Optimization Models

EECS 127 / EECS 227AT

Laurent El Ghaoui

EECS department
UC Berkeley

Fall 2018

LECTURE 3

Matrices and Linear Maps

The Matrix is everywhere. It is all around us.

Morpheus

Outline

Introduction

- A matrix is a collection of numbers, arranged in columns and rows in a tabular format.
- Suitably defining operations such as sum, product and norms on matrices, we can treat matrices as elements of a vector space.
- A matrix defines a *linear map* between an input and an output space. This leads to the introduction of concepts such as range, rank, nullspace, eigenvalues and eigenvectors, that permit a complete analysis of (finite dimensional) linear maps.
- Matrices are an ubiquitous tool in engineering for organizing and manipulating data. They constitute the fundamental building block of numerical computation methods.

A data matrix

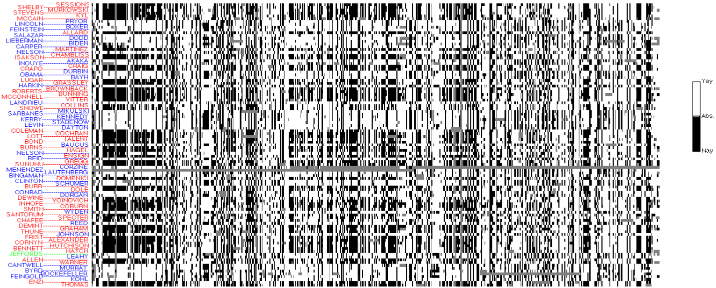


Figure: Votes of US Senators, 2002-2004.

Is there anything beyond just an array of numbers?

Basics

- We shall mainly deal with matrices whose elements are real (or sometimes complex) numbers, that is with arrays of the form

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}.$$

  This matrix has m *rows* and n *columns*.

- In the case of real elements, we say that $A \in \mathbb{R}^{m,n}$, resp. $A \in \mathbb{C}^{m,n}$ in the case of complex elements.
- The i-th row of A is the (row) vector $[a_{i1} \ \cdots \ a_{in}]$; the j-th column of A is the (column) vector $[a_{1j} \ \cdots \ a_{mj}]^\top$.
- The transposition operation works on matrices by exchanging rows and columns, that is

$$[A^\top]_{ij} = [A]_{ji},$$

  where the notation $[A]_{ij}$ (or sometimes also simply A_{ij}) refers to the element of A positioned in row i and column j.

Example

Matrices for networks

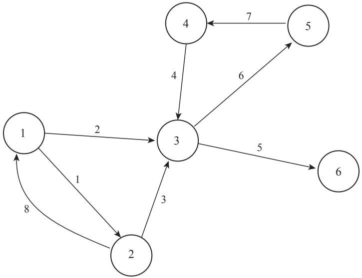
A network can be represented as a graph of m nodes connected by n directed arcs. Here, we assume that arcs are ordered pairs of nodes, with at most one arc joining any two nodes; we also assume that there are no self-loops (arcs from a node to itself).

We can fully describe such kind of network via the so-called (directed) arc-node incidence matrix, which is an $m \times n$ matrix defined as follows:

$$A_{ij} = \begin{cases} 1 & \text{if arc } j \text{ starts at node } i \\ -1 & \text{if arc } j \text{ ends at node } i \\ 0 & \text{otherwise.} \end{cases}, \quad 1 \leq i \leq m, \;\; 1 \leq j \leq n. \tag{1}$$

Example

Matrices for networks: example



A network with $m = 6$ nodes and $n = 8$ arcs, with (directed) arc-node incidence matrix

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

Basics

Matrix products

- Two matrices can be multiplied if conformably sized, i.e., if $A \in \mathbb{R}^{m,n}$ and $B \in \mathbb{R}^{n,p}$, then the matrix product $AB \in \mathbb{R}^{m,p}$ is defined as a matrix whose (i, j)-th entry is

$$[AB]_{ij} = \sum_{k=1}^{n} A_{ik} B_{kj}.$$

  The matrix product is non-commutative, meaning that, in general, $AB \neq BA$.

- The $n \times n$ *identity matrix* (often denoted I_n, or simply I, depending on context), is a matrix with all zero elements, except for the elements on the diagonal (that is, the elements with row index equal to the column index), which are equal to one. This matrix satisfies $AI_n = A$ for every matrix A with n columns, and $I_nB = B$ for every matrix B with n rows.

