Corollary: Basis for $\mathbb{C}^m$

Corollary

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

I.e. we can write any vector $\vec{v} \in \mathbb{C}^m$ as a unique linear combination

$$\vec{v} = \sum_{i=1}^{m} a_i \vec{s}_i, \quad \text{where} \quad a_i = \frac{\langle \vec{s}_i, \vec{v} \rangle}{\|\vec{s}_i\|^2}.$$

We can view the computation of a_i as a **projection** of the vector $\vec{v}$ onto the direction $\vec{s}_i$.

We can use this in order to decompose arbitrary vectors into orthogonal components...

Vector Components

Suppose we have an **orthonormal set** of vectors $\{\vec{q}_1, \vec{q}_2, \ldots, \vec{q}_n\}$, $\vec{q}_i \in \mathbb{C}^m$, $n \leq m$.

Now, for any vector $\vec{v} \in \mathbb{C}^m$, the vector

$$\vec{r} = \vec{v} - \sum_{i=1}^{n} \langle \vec{q}_i, \vec{v} \rangle \vec{q}_i$$

is orthogonal to $\{\vec{q}_1, \vec{q}_2, \ldots, \vec{q}_n\}$:

$$\langle \vec{q}_k, \vec{r} \rangle = \langle \vec{q}_k, \vec{v} \rangle - \underbrace{\sum_{i=1}^{n} \langle \vec{q}_i, \vec{v} \rangle \langle \vec{q}_k, \vec{q}_i \rangle}_{\langle \vec{q}_k, \vec{v} \rangle \underbrace{\langle \vec{q}_k, \vec{q}_k \rangle}_{1}} = 0.$$

Vector Components

We see that by applying this procedure, we have decomposed the vector $\vec{v}$ into $n+1$ orthogonal components:

$$\vec{v} = \vec{r} + \sum_{i=1}^{n} \langle\, \vec{q}_i,\, \vec{v}\,\rangle \vec{q}_i$$

If $\{\vec{q}_i\}$ is a basis for $\mathbb{C}^m$, then $n = m$ and $\vec{r} = \vec{0}$, *i.e.*

$$\vec{v} = \sum_{i=1}^{n} \langle\, \vec{q}_i,\, \vec{v}\,\rangle \vec{q}_i = \sum_{i=1}^{n} (\vec{q}_i^{\,*} \vec{v}) \vec{q}_i = \sum_{i=1}^{n} \vec{q}_i (\vec{q}_i^{\,*} \vec{v}) = \sum_{i=1}^{n} (\vec{q}_i \vec{q}_i^{\,*}) \vec{v}$$

Vector Components

$$\vec{v} = \sum_{i=1}^{n} \langle \vec{q}_i, \vec{v} \rangle \vec{q}_i = \sum_{i=1}^{n} (\vec{q}_i^* \vec{v}) \vec{q}_i = \sum_{i=1}^{n} \vec{q}_i (\vec{q}_i^* \vec{v}) = \sum_{i=1}^{n} (\vec{q}_i \vec{q}_i^*) \vec{v}$$

In the expression $\vec{v} = \sum_{i=1}^{n} (\vec{q}_i^* \vec{v}) \vec{q}_i$ we view $\vec{v}$ as a sum of coefficients (circled) times vectors $\vec{q}_i$, whereas in the equivalent expression $\vec{v} = \sum_{i=1}^{n} (\vec{q}_i \vec{q}_i^*) \vec{v}$, we view $\vec{v}$ as a sum of **orthogonal projections** onto the various directions $\vec{q}_i$.

We will return to the issue of projection matrices of the form $\vec{q}_i \vec{q}_i^*$ soon.

Unitary Matrices

A square matrix $Q \in \mathbb{C}^{m \times m}$ is **unitary** (in the real case "orthogonal") if

$$Q^* = Q^{-1} \quad \Leftrightarrow \quad Q^* Q = I$$

In terms of the columns, $\vec{q}_i$ of Q this looks like

$$\begin{bmatrix} \text{---} & \vec{q}_1^{\,*} & \text{---} \\ \text{---} & \vec{q}_2^{\,*} & \text{---} \\ & \vdots & \\ \text{---} & \vec{q}_n^{\,*} & \text{---} \end{bmatrix} \begin{bmatrix} | & | & & | \\ \vec{q}_1 & \vec{q}_2 & \cdots & \vec{q}_n \\ | & | & & | \end{bmatrix} = \begin{bmatrix} 1 & & & \\ & 1 & & \\ & & \ddots & \\ & & & 1 \end{bmatrix}$$

We have $\vec{q}_i^{\,*} \vec{q}_j = \delta_{ij}$, the **Kronecker delta**, equal to 1 if and only if $i = j$, and 0 otherwise.

