San José State University

Math 253: Mathematical Methods for Data Visualization

Lecture 1: Review of Linear Algebra and Multivariable Calculus

Dr. Guangliang Chen

Outline

- **Matrix algebra**
  - Multiplication
  - Rank
  - Trace
  - Determinant
  - Eigenvalues and eigenvectors
  - Diagonalization of square matrices
- **Constrained optimization** (with equality constraints)

Notation: vectors

Vectors are denoted by **boldface** lowercase letters (such as $\mathbf{a}, \mathbf{b}, \mathbf{u}, \mathbf{v}, \mathbf{x}, \mathbf{y}$):

- They are assumed to be in column form, e.g., $\mathbf{a} = (1, 2, 3)^T = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$
- To indicate their dimensions, we use notation like $\mathbf{x} \in \mathbb{R}^n$.
- The ith element of a vector $\mathbf{a}$ is written as a_i or $\mathbf{a}(i)$.

We introduce the following notation for denoting two constant vectors (with their dimensions implied by the context):

$$\mathbf{0} = [0, 0, \ldots, 0]^T, \qquad \mathbf{1} = [1, 1, \ldots, 1]^T$$

Notation: matrices

Matrices are denoted by **boldface** UPPERCASE letters (such as $\mathbf{A}, \mathbf{B}, \mathbf{U}, \mathbf{V}, \mathbf{P}, \mathbf{Q}$).

Similarly, we write $\mathbf{A} \in \mathbb{R}^{m \times n}$ to indicate its size.

The (i, j) entry of $\mathbf{A}$ is denoted by a_{ij} or $\mathbf{A}(i, j)$.

The ith row of $\mathbf{A}$ is denoted by $\mathbf{A}(i, :)$ while its columns are written as $\mathbf{A}(:, j)$, as in MATLAB.

We use $\mathbf{I}$ to denote the identity matrix, and $\mathbf{O}$ the zero matrix, with their sizes implied by the context.

Description of matrix shape

A matrix is a rectangular array of numbers arranged in rows and columns.

We say that a matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$ is

- a **square** matrix, if $m = n$.
- a **long** matrix, if $m < n$
- a **tall** matrix, if $m > n$

A **diagonal** matrix is a square matrix $\mathbf{A} \in \mathbb{R}^{n\times n}$ whose off diagonal entries are all zero ($a_{ij} = 0$ for all $i \neq j$):

$$\mathbf{A} = \begin{pmatrix} a_{11} & & \\ & \ddots & \\ & & a_{nn} \end{pmatrix}$$

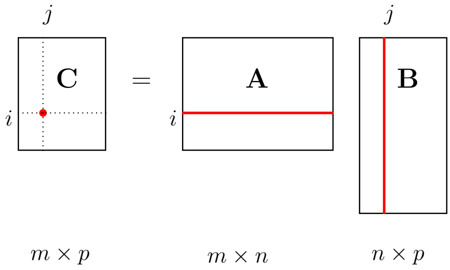
A diagonal matrix is uniquely defined through a vector that contains all the diagonal entries, and denoted as follows:

$$\mathbf{A} = \mathrm{diag}(\underbrace{1, 2, 3, 4, 5}_{\mathbf{a}}) = \mathrm{diag}(\mathbf{a}).$$

Matrix multiplication

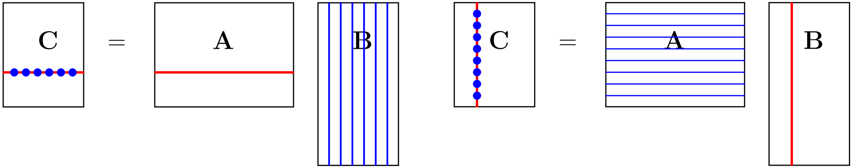
Let $\mathbf{A} \in \mathbb{R}^{m\times n}$ and $\mathbf{B} \in \mathbb{R}^{n\times p}$. Their matrix product is an $m \times p$ matrix

$$\mathbf{A}\mathbf{B} = \mathbf{C} = (c_{ij}), \quad c_{ij} = \sum_{k=1}^{n} a_{ik} b_{kj} = \mathbf{A}(i,:) \cdot \mathbf{B}(:,j).$$