Basics

Matrix-vector product

- Let $A \in \mathbb{R}^{m,n}$ be a matrix with columns $a_1, \ldots, a_n \in \mathbb{R}^m$ and $b \in \mathbb{R}^n$ a vector. We define the matrix-vector product by

$$Ab = \sum_{k=1}^{n} a_k b_k, \quad A \in \mathbb{R}^{m,n}, b \in \mathbb{R}^n.$$

  That is, Ab is a vector in $\mathbb{R}^m$ obtained by forming a linear combination of the columns of A, using the elements in b as coefficients.

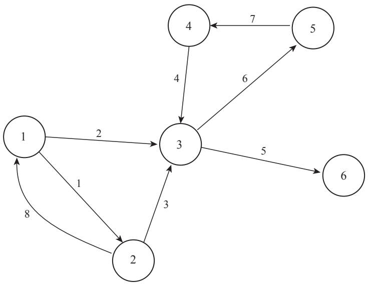
- Similarly, we can multiply matrix $A \in \mathbb{R}^{m,n}$ on the left by (the transpose of) vector $c \in \mathbb{R}^m$ as follows

$$c^\top A = \sum_{k=1}^{m} c_k \alpha_k^\top, \quad A \in \mathbb{R}^{m,n}, c \in \mathbb{R}^m.$$

  That is, $c^\top A$ is a vector in $\mathbb{R}^{1,n}$ obtained by forming a linear combination of the rows α_k of A, using the elements in c as coefficients.

Matrix-vector product

For a network incidence matrix



We describe a flow (of goods, traffic, charge, information, etc) across the network as a vector $x \in \mathbb{R}^n$, where the j-th component of x denotes the amount flowing through arc j. By convention, we use positive values when the flow is in the direction of the arc, and negative ones in the opposite case.

The total flow leaving a given node i is then

$$\sum_{j=1}^{n} A_{ij} x_j = [Ax]_i,$$

where $[Ax]_i$ denotes the i-th component of vector Ax.

Basics

Matrix products

- A matrix $A \in \mathbb{R}^{m,n}$ can also be seen as a collection of columns, each column being a vector, or as a collection of rows, each row being a (transposed) vector:

$$A = \begin{bmatrix} a_1 & a_2 & \cdots & a_n \end{bmatrix}, \quad \text{or} \quad A = \begin{bmatrix} \alpha_1^\top \\ \alpha_2^\top \\ \vdots \\ \alpha_m^\top \end{bmatrix},$$

  where $a_1, \ldots, a_n \in \mathbb{R}^m$ denote the columns of A, and $\alpha_1^\top, \ldots, \alpha_m^\top \in \mathbb{R}^n$ denote the rows of A.

- If the columns of B are given by the vectors $b_i \in \mathbb{R}^n$, $i = 1, \ldots, p$, so that $B = [b_1 \cdots b_p]$, then AB can be written as

$$AB = A \begin{bmatrix} b_1 & \ldots & b_p \end{bmatrix} = \begin{bmatrix} Ab_1 & \ldots & Ab_p \end{bmatrix}.$$

  In other words, AB results from transforming each column b_i of B into Ab_i.

Basics

Matrix products

- The matrix-matrix product can also be interpreted as an operation on the rows of A. Indeed, if A is given by its rows $\alpha_i^\top$, $i = 1, \ldots, m$, then AB is the matrix obtained by transforming each one of these rows into $\alpha_i^\top B$, $i = 1, \ldots, m$:

$$AB = \begin{bmatrix} \alpha_1^\top \\ \vdots \\ \alpha_m^\top \end{bmatrix} B = \begin{bmatrix} \alpha_1^\top B \\ \vdots \\ \alpha_m^\top B \end{bmatrix}.$$

- Finally, the product AB can be given the interpretation as the sum of so-called *dyadic* matrices (matrices of rank one, of the form $a_i \beta_i^\top$, where $\beta_i^\top$ denote the rows of B:

$$AB = \sum_{i=1}^{n} a_i \beta_i^\top, \quad A \in \mathbb{R}^{m,n}, B \in \mathbb{R}^{n,p}.$$