Multiplication by a Unitary Matrix

Since the norm of the columns of a unitary matrix is 1, multiplication by a unitary matrix preserves the Euclidean norm in the following sense:

For a unitary Q:

$$(1) \qquad \langle\, Q\vec{x},\, Q\vec{y}\,\rangle = (Q\vec{x})^*(Q\vec{y}) = \vec{x}^*\, \underbrace{Q^*\, Q}_{I}\, \vec{y} = \vec{x}^*\vec{y} = \langle\, \vec{x},\, \vec{y}\,\rangle$$

$$(2) \qquad \|Q\vec{x}\| = \|\vec{x}\|$$

The invariance of inner products mean that angles between vectors are preserved.

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

Vector Norms

Norms give us the essential notion of size and distance in a vector space — these are our tools for measuring the quality of approximations and convergence in our algorithms.

Definition (Norm)

A **norm** is a function $\|\cdot\| : \mathbb{C}^m \to \mathbb{R}$ that assigns a real-valued (length) to each vector. A norm must satisfy the following three conditions for all vectors $\vec{x}, \vec{y} \in \mathbb{C}^m$, and scalars $\alpha \in \mathbb{C}$,

$$\begin{array}{ll} (1) & \|\vec{x}\| \geq 0, \text{ and } \|\vec{x}\| = 0 \text{ only if } \vec{x} = 0 \\ (2) & \|\vec{x}+\vec{y}\| \leq \|\vec{x}\| + \|\vec{y}\| \\ (3) & \|\alpha\vec{x}\| = |\alpha| \, \|\vec{x}\| \end{array}$$

(2) is known as the **"triangle inequality."**

The p-norms

The p-norms (sometimes referred to as the ℓ_p-norms), parametrized by p are defined by

$$\|\vec{x}\|_p = \left[\sum_{i=1}^{m} |x_i|^p\right]^{1/p}$$

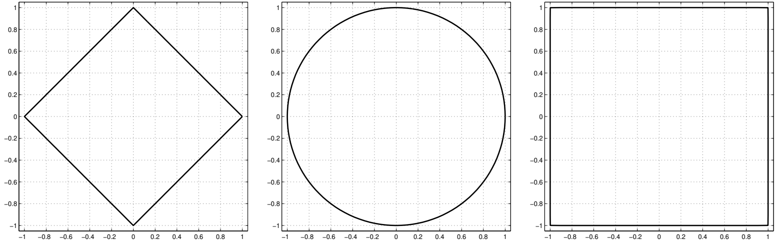
As an illustration, the **unit sphere** $\|\vec{x}\|_p = 1$, $\vec{x} \in \mathbb{R}^2$ is illustrated for some common (and uncommon) p-norms, on the following slides.

The 2-norm is the standard Euclidean length function.

The 1-norm is sometimes referred to as the Manhattan/taxicab-distance.

The p-norms

Some commonly used p-norms

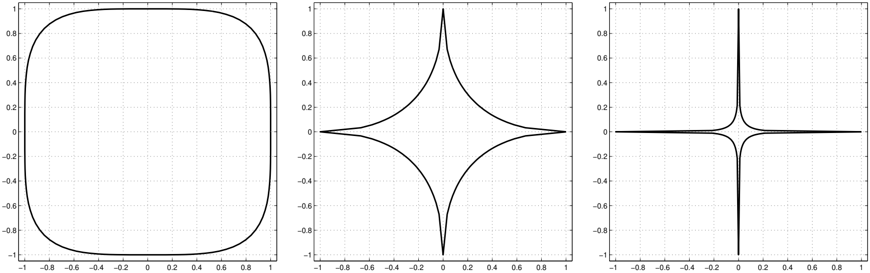


$$\|\vec{x}\|_1 = \sum_{i=1}^{m} |x_i|, \quad \|\vec{x}\|_2 = \left[\sum_{i=1}^{m} |x_i|^2\right]^{1/2}, \quad \|\vec{x}\|_\infty = \max_{i=1\ldots m} |x_i|$$

San Diego State University

The p-norms

Some exotic p-{norms,non-norms}



$$\|\vec{x}\|_4 = \left[\sum_{i=1}^{m} |x_i|^4\right]^{1/4}, \quad \|\vec{x}\|_{1/2} = \left[\sum_{i=1}^{m} |x_i|^{1/2}\right]^{2}, \quad \|\vec{x}\|_{1/4} = \left[\sum_{i=1}^{m} |x_i|^{1/4}\right]^{4}$$

Note: when $p < 1$ the "norms" are not convex; which means the triangle inequality will not hold; and strictly speaking these are not norms...

$\exists$ **Movie**.