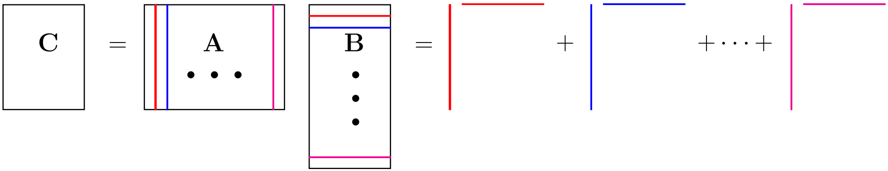
It is possible to obtain one full row (or column) of $\mathbf{C}$ at a time via matrix-vector multiplication:

$$\mathbf{C}(i,:) = \mathbf{A}(i,:) \cdot \mathbf{B}, \qquad \mathbf{C}(:,j) = \mathbf{A} \cdot \mathbf{B}(:,j)$$



The full matrix $\mathbf{C}$ can be written as a sum of rank-1 matrices:

$$\mathbf{C} = \sum_{k=1}^{n} \mathbf{A}(:,k) \cdot \mathbf{B}(k,:).$$



When one of the matrices is a diagonal matrix, we have the following rules:

$$\underbrace{\mathbf{A}}_{\text{diagonal}}\mathbf{B} = \begin{pmatrix} a_1 & & \\ & \ddots & \\ & & a_n \end{pmatrix} \begin{pmatrix} \mathbf{B}(1,:) \\ \vdots \\ \mathbf{B}(n,:) \end{pmatrix} = \begin{pmatrix} a_1\mathbf{B}(1,:) \\ \vdots \\ a_n\mathbf{B}(n,:) \end{pmatrix}$$

$$\begin{aligned} \mathbf{A}\underbrace{\mathbf{B}}_{\text{diagonal}} &= [\mathbf{A}(:,1)\ldots\mathbf{A}(:,n)] \begin{pmatrix} b_1 & & \\ & \ddots & \\ & & b_n \end{pmatrix} \\ &= [b_1\mathbf{A}(:,1)\ldots b_n\mathbf{A}(:,n)] \end{aligned}$$

Finally, below are some identities involving the vector $\mathbf{1} \in \mathbb{R}^n$:

$$\mathbf{1}\mathbf{1}^T = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} \begin{pmatrix} 1 & 1 & \dots & 1 \end{pmatrix} = \begin{pmatrix} 1 & 1 & \dots & 1 \\ 1 & 1 & \dots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \dots & 1 \end{pmatrix},$$

$$\mathbf{1}^T\mathbf{1} = n,$$

$$\mathbf{A}\mathbf{1} = \sum_j \mathbf{A}(:,j), \qquad \text{(vector of row sums)}$$

$$\mathbf{1}^T\mathbf{A} = \sum_i \mathbf{A}(i,:), \qquad \text{(horizontal vector of column sums)}$$

$$\mathbf{1}^T\mathbf{A}\mathbf{1} = \sum_i \sum_j \mathbf{A}(i,j) \qquad \text{(total sum of all entries)}$$

Example 0.1. Let

$$\mathbf{A} = \begin{pmatrix} 3 & 0 & 0 \\ 5 & 1 & -1 \\ -2 & 2 & 4 \end{pmatrix}, \ \mathbf{B} = \begin{pmatrix} 1 & 0 \\ 0 & -1 \\ 2 & 3 \end{pmatrix}, \ \mathbf{\Lambda}_1 = \begin{pmatrix} 1 & & \\ & 0 & \\ & & -1 \end{pmatrix}, \ \mathbf{\Lambda}_2 = \begin{pmatrix} 2 & \\ & -3 \end{pmatrix}.$$

Find the products $\mathbf{AB}, \mathbf{\Lambda}_1\mathbf{B}, \mathbf{B}\mathbf{\Lambda}_2, \mathbf{1}^T\mathbf{B}, \mathbf{B1}$ and verify the above rules.