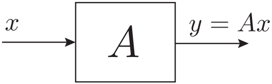
- For any two conformably sized matrices A, B, it holds that

$$(AB)^\top = B^\top A^\top,$$

Matrices as linear maps

- We can interpret matrices as linear maps (vector-valued functions), or "operators," acting from an "input" space to an "output" space.
- We recall that a map $f : \mathcal{X} \to \mathcal{Y}$ is linear if any points x and z in $\mathcal{X}$ and any scalars λ, μ satisfy $f(\lambda x + \mu z) = \lambda f(x) + \mu f(z)$.
- Any linear map $f : \mathbb{R}^n \to \mathbb{R}^m$ can be represented by a matrix $A \in \mathbb{R}^{m,n}$, mapping input vectors $x \in \mathbb{R}^n$ to output vectors $y \in \mathbb{R}^m$:

$$y = Ax.$$



- Affine maps are simply linear functions plus a constant term, thus any affine map $f : \mathbb{R}^n \to \mathbb{R}^m$ can be represented as

$$f(x) = Ax + b,$$

for some $A \in \mathbb{R}^{m,n}$, $b \in \mathbb{R}^m$.

Range, rank, and nullspace

- Consider a $m \times n$ matrix A, and denote by a_i, $i = 1, \ldots, n$, its i-th column, so that $A = [a_1 \ \ldots \ a_n]$.
- The set of vectors y obtained as a linear combination of the a_i's are of the form $y = Ax$ for some vector $x \in \mathbb{R}^n$. This set is commonly known as the *range* of A, and is denoted $\mathcal{R}(A)$:
$$\mathcal{R}(A) = \{Ax \ : \ x \in \mathbb{R}^n\}.$$
- By construction, the range is a subspace. The dimension of $\mathcal{R}(A)$ is called the *rank* of A and denoted with $\mathrm{rank}(A)$; by definition the rank represents the number of linearly independent columns of A.
- The rank is also equal to the number of linearly independent rows of A; that is, the rank of A is the same as that of its transpose $A^\top$. Proof here: `https://en.wikipedia.org/wiki/Rank_(linear_algebra)`
- As a consequence, we always have the bounds $1 \leq \mathrm{rank}(A) \leq \min(m, n)$.

Range, rank, and nullspace

- The nullspace of the matrix $A \in \mathbb{R}^{m,n}$ is the set of vectors in the input space that are mapped to zero, and is denoted $\mathcal{N}(A)$:

$$\mathcal{N}(A) = \{x \in \mathbb{R}^n : \; Ax = 0\}.$$

  This set is again a subspace.
- $\mathcal{R}(A^\top)$ and $\mathcal{N}(A)$ are mutually orthogonal subspaces, i.e., $\mathcal{N}(A) \perp \mathcal{R}(A^\top)$.
- The direct sum of a subspace and its orthogonal complement equals the whole space, thus,

$$\mathbb{R}^n = \mathcal{N}(A) \oplus \mathcal{N}(A)^\perp = \mathcal{N}(A) \oplus \mathcal{R}(A^\top).$$

Fundamental theorem of linear algebra

Theorem 1

For any given matrix $A \in \mathbb{R}^{m,n}$, *it holds that* $\mathcal{N}(A) \perp \mathcal{R}(A^\top)$ *and* $\mathcal{R}(A) \perp \mathcal{N}(A^\top)$, *hence*

$$\begin{aligned} \mathcal{N}(A) \oplus \mathcal{R}(A^\top) &= \mathbb{R}^n \\ \mathcal{R}(A) \oplus \mathcal{N}(A^\top) &= \mathbb{R}^m. \end{aligned}$$

Consequently, we can decompose any vector $x \in \mathbb{R}^n$ *as the sum of two vectors orthogonal to each other, one in the range of* $A^\top$, *and the other in the nullspace of* A:

$$x = A^\top \xi + z, \quad z \in \mathcal{N}(A).$$

Similarly, we can decompose any vector $w \in \mathbb{R}^m$ *as the sum of two vectors orthogonal to each other, one in the range of* A, *and the other in the nullspace of* $A^\top$:

$$w = A\varphi + \zeta, \quad \zeta \in \mathcal{N}(A^\top).$$

Fundamental theorem of linear algebra

Geometry

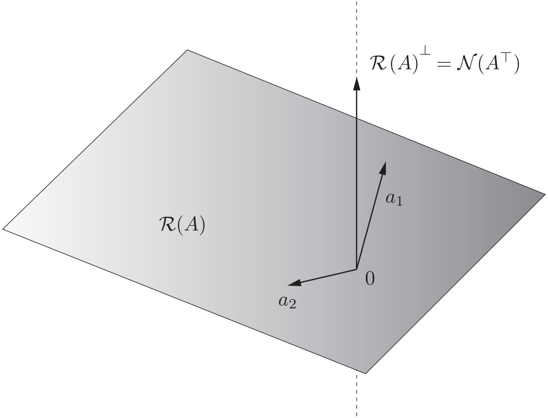


Figure: Illustration of the fundamental theorem of linear algebra in $\mathbb{R}^3$. Here, $A = [a_1\ a_2]$. Any vector in $\mathbb{R}^3$ can be written as the sum of two orthogonal vectors, one in the range of A, the other in the nullspace of $A^{\top}$.

Determinants

- The determinant of a generic (square) matrix $A \in \mathbb{R}^{n,n}$ can be computed by defining $\det a = a$ for a scalar a, and then applying the following inductive formula (Laplace's determinant expansion):

$$\det(A) = \sum_{j=1}^{n} (-1)^{i+j} a_{ij} \det A_{(i,j)},$$

  where i is any row, chosen at will (e.g., one may choose $i = 1$), and $A_{(i,j)}$ denotes a $(n-1) \times (n-1)$ submatrix of A obtained by eliminating row i and column j from A.

- $$A \in \mathbb{R}^{n,n} \text{ is singular } \Leftrightarrow \det A = 0 \Leftrightarrow \mathcal{N}(A) \text{ is not equal to } \{0\}.$$

- For any square matrices $A, B \in \mathbb{R}^{n,n}$ and scalar α:

$$\begin{aligned} \det A &= \det A^\top \\ \det AB &= \det BA = \det A \det B \\ \det \alpha A &= \alpha^n \det A. \end{aligned}$$

Determinant

Geometry

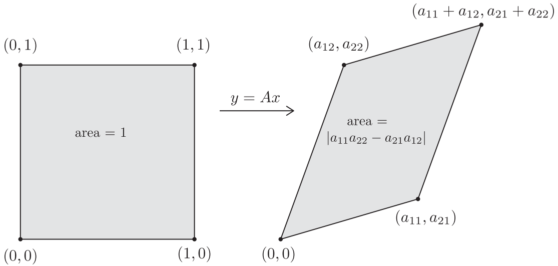


Figure: Linear mapping of the unit square. The absolute value of the determinant equals the area of the transformed unit square.

Matrix inverses

- If $A \in \mathbb{R}^{n,n}$ is nonsingular (i.e., $\det A \neq 0$), then we define the *inverse matrix* A^{-1} as the unique $n \times n$ matrix such that

$$AA^{-1} = A^{-1}A = I_n.$$

- If A, B are square and nonsingular, then it holds for the inverse of the product that $(AB)^{-1} = B^{-1}A^{-1}$.
- If A is square and nonsingular, then

$$\begin{aligned} (A^\top)^{-1} &= (A^{-1})^\top \\ \det A &= \det A^\top = \frac{1}{\det A^{-1}}. \end{aligned}$$

- For a generic matrix $A \in \mathbb{R}^{m,n}$, a generalized inverse (or, *pseudoinverse*) can be defined:

$$\text{if } m \geq n, \text{ then } A^{\mathrm{li}} \text{ is a \textit{left inverse} of } A, \text{ if } A^{\mathrm{li}}A = I_n.$$
$$\text{if } n \geq m, \text{ then } A^{\mathrm{ri}} \text{ is a \textit{right inverse} of } A, \text{ if } AA^{\mathrm{ri}} = I_m.$$

- In general, matrix A^{pi} is a pseudoinverse of A, if $AA^{\mathrm{pi}}A = A$.