Weighted p-norms

The **weighted p-norms** $\|\cdot\|_{W,p}$ are derived from the p-norms:

$$\|\vec{x}\|_{W,p} = \|W\vec{x}\|_p$$

where W is *e.g.* a diagonal matrix, in which the ith diagonal entry is the weight $w_i \neq 0$:

$$\|\vec{x}\|_{W,p} = \left[\sum_{i=1}^{m} |w_i x_i|^p\right]^{1/p}$$

Weighted p-norms

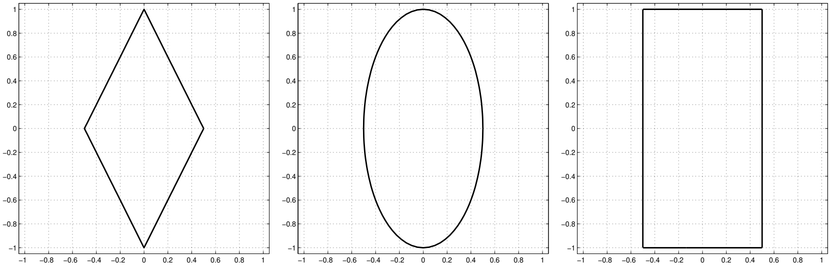


Figure: Visualization of the unit-sphere for the weighted 1-, 2- and ∞-norms, where $W = \mathrm{diag}(2, 1)$.

The concept of weighted p-norms can be generalized to arbitrary non-singular weight matrices W.

Weighted p-norms

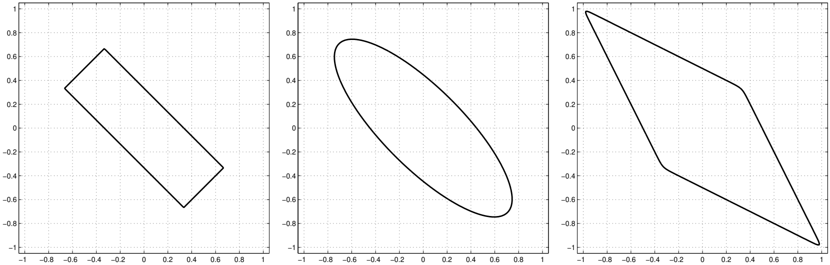


Figure: Visualization of the unit-sphere for the weighted 1-, 2- and ∞-norms, where $W = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$.

$\exists$ Movie.

Matrix Norms — Induced by Vector Norms

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

$$\|A\|_{(m,n)} = \sup_{\vec{x}\in\mathbb{C}^n-\{\vec{0}\}} \left[\frac{\|A\vec{x}\|_{(m)}}{\|\vec{x}\|_{(n)}}\right]$$

In any sane application, both $\|\cdot\|_{(m)}$ and $\|\cdot\|_{(n)}$ will be of the same type, *i.e.* the p-norms (with the same p).

Due to the linearity of norms — the third norm-condition — it is sufficient to maximize the matrix norm over $\vec{x} \in \mathbb{C}^n : \|\vec{x}\| = 1\ldots$

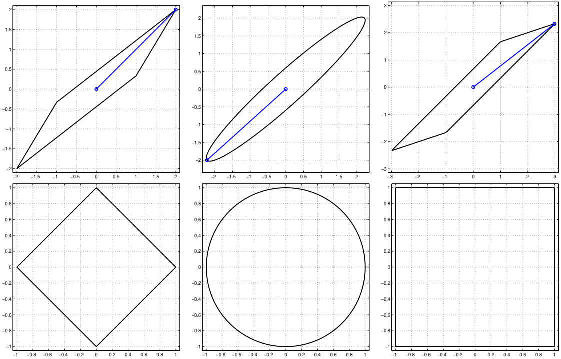
Most of the time the norms with $p = 2$ are used. Indeed, if nothing else is specified, this is usually implied.

SAN DIEGO STATE UNIVERSITY

Illustration: Matrix Norms

$$A = \begin{bmatrix} 1 & 2 \\ 1/3 & 2 \end{bmatrix}, \qquad \begin{aligned} \lambda(A) &= \{2.45743, 0.54257\} \\ \sigma(A) &= \{2.98523, 0.44664\} \end{aligned}$$

eigenvalues
singular values



$\|A\|_1 = 4$ $\|A\|_2 \approx 2.9852$ $\|A\|_\infty = 3$

Special Cases: Matrix p-norms

If D is a diagonal matrix, then

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

The 1-norm of a matrix is the maximal column-abs-sum:

$$\|A\|_1 = \max_{1 \le j \le n} \|\vec{a}_j\|_1$$

The ∞-norm of a matrix is the maximal row-abs-sum:

$$\|A\|_\infty = \max_{1 \le i \le m} \|\vec{a}_i^*\|_1$$

Next Time

- Finish up the discussion on norms:
  - Inequalities, General matrix norms, The Frobenius norm, Bounds on norms of products of matrices.
- The Singular Value Decomposition (SVD).