The Hadamard product

Another way to multiply two matrices of the same size, say $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{m\times n}$, is through the Hadamard product, also called the entrywise product:

$$\mathbf{C} = \mathbf{A} \circ \mathbf{B} \in \mathbb{R}^{m\times n}, \qquad \text{with} \quad c_{ij} = a_{ij} b_{ij}.$$

For example,

$$\begin{pmatrix} 0 & 2 & -3 \\ -1 & 0 & -4 \end{pmatrix} \circ \begin{pmatrix} 1 & 0 & -3 \\ 2 & 1 & -1 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 9 \\ -2 & 0 & 4 \end{pmatrix}.$$

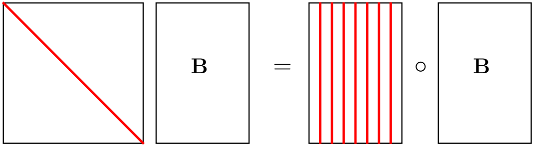
An important application of the entrywise product is in efficiently computing the product of a diagonal matrix and a rectangular matrix by software.

Let $\mathbf{A} = \text{diag}(a_1, \ldots, a_n) \in \mathbb{R}^{n\times n}$ and $\mathbf{B} \in \mathbb{R}^{n\times k}$. Define also a vector $\mathbf{a} = (a_1, \ldots, a_n)^T \in \mathbb{R}^n$, which represents the diagonal of $\mathbf{A}$.

Then

$$\mathbf{AB} = \underbrace{[\mathbf{a} \ldots \mathbf{a}]}_{k \text{ copies}} \circ \mathbf{B}.$$

The former takes $\mathcal{O}(n^2k)$ operations, while the latter takes only $\mathcal{O}(nk)$ operations, which is one magnitude faster.



Matrix transpose

The transpose of a matrix $\mathbf{A} \in \mathbb{R}^{m\times n}$ is another matrix $\mathbf{B} \in \mathbb{R}^{n\times m}$ with $b_{ij} = a_{ji}$ for all i, j. We denote the transpose of $\mathbf{A}$ by $\mathbf{A}^T$.

A square matrix $\mathbf{A} \in \mathbb{R}^{n\times n}$ is said to be **symmetric** if $\mathbf{A}^T = \mathbf{A}$.

Let $\mathbf{A} \in \mathbb{R}^{m\times n}, \mathbf{B} \in \mathbb{R}^{n\times p}$, and $k \in \mathbb{R}$. Then

- $(\mathbf{A}^T)^T = \mathbf{A}$
- $(k\mathbf{A})^T = k\mathbf{A}^T$
- $(\mathbf{A}+\mathbf{B})^T = \mathbf{A}^T + \mathbf{B}^T$
- $(\mathbf{A}\mathbf{B})^T = \mathbf{B}^T\mathbf{A}^T$

Matrix inverse

A square matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$ is said to be invertible if there exists another square matrix of the same size $\mathbf{B}$ such that $\mathbf{AB} = \mathbf{BA} = \mathbf{I}$.

In this case, $\mathbf{B}$ is called the matrix inverse of $\mathbf{A}$ and denoted as $\mathbf{B} = \mathbf{A}^{-1}$.

Let $\mathbf{A}, \mathbf{B}$ be two invertible matrices of the same size, and $k \neq 0$. Then

- $(k\mathbf{A})^{-1} = \frac{1}{k}\mathbf{A}^{-1}$
- $(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}$
- $(\mathbf{A}^T)^{-1} = (\mathbf{A}^{-1})^T$

(Note that $\mathbf{A} + \mathbf{B}$ is not necessarily still invertible)

Matrix trace

The **trace** of a square matrix $\mathbf{A} \in \mathbb{R}^{n\times n}$ is defined as the sum of the entries in its diagonal:

$$\operatorname{trace}(\mathbf{A}) = \sum_i a_{ii}.$$

Clearly, $\operatorname{trace}(\mathbf{A}) = \operatorname{trace}(\mathbf{A}^T)$.

Trace is a **linear** operator: $\operatorname{trace}(k\mathbf{A}) = k \operatorname{trace}(A)$ and $\operatorname{trace}(\mathbf{A}+\mathbf{B}) = \operatorname{trace}(\mathbf{A}) + \operatorname{trace}(\mathbf{B})$.