Similar matrices

- Two matrices $A, B \in \mathbb{R}^{n,n}$ are said to be *similar* if there exist a nonsingular matrix $P \in \mathbb{R}^{n,n}$ such that
$$B = P^{-1}AP.$$
- Similar matrices are related to different representation of the same linear map, under a change of basis in the underlying space.
- Consider the linear map $y = Ax$ mapping $\mathbb{R}^n$ into itself. Since $P \in \mathbb{R}^{n,n}$ is nonsingular, its columns are linearly independent, hence they represent a basis for $\mathbb{R}^n$. Vectors x and y can thus be represented in this basis as linear combinations of the columns of P, that is there exist vectors $\tilde{x}$, $\tilde{y}$ such that
$$x = P\tilde{x}, \quad y = P\tilde{y}.$$
- Writing the relation $y = Ax$, substituting the representations of x, y in the new basis, we obtain
$$P\tilde{y} = AP\tilde{x} \quad \Rightarrow \quad \tilde{y} = P^{-1}AP\tilde{x} = B\tilde{x},$$
that is, matrix $B = P^{-1}AP$ represents the linear map $y = Ax$, in the new basis defined by the columns of P.

Eigenvalues/eigenvectors

- We say that $\lambda \in \mathbb{C}$ is an *eigenvalue* of matrix $A \in \mathbb{R}^{n,n}$, and $u \in \mathbb{C}^n$ is a corresponding *eigenvector*, if it holds that

$$Au = \lambda u, \quad u \neq 0,$$

  or, equivalently, $(\lambda I_n - A)u = 0$, $u \neq 0$.

- Eigenvalues can be characterized as those real or complex numbers that satisfy the equation

$$p(\lambda) \doteq \det(\lambda I_n - A) = 0,$$

  where $p(\lambda)$ is a polynomial of degree n in λ, known as the *characteristic polynomial* of A.

- Any matrix $A \in \mathbb{R}^{n,n}$ has n eigenvalues λ_i, $i = 1, \ldots, n$, counting multiplicities.

- To each distinct eigenvalue λ_i, $i = 1, \ldots, k$, there corresponds a whole subspace $\phi_i \doteq \mathcal{N}(\lambda_i I_n - A)$ of eigenvectors associated to this eigenvalue, called the *eigenspace*.

Diagonalizable matrices

Theorem 2

Let λ_i*,* $i = 1, \ldots, k \le n$*, be the distinct eigenvalues of* $A \in \mathbb{R}^{n,n}$*,*
let μ_i*,* $i = 1, \ldots, k$*, denote the corresponding algebraic multiplicities,*
let $\phi_i = \mathcal{N}(\lambda_i I_n - A)$*, and* $U^{(i)} = [u_1^{(i)} \cdots u_{\nu_i}^{(i)}]$ *be a matrix containing by columns a basis of* ϕ_i*, being* $\nu_i \doteq \dim \phi_i$*.*
It holds that $\nu_i \le \mu_i$ *and, if* $\nu_i = \mu_i$*,* $i = 1, \ldots, k$*, then*

$$U = [U^{(1)} \cdots U^{(k)}]$$

is invertible, and

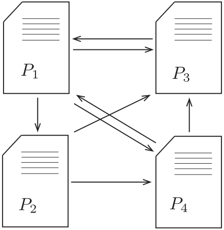
$$A = U \Lambda U^{-1}$$

where

$$\Lambda = \begin{bmatrix} \lambda_1 I_{\mu_1} & 0 & \cdots & 0 \\ 0 & \lambda_2 I_{\mu_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & \lambda_k I_{\mu_k} \end{bmatrix}.$$

Eigenvectors and Google's PageRank

- The effectiveness of Google's search engine largely relies on its PageRank (so named after Google's founder Larry Page) algorithm, which quantitatively ranks the importance of each page on the web, allowing Google to present to the user the more important (and typically most relevant and helpful) pages first.
- If the web of interest is composed of n pages, each labelled with integer k, $k = 1, \ldots, n$, we can model this web as a directed graph, where pages are the nodes of the graph, and a directed edge exists pointing from node k_1 to node k_2 if the web page k_1 contains a link to k_2.
- We denote by x_k, $k = 1, \ldots, n$ the importance *score* of page k.