If $\mathbf{A}$ is an $m \times n$ matrix and $\mathbf{B}$ is an $n \times m$ matrix, then

$$\boxed{\operatorname{trace}(\mathbf{AB}) = \operatorname{trace}(\mathbf{BA})}$$

Note that as matrices, $\mathbf{AB}$ is not necessarily equal to $\mathbf{BA}$.

Matrix rank

Let $\mathbf{A} \in \mathbb{R}^{m\times n}$. The largest number of linearly independent rows (or columns) contained in the matrix is called the rank of $\mathbf{A}$, and often denoted as $\text{rank}(\mathbf{A})$.

A square matrix $\mathbf{P} \in \mathbb{R}^{n\times n}$ is said to be **of full rank** (or nonsingular) if $\text{rank}(\mathbf{P}) = n$; otherwise, it is said to be **rank deficient** (or singular).

A rectangular matrix $\mathbf{A} \in \mathbb{R}^{m\times n}$ is said to have **full column rank** if $\text{rank}(\mathbf{B}) = n$.

Similarly, a rectangular matrix $\mathbf{A} \in \mathbb{R}^{m\times n}$ is said to have **full row rank** if $\text{rank}(\mathbf{B}) = m$.

Some useful properties of the matrix rank:

- For any $\mathbf{A} \in \mathbb{R}^{m \times n}$, $0 \leq \operatorname{rank}(\mathbf{A}) = \operatorname{rank}(\mathbf{A}^T) \leq \min(m, n)$, and $\operatorname{rank}(\mathbf{A}) = 0$ if and only if $\mathbf{A} = \mathbf{O}$.
- Any nonzero row or column vector has rank 1 (as a matrix).
- For any (column) vectors $\mathbf{u}, \mathbf{v}$, $\operatorname{rank}(\mathbf{u}\mathbf{v}^T) = 1$.
- For any two matrices $\mathbf{A} \in \mathbb{R}^{m \times n}, \mathbf{B} \in \mathbb{R}^{n \times p}$,
$$\operatorname{rank}(\mathbf{AB}) \leq \min(\operatorname{rank}(\mathbf{A}), \operatorname{rank}(\mathbf{B})).$$
- For any $\mathbf{A} \in \mathbb{R}^{m \times n}$ and square, nonsingular $\mathbf{P} \in \mathbb{R}^{m \times m}, \mathbf{Q} \in \mathbb{R}^{n \times n}$,
$$\operatorname{rank}(\mathbf{PA}) = \operatorname{rank}(\mathbf{A}) = \operatorname{rank}(\mathbf{AQ}).$$

Matrix determinant

The matrix determinant is a rule to evaluate square matrices to numbers (in order to determine if they are nonsingular):

$$\det : \mathbf{A} \in \mathbb{R}^{n\times n} \mapsto \det(\mathbf{A}) \in \mathbb{R}.$$

A remarkable property is that $\mathbf{A} \in \mathbb{R}^{n\times n}$ is invertible or nonsingular (i.e., of full rank) if and only if $\det(\mathbf{A}) \neq 0$.

Let $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{n\times n}$ and $k \in \mathbb{R}$. Then

- $\det(k\mathbf{A}) = k^n \det(\mathbf{A})$
- $\det(\mathbf{A}^T) = \det(\mathbf{A})$
- $\det(\mathbf{AB}) = \det(\mathbf{A}) \det(\mathbf{B})$

Example 0.2. For the matrix

$$\mathbf{A} = \begin{pmatrix} 3 & 0 & 0 \\ 5 & 1 & -1 \\ -2 & 2 & 4 \end{pmatrix},$$

find its rank, trace and determinant.

Answer. $\text{rank}(\mathbf{A}) = 3, \ \text{trace}(\mathbf{A}) = 8, \ \det(\mathbf{A}) = 18$

Eigenvalues and eigenvectors

Let $\mathbf{A} \in \mathbb{R}^{n\times n}$. The **characteristic polynomial** of $\mathbf{A}$ is

$$p(\lambda) = \det(\mathbf{A} - \lambda\mathbf{I}).$$

The roots of the characteristic equation $p(\lambda) = 0$ are called **eigenvalues** of $\mathbf{A}$.

For a specific eigenvalue λ_i, any nonzero vector $\mathbf{v}_i$ satisfying

$$(\mathbf{A} - \lambda_i\mathbf{I})\mathbf{v}_i = \mathbf{o}$$

or equivalently,

$$\mathbf{A}\mathbf{v}_i = \lambda_i\mathbf{v}_i$$

is called an **eigenvector** of $\mathbf{A}$ (associated to the eigenvalue λ_i).

Example 0.3. For the matrix $\mathbf{A} = \begin{pmatrix} 3 & 0 & 0 \\ 5 & 1 & -1 \\ -2 & 2 & 4 \end{pmatrix}$, find its eigenvalues and associated eigenvectors.

Answer. The eigenvalues are $\lambda_1 = 3, \lambda_2 = 2$ with corresponding eigenvectors $\mathbf{v}_1 = (0, 1, -2)^T, \mathbf{v}_2 = (0, 1, -1)^T$.

All eigenvectors associated to λ_i span a linear subspace, called the **eigenspace**:

$$\mathrm{E}(\lambda_i) = \{\mathbf{v} \in \mathbb{R}^n : (\mathbf{A} - \lambda_i \mathbf{I})\mathbf{v} = \mathbf{o}\}.$$

The dimension g_i of $\mathrm{E}(\lambda_i)$ is called the **geometric multiplicity** of λ_i, while the degree a_i of the factor $(\lambda - \lambda_i)^{a_i}$ in $p(\lambda)$ is called the **algebraic multiplicity** of λ_i.

Note that we must have $\sum a_i = n$ and for all i, $1 \leq g_i \leq a_i$.

Example 0.4. For the eigenvalues of the matrix on previous slide, find their algebraic and geometric multiplicities.

Answer. $a_1 = 2, a_2 = 1$ and $g_1 = g_2 = 1$.

The following theorem indicates that the trace and determinant of a square matrix can both be computed from the eigenvalues of the matrix.

Theorem 0.1. Let $\mathbf{A}$ be a real square matrix whose eigenvalues are $\lambda_1, \ldots, \lambda_n$ (possibly with repetitions). Then

$$\det(\mathbf{A}) = \prod_{i=1}^{n} \lambda_i \qquad \text{and} \qquad \text{trace}(\mathbf{A}) = \sum_{i=1}^{n} \lambda_i.$$

Example 0.5. For the matrix $\mathbf{A}$ defined previously, verify the identities in the above theorem.

Similar matrices

Two square matrices of the same size $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{n\times n}$ if there exists an invertible matrix $\mathbf{P} \in \mathbb{R}^{n\times n}$ such that

$$\mathbf{B} = \mathbf{P}\mathbf{A}\mathbf{P}^{-1}$$

Similar matrices have the same

- rank
- trace
- determinant
- characteristic polynomial
- eigenvalues and their multiplicities (but not eigenvectors)

Diagonalizability of square matrices

Definition 0.1. A square matrix $\mathbf{A}$ is **diagonalizable** if it is similar to a diagonal matrix, i.e., there exist an invertible matrix $\mathbf{P}$ and a diagonal matrix $\mathbf{\Lambda}$ such that

$$\mathbf{A} = \mathbf{P}\mathbf{\Lambda}\mathbf{P}^{-1}, \qquad \text{or equivalently,} \quad \mathbf{P}^{-1}\mathbf{A}\mathbf{P} = \mathbf{\Lambda}.$$

Remark. If we write $\mathbf{P} = (\mathbf{p}_1, \ldots, \mathbf{p}_n)$ and $\mathbf{\Lambda} = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$, then the above equation can be rewritten as

$$\mathbf{A}\mathbf{P} = \mathbf{P}\mathbf{\Lambda}, \qquad \text{or in columns,} \quad \mathbf{A}\mathbf{p}_i = \lambda_i \mathbf{p}_i,\, 1 \le i \le n.$$

This shows that each λ_i is an eigenvalue of $\mathbf{A}$ and $\mathbf{p}_i$ the corresponding eigenvector. Thus, the above factorization is called the **eigenvalue decomposition** of $\mathbf{A}$, or sometimes the **spectral decomposition** of $\mathbf{A}$.