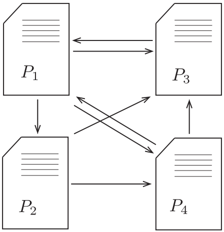


Eigenvectors and Google's PageRank

- Each page j has a score x_j and n_j outgoing links; as an assumption, we do not allow links from a page to itself, and we do not allow for dangling pages, that is pages with no outgoing links, therefore $n_j > 0$ for all j.
- The score x_j represents the total "voting'" power of node j, which is to be evenly subdivided among the n_j outgoing links; each outgoing link thus carries x_j/n_j units of vote.
- Let B_k denote the set of labels of the pages that point to page k, i.e., B_k is the set of backlinks for page k. Then, the score of page k is computed as

$$x_k = \sum_{j \in B_k} \frac{x_j}{n_j}, \quad k = 1, \dots, n.$$

Eigenvectors and Google's PageRank



For the example in the figure, we have $n_1 = 3$, $n_2 = 2$, $n_3 = 1$, $n_4 = 2$, hence

$$
\begin{aligned}
x_1 &= x_3 + \frac{1}{2}x_4 \\
x_2 &= \frac{1}{3}x_1 \\
x_3 &= \frac{1}{3}x_1 + \frac{1}{2}x_2 + \frac{1}{2}x_4 \\
x_4 &= \frac{1}{3}x_1 + \frac{1}{2}x_2.
\end{aligned}
$$

Eigenvectors and the Google PageRank

- We can write this system of equations in compact form exploiting the matrix-vector product rule, as follows

$$x = Ax, \quad A = \begin{bmatrix} 0 & 0 & 1 & \frac{1}{2} \\ \frac{1}{3} & 0 & 0 & 0 \\ \frac{1}{3} & \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{3} & \frac{1}{2} & 0 & 0 \end{bmatrix}, \quad x = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix}.$$

- Computing the web pages' scores thus amounts to finding x such that $Ax = x$: this is an eigenvalue/eigenvector problem and, in particular, x is an eigenvector of A associated with the eigenvalue $\lambda = 1$. A is called the *link matrix* of the network.
- In this example, the eigenspace $\phi_1 = \mathcal{N}(I_n - A)$ associated with the eigenvalue $\lambda = 1$ has dimension one, and it is given by

$$\phi_1 = \mathcal{N}(I_n - A) = \mathrm{span}\left(\begin{bmatrix} 12 & 4 & 9 & 6 \end{bmatrix}^\top \right)$$

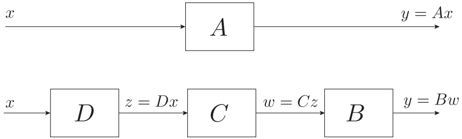
- Page 1 thus appears to be the most relevant, according to the PageRank scoring.

Matrices with special structure

- Square, diagonal, triangular (upper or lower)
- Symmetric: a square matrix A such that $A = A^\top$
- Orthogonal: a square matrix A such that $AA^\top = A^\top A = I$
- Dyad: a rank-one matrix A, which can be written as $A = uv^\top$, where u, v are vectors
- Block-structured matrices: block diagonal, block triangular, etc.

Matrix factorizations

- Given a matrix $A \in \mathbb{R}^{m,n}$, write this matrix as the product of two or more matrices with special structure.
- Usually, once a matrix is suitably factorized, several quantities of interest become readily accessible, and subsequent computations are greatly simplified.
- In terms of the linear map defined by a matrix A, a factorization can be interpreted as a decomposition of the map into a series of successive stages.



Matrix factorizations

More on this later

- Orthogonal-triangular decomposition (QR). Any square $A \in \mathbb{R}^{n,n}$ can be decomposed as
$$A = QR,$$
where Q is an orthogonal matrix, and R is an upper triangular matrix. If A is nonsingular, then the factors Q, R are uniquely defined, if the diagonal elements in R are imposed to be positive.
- Singular value decomposition (SVD). Any non-zero $A \in \mathbb{R}^{m,n}$ can be decomposed as
$$A = U\tilde{\Sigma}V^\top,$$
where $V \in \mathbb{R}^{n,n}$ and $U \in \mathbb{R}^{m,m}$ are orthogonal matrices, and
$$\tilde{\Sigma} = \begin{bmatrix} \Sigma & 0_{r,n-r} \\ 0_{m-r,r} & 0_{m-r,n-r} \end{bmatrix}, \quad \Sigma = \mathrm{diag}\left(\sigma_1, \ldots, \sigma_r\right),$$
where r is the rank of A, and the scalars $\sigma_i > 0$, $i = 1, \ldots, r$, are called the *singular values* of A. The first r columns $u_1, \ldots, u_r$ of U (resp. $v_1, \ldots, v_r$ of V) are called the left (resp. right) singular vectors, and satisfy
$$Av_i = \sigma_i u_i, \quad A^\top u_i = \sigma_i v_i, \quad i = 1, \ldots, r.$$