Example 0.6. The matrix

$$\mathbf{A} = \begin{pmatrix} 0 & 1 \\ 3 & 2 \end{pmatrix}$$

is diagonalizable because

$$\begin{pmatrix} 0 & 1 \\ 3 & 2 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 3 & -1 \end{pmatrix} \begin{pmatrix} 3 & \\ & -1 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 3 & -1 \end{pmatrix}^{-1}$$

but the matrix

$$\mathbf{B} = \begin{pmatrix} 0 & 1 \\ -1 & 2 \end{pmatrix}$$

is not.

Why is diagonalization important?

We can use instead the diagonal matrix (that is similar to the given matrix) to compute the determinant and eigenvalues (and their algebraic multiplicities), which is a lot simpler.

Additionally, it can help compute **matrix powers** ($\mathbf{A}^k$). To see this, suppose $\mathbf{A}$ is diagonalizable, that is, $\mathbf{A} = \mathbf{PDP}^{-1}$ for some invertible matrix $\mathbf{P}$ and a diagonal matrix $\mathbf{\Lambda}$. Then

$$\begin{aligned}
\mathbf{A}^2 &= \mathbf{P\Lambda P}^{-1} \cdot \mathbf{P\Lambda P}^{-1} = \mathbf{P\Lambda}^2\mathbf{P}^{-1} \\
\mathbf{A}^3 &= \mathbf{P\Lambda P}^{-1} \cdot \mathbf{P\Lambda P}^{-1} \cdot \mathbf{P\Lambda P}^{-1} = \mathbf{P\Lambda}^3\mathbf{P}^{-1} \\
\mathbf{A}^k &= \mathbf{P\Lambda}^k\mathbf{P}^{-1} \quad \text{(for any positive integer } k\text{)}
\end{aligned}$$

Checking diagonalizability of a square matrix

Theorem 0.2. A matrix $\mathbf{A} \in \mathbb{R}^{n\times n}$ is diagonalizable if and only if it has n linearly independent eigenvectors (i.e., $\sum g_i = n$).

Corollary 0.3. The following matrices are diagonalizable:

- Any matrix whose eigenvalues all have identical geometric and algebraic multiplicities, i.e., $g_i = a_i$ for all i;
- Any matrix with n distinct eigenvalues ($g_i = a_i = 1$ for all i);

Example 0.7. The matrix $\mathbf{B} = \begin{pmatrix} 0 & 1 \\ -1 & 2 \end{pmatrix}$ is not diagonalizable because it has only one distinct eigenvalue $\lambda_1 = 1$ with $a_1 = 2$ and $g_1 = 1$.

Orthogonal matrices

An **orthogonal matrix** is a square matrix $\mathbf{Q} \in \mathbb{R}^{n\times n}$ whose inverse equals its transpose, i.e., $\mathbf{Q}^{-1} = \mathbf{Q}^T$.

In other words, orthogonal matrices $\mathbf{Q}$ satisfy $\mathbf{Q}\mathbf{Q}^T = \mathbf{Q}^T\mathbf{Q} = \mathbf{I}$.

For example, the following are orthogonal matrices:

$$\begin{pmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix}, \quad \begin{pmatrix} \frac{2}{\sqrt{6}} & 0 & \frac{1}{\sqrt{3}} \\ -\frac{1}{\sqrt{6}} & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{3}} \end{pmatrix}$$

Theorem 0.4. A square matrix $\mathbf{Q} = [\mathbf{q}_1 \ldots \mathbf{q}_n]$ is orthogonal if and only if its columns form an orthonormal basis for $\mathbb{R}^n$. That is,

$$\mathbf{q}_i^T \mathbf{q}_j = \begin{cases} 0, & i \neq j \\ 1, & i = j \end{cases}$$

Proof. This is a direct consequence of the following identity

$$\mathbf{Q}^T \mathbf{Q} = \begin{bmatrix} \mathbf{q}_1^T \\ \vdots \\ \mathbf{q}_n^T \end{bmatrix} \begin{bmatrix} \mathbf{q}_1 & \ldots & \mathbf{q}_n \end{bmatrix} = (\mathbf{q}_i^T \mathbf{q}_j)$$

Remark. Geometrically, an orthogonal matrix multiplying a vector (i.e., $\mathbf{Qx} \in \mathbb{R}^n$) represents an rotation of the vector in the space.