Matrix norms

- A function $f : \mathbb{R}^{m,n} \to \mathbb{R}$ is a matrix norm if, analogously to the vector case, it satisfies three standard axioms. Namely, for all $A, B \in \mathbb{R}^{m,n}$ and all $\alpha \in \mathbb{R}$:
  - $f(A) \geq 0$, and $f(A) = 0$ if and only if $A = 0$;
  - $f(\alpha A) = |\alpha| f(A)$;
  - $f(A + B) \leq f(A) + f(B)$.
- Many of the popular matrix norms also satisfy a fourth condition called *sub-multiplicativity*: for any conformably sized matrices A, B

$$f(AB) \leq f(A)f(B).$$

Matrix norms

Frobenius norm

- The Frobenius norm $\|A\|_F$ is nothing but the standard Euclidean (ℓ_2) vector norm applied to the vector formed by all elements of $A \in \mathbb{R}^{m,n}$:

$$\|A\|_F = \sqrt{\operatorname{trace} AA^\top} = \sqrt{\sum_{i=1}^{m} \sum_{j=1}^{n} |a_{ij}|^2}.$$

- The Frobenius norm also has an interpretation in terms of the eigenvalues of the symmetric matrix $AA^\top$:

$$\|A\|_F = \sqrt{\operatorname{trace} AA^\top} = \sqrt{\sum_{i=1}^{m} \lambda_i(AA^\top)}.$$

- For any $x \in \mathbb{R}^n$, it holds that $\|Ax\|_2 \leq \|A\|_F \|x\|_2$. (this is a consequence of the Cauchy-Schwartz inequality applied to $|a_i^\top x|$).
- The Frobenius norm is sub-multiplicative: for any $B \in \mathbb{R}^{n,p}$, it holds that

$$\|AB\|_F \leq \|A\|_F \|B\|_F.$$

Matrix norms

Operator norms

- The so-called operator norms give a characterization of the *maximum* input-output *gain* of the linear map $u \to y = Au$. Choosing to measure both inputs and outputs in terms of a given ℓ_p norm, with typical values $p = 1, 2, \infty$, leads to the definition

$$\|A\|_p \doteq \max_{u \neq 0} \frac{\|Au\|_p}{\|u\|_p} = \max_{\|u\|=1} \|Au\|_p,$$

- By definition, for every u, $\|Au\|_p \leq \|A\|_p \|u\|_p$. From this property follows that any operator norm is sub-multiplicative, that is, for any two conformably sized matrices A, B, it holds that

$$\|AB\|_p \leq \|A\|_p \|B\|_p.$$

- This fact is easily seen by considering the product AB as the series connection of the two operators B, A:

$$\|Bu\|_p \leq \|B\|_p \|u\|_p, \quad \|ABu\|_p \leq \|A\|_p \|Bu\|_p \leq \|A\|_p \|B\|_p \|u\|_p,$$

Matrix norms

Operator norms

For the typical values of $p = 1, 2, \infty$, we have the following results:

- The ℓ_1-induced norm corresponds to the largest absolute column sum:

$$\|A\|_1 = \max_{\|u\|_1 = 1} \|Au\|_1 = \max_{j=1,\dots,n} \sum_{i=1}^{m} |a_{ij}|.$$

- The ℓ_∞-induced norm corresponds to the largest absolute row sum:

$$\|A\|_\infty = \max_{\|u\|_\infty = 1} \|Au\|_\infty = \max_{i=1,\dots,m} \sum_{j=1}^{n} |a_{ij}|.$$

- The ℓ_2-induced norm (sometimes referred to as the *spectral* norm) corresponds to the square-root of the largest eigenvalue $\lambda_{\max}$ of $A^\top A$:

$$\|A\|_2 = \max_{\|u\|_2 = 1} \|Au\|_2 = \sqrt{\lambda_{\max}(A^\top A)}.$$

  The latter identity follows from the variational characterization of the eigenvalues of a symmetric matrix. In lecture 5 we revisit this.