Spectral decomposition of symmetric matrices

Theorem 0.5. Let $\mathbf{A} \in \mathbb{R}^{n\times n}$ be a symmetric matrix. Then there exist an orthogonal matrix $\mathbf{Q} = [\mathbf{q}_1 \dots \mathbf{q}_n]$ and a diagonal matrix $\mathbf{\Lambda} = \mathrm{diag}(\lambda_1, \dots, \lambda_n)$, such that

$$\mathbf{A} = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^T \qquad \text{(we say that } \mathbf{A} \text{ is orthogonally diagonalizable)}$$

Remark. Note that the above equation is equivalent to $\mathbf{AQ} = \mathbf{Q\Lambda}$, or in columns,

$$\mathbf{A}\mathbf{q}_i = \lambda_i \mathbf{q}_i, \quad i = 1, \dots, n$$

Therefore, the λ_i's represent eigenvalues of $\mathbf{A}$ while the $\mathbf{q}_i$'s are the associated eigenvectors (with unit norm).

Remark. One can rewrite the matrix decomposition

$$\mathbf{A} = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^T$$

into a sum of rank-1 matrices:

$$\mathbf{A} = \begin{pmatrix} \mathbf{q}_1 & \dots & \mathbf{q}_n \end{pmatrix} \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix} \begin{pmatrix} \mathbf{q}_1^T \\ \vdots \\ \mathbf{q}_n \end{pmatrix} = \sum_{i=1}^n \lambda_i \mathbf{q}_i \mathbf{q}_i^T$$

For convenience, the diagonal elements of $\mathbf{\Lambda}$ are often assumed to be sorted in decreasing order:

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$$

Example 0.8. Find the spectral decomposition of the following matrix

$$\mathbf{A} = \begin{pmatrix} 0 & 2 \\ 2 & 3 \end{pmatrix}$$

Answer.

$$\mathbf{A} = \underbrace{\frac{1}{\sqrt{5}}\begin{pmatrix} 1 & -2 \\ 2 & 1 \end{pmatrix}}_{\mathbf{Q}} \cdot \underbrace{\begin{pmatrix} 4 & \\ & -1 \end{pmatrix}}_{\mathbf{\Lambda}} \cdot \underbrace{\frac{1}{\sqrt{5}}\begin{pmatrix} 1 & -2 \\ 2 & 1 \end{pmatrix}^T}_{\mathbf{Q}^T}$$

$$= 4\begin{pmatrix} \frac{1}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} \end{pmatrix}\begin{pmatrix} \frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} \end{pmatrix} + (-1)\begin{pmatrix} -\frac{2}{\sqrt{5}} \\ \frac{1}{\sqrt{5}} \end{pmatrix}\begin{pmatrix} -\frac{2}{\sqrt{5}} & \frac{1}{\sqrt{5}} \end{pmatrix}$$

Positive (semi)definite matrices

Definition 0.2. A symmetric matrix $\mathbf{A} \in \mathbb{R}^{n\times n}$ is said to be **positive semidefinite** if $\mathbf{x}^T\mathbf{A}\mathbf{x} \geq 0$ for all $\mathbf{x} \in \mathbb{R}^n$.

If the equality holds true only for $\mathbf{x} = \mathbf{0}$ (i.e., $\mathbf{x}^T\mathbf{A}\mathbf{x} > 0$ for all $\mathbf{x} \neq \mathbf{0}$), then $\mathbf{A}$ is said to be **positive definite**.

Example 0.9. For any rectangular matrix $\mathbf{A} \in \mathbb{R}^{m\times n}$, show that both of the matrices $\mathbf{A}\mathbf{A}^T \in \mathbb{R}^{m\times m}$ and $\mathbf{A}^T\mathbf{A} \in \mathbb{R}^{n\times n}$ are positive semidefinite.

Theorem. A symmetric matrix is positive definite (semidefinite) if and only if all of its eigenvalues are positive (nonnegative).

Review of multivariable calculus

First, consider the following **constrained optimization** problem with an equality constraint in $\mathbb{R}^n$ (i.e., $\mathbf{x} \in \mathbb{R}^n$):

$$\mathbf{max} \,/\, \mathbf{min}\ \mathbf{f}(\mathbf{x}) \qquad \text{subject to} \quad \boldsymbol{g(\mathbf{x}) = b}$$

For example,[1] consider

$$\max \,/\, \min\ \underbrace{8x^2 - 2y}_{f(x,y)} \qquad \text{subject to} \quad \underbrace{x^2 + y^2}_{g(x,y)} = \underbrace{1}_{b}$$

which can be interpreted as finding the extreme values of $f(x, y) = 8x^2 - 2y$ over the unit circle in $\mathbb{R}^2$.

[1] http://tutorial.math.lamar.edu/Classes/CalcIII/LagrangeMultipliers.aspx

To solve such problems, one can use the **method of Lagrange multipliers**:

1. Form the Lagrange function (by introducing an extra variable λ)

$$L(\mathbf{x}, \lambda) = f(\mathbf{x}) - \lambda(g(\mathbf{x}) - b)$$

2. Find all critical points of the Lagrangian L by solving

$$\frac{\partial}{\partial \mathbf{x}} L = \mathbf{0} \longrightarrow \nabla f(\mathbf{x}) = \lambda \nabla g(\mathbf{x})$$
$$\frac{\partial}{\partial \lambda} L = 0 \longrightarrow g(\mathbf{x}) = b$$

3. Evaluate and compare the function f at all the critical points to select the maximum and/or minimum.

Now, let us solve

$$\max / \min \underbrace{8x^2 - 2y}_{f(x,y)} \qquad \text{subject to} \quad \underbrace{x^2 + y^2}_{g(x,y)} = \underbrace{1}_{b}$$

By the method of Lagrange multipliers, we have

$$\begin{aligned} \frac{\partial f}{\partial x} = \lambda \frac{\partial g}{\partial x} \longrightarrow &\quad 16x = \lambda(2x) \\ \frac{\partial f}{\partial y} = \lambda \frac{\partial g}{\partial y} \longrightarrow &\quad -2 = \lambda(2y) \\ g(\mathbf{x}) = b \longrightarrow &\quad x^2 + y^2 = 1 \end{aligned}$$

from which we obtain 4 critical points with corresponding function values:

$$f(0,-1) = 2, \quad f(0,1) = \underbrace{-2}_{\min}, \quad f(-\frac{3}{8}\sqrt{7}, -\frac{1}{8}) = \underbrace{\frac{65}{8}}_{\max}$$

Question 1: What if there are multiple equality constraints?

$$\max / \min f(\mathbf{x}) \qquad \text{subject to} \quad g_1(\mathbf{x}) = b_1, \ldots, g_k(\mathbf{x}) = b_k$$

The method of Lagrange multipliers works very similarly:

1. Form the Lagrange function

$$L(\mathbf{x}, \lambda_1, \ldots, \lambda_k) = f(\mathbf{x}) - \lambda_1(g_1(\mathbf{x}) - b_1) - \cdots - \lambda_k(g_k(\mathbf{x}) - b_k)$$

2. Find all critical points by solving

$$\nabla_{\mathbf{x}} L = \mathbf{0} \longrightarrow \nabla f(\mathbf{x}) = \lambda_1 \nabla g_1(\mathbf{x}) + \cdots + \lambda_k \nabla g_k(\mathbf{x})$$
$$\frac{\partial L}{\partial \lambda_1} = 0, \ldots, \frac{\partial L}{\partial \lambda_k} = 0 \longrightarrow g_1(\mathbf{x}) - b_1 = 0, \ \ldots, \ g_k(\mathbf{x}) - b_k = 0$$

3. Evaluate and compare the function at all the critical points found above.

Question 2: What if inequality constraints?

$$\min f(\mathbf{x}) \qquad \text{subject to} \quad g(\mathbf{x}) \geq b$$

We will review this part later this semester when it is needed.

Next time: Matrix Computing in MATLAB

Make sure to do the following before Tuesday:

- Install MATLAB on your laptop
- Complete the 2-hour **MATLAB Onramp** tutorial[2]
- Explore the MATLAB documentation - **Getting Started with MATLAB**[3]

Lastly, bring your laptop to class next Tuesday.

[2] https://www.mathworks.com/learn/tutorials/matlab-onramp.html

[3] https://www.mathworks.com/help/matlab/getting-started-with-matlab.html