Spectral radius

- The spectral radius $\rho(A)$ of a matrix $A \in \mathbb{R}^{n,n}$ is defined as the maximum modulus of the eigenvalues of A, that is

$$\rho(A) \doteq \max_{i=1,\dots,n} |\lambda_i(A)|.$$

- Clearly, $\rho(A) \geq 0$ for all A, and $A = 0$ implies $\rho(A) = 0$. However, the converse is not true, since $\rho(A) = 0$ does not imply necessarily that $A = 0$, hence $\rho(A)$ is not a matrix norm.
- However, for any induced matrix norm $\|\cdot\|_p$, it holds that

$$\rho(A) \leq \|A\|_p.$$

- It follows, in particular, that $\rho(A) \leq \min(\|A\|_1, \|A\|_\infty)$, that is $\rho(A)$ is no larger than the maximum row or column sum of $|A|$ (the matrix whose entries are the absolute values of the entries in A).

Matrix functions

Matrix powers and polynomials

- The integer power function

$$f(X) = X^k, \quad k = 0, 1, \ldots$$

  can be quite naturally defined via the matrix product, by observing that $X^k = XX \cdots X$ (k times; we take the convention that $X^0 = I_n$).
- Similarly, negative integer power functions can be defined over nonsingular matrices as integer powers of the inverse:

$$f(X) = X^{-k} = (X^{-1})^k, \quad k = 0, 1, \ldots$$

- A polynomial matrix function of degree $m \geq 0$ can hence be naturally defined as

$$p(X) = a_m X^m + a_{m-1} X^{m-1} + \cdots + a_1 X + a_0 I_n,$$

  where a_i, $i = 0, 1, \ldots, m$, are the scalar coefficients of the polynomial.

Matrix functions

Diagonal factorization of a matrix polynomial

- Let $X \in \mathbb{R}^{n,n}$ admit a diagonal factorization

$$X = U\Lambda U^{-1},$$

where Λ is a diagonal matrix containing the eigenvalues of X, and U is a matrix containing by columns the corresponding eigenvectors. Let $p(t)$, $t \in \mathbb{R}$, be a polynomial

$$p(t) = a_m t^m + a_{m-1} t^{m-1} + \cdots + a_1 t + a_0.$$

Then,

$$p(X) = U p(\Lambda) U^{-1},$$

where

$$p(\Lambda) = \mathrm{diag}\left(p(\lambda_1), \ldots, p(\lambda_n)\right).$$

- More generally, if λ, u is an eigenvalue/eigenvector pair for X, then

$$p(X)u = p(\lambda)u.$$

Matrix functions

Non-polynomial matrix functions

- Let $f : \mathbb{R} \to \mathbb{R}$ be an *analytic* function, that is, a function which is locally representable by a power series $f(t) = \sum_{k=0}^{\infty} a_k t^k$, which is convergent for all t such that $|t| \leq R$, $R > 0$.
- If $\rho(X) < R$ (where $\rho(X)$ is the spectral radius of X), then the value of the matrix function $f(X)$ can be defined as the sum of the convergent series

$$f(X) = \sum_{k=0}^{\infty} a_k X^k.$$

- Moreover, if X is diagonalizable, then $X = U \Lambda U^{-1}$, and

$$f(X) \;=\; \sum_{k=0}^{\infty} a_k X^k = U f(\Lambda) U^{-1}.$$

- This equation states in particular that the *spectrum* (i.e., the set of eigenvalues) of $f(A)$ is the image of the spectrum of A under f. This fact is known as the *spectral mapping theorem*.

Matrix functions

Examples

- The *matrix exponential*: the function $f(t) = \mathrm{e}^t$ has a power series representation which is globally convergent

$$\mathrm{e}^t = \sum_{k=0}^{\infty} \frac{1}{k!} t^k,$$

hence, for any diagonalizable $X \in \mathbb{R}^{n,n}$, we have

$$\mathrm{e}^X \doteq \sum_{k=0}^{\infty} \frac{1}{k!} X^k = U \mathrm{diag}\left(\mathrm{e}^{\lambda_1}, \ldots, \mathrm{e}^{\lambda_n}\right) U^{-1}.$$

- Another example is given by the geometric series

$$f(t) = (1-t)^{-1} = \sum_{k=0}^{\infty} t^k, \quad \text{for } |t| < 1 = R,$$

from which we obtain that

$$f(X) = (I-X)^{-1} = \sum_{k=0}^{\infty} X^k, \quad \text{for } \rho(X) < 1.$$