Numerical Analysis
(Review of) Linear Algebra

Aleksandar Donev
Courant Institute, NYU[1]
donev@courant.nyu.edu

[1] Course MATH-UA.0252/MA-UY_4424, Spring 2021

Spring 2021

Outline

Outline

Linear Spaces

- A **vector space** $\mathcal{V}$ is a set of elements called **vectors** $\mathbf{x} \in \mathcal{V}$ that may be multiplied by a **scalar** c and added, e.g.,

$$\mathbf{z} = \alpha \mathbf{x} + \beta \mathbf{y}$$

- I will denote scalars with lowercase letters and vectors with lowercase bold letters.
- Prominent examples of vector spaces are $\mathbb{R}^n$ (or more generally $\mathbb{C}^n$), but there are many others, for example, the set of polynomials in x.
- A **subspace** $\mathcal{V}' \subseteq \mathcal{V}$ of a vector space is a subset such that sums and multiples of elements of $\mathcal{V}'$ remain in $\mathcal{V}'$ (i.e., it is closed).
- An example is the set of vectors in $x \in \mathbb{R}^3$ such that $x_3 = 0$.

Image Space

- Consider a set of n vectors $\mathbf{a}_1, \mathbf{a}_2, \cdots, \mathbf{a}_n \in \mathbb{R}^m$ and form a **matrix** by putting these vectors as columns

$$\mathbf{A} = [\mathbf{a}_1 \,|\, \mathbf{a}_2 \,|\, \cdots \,|\, \mathbf{a}_m] \in \mathbb{R}^{m,n}.$$

- I will denote matrices with bold capital letters, and sometimes write $\mathbf{A} = [m, n]$ to indicate dimensions.
- The **matrix-vector product** is defined as a **linear combination** of the columns:

$$\mathbf{b} = \mathbf{A}\mathbf{x} = x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + \cdots + x_n\mathbf{a}_n \in \mathbb{R}^m.$$

- The **image** $\text{im}(\mathbf{A})$ or **range** $\text{range}(\mathbf{A})$ of a matrix is the subspace of all linear combinations of its columns, i.e., the set of all $\mathbf{b}'s$.
  It is also sometimes called the **column space** of the matrix.

Dimension

- The set of vectors $\mathbf{a}_1, \mathbf{a}_2, \cdots, \mathbf{a}_n$ are **linearly independent** or form a **basis** for $\mathbb{R}^m$ if $\mathbf{b} = \mathbf{A}\mathbf{x} = \mathbf{0}$ implies that $\mathbf{x} = \mathbf{0}$.
- The **dimension** $r = \dim \mathcal{V}$ of a vector (sub)space $\mathcal{V}$ is the number of elements in a basis. This is a property of $\mathcal{V}$ itself and *not* of the basis, for example,

$$\dim \mathbb{R}^n = n$$

- Given a basis $\mathbf{A}$ for a vector space $\mathcal{V}$ of dimension n, every vector of $\mathbf{b} \in \mathcal{V}$ can be uniquely represented as the vector of coefficients $\mathbf{x}$ in that particular basis,

$$\mathbf{b} = x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + \cdots + x_n\mathbf{a}_n.$$

- A simple and common basis for $\mathbb{R}^n$ is $\{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$, where $\mathbf{e}_k$ has all components zero except for a single 1 in position k.
  With this choice of basis the coefficients are simply the entries in the vector, $\mathbf{b} \equiv \mathbf{x}$.

Kernel Space

- The dimension of the column space of a matrix is called the **rank** of the matrix $\mathbf{A} \in \mathbb{R}^{m,n}$,
$$r = \operatorname{rank} \mathbf{A} \leq \min(m, n).$$
- If $r = \min(m, n)$ then the matrix is of **full rank**.
- The **nullspace** null($\mathbf{A}$) or **kernel** ker($\mathbf{A}$) of a matrix $\mathbf{A}$ is the subspace of vectors $\mathbf{x}$ for which
$$\mathbf{A}\mathbf{x} = \mathbf{0}.$$
- The dimension of the nullspace is called the **nullity** of the matrix.
- For a basis $\mathbf{A}$ the nullspace is $\text{null}(\mathbf{A}) = \{\mathbf{0}\}$ and the nullity is zero.

Orthogonal Spaces

- An inner-product space is a vector space together with an **inner or dot product**, which must satisfy some properties.
- The standard dot-product in $\mathbb{R}^n$ is denoted with several different notations:

$$\mathbf{x} \cdot \mathbf{y} = (\mathbf{x}, \mathbf{y}) = \langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{x}^T \mathbf{y} = \sum_{i=1}^{n} x_i y_i .$$

- For $\mathbb{C}^n$ we need to add complex conjugates (here $\star$ denotes a complex conjugate transpose, or **adjoint**),

$$\mathbf{x} \cdot \mathbf{y} = \mathbf{x}^\star \mathbf{y} = \sum_{i=1}^{n} \bar{x}_i y_i .$$

- Two vectors $\mathbf{x}$ and $\mathbf{y}$ are **orthogonal** if $\mathbf{x} \cdot \mathbf{y} = 0$.

Part I of Fundamental Theorem

- One of the most important theorems in linear algebra is that the sum of rank and nullity is equal to the number of columns: For $\mathbf{A} \in \mathbb{R}^{m,n}$

$$\operatorname{rank} \mathbf{A} + \operatorname{nullity} \mathbf{A} = n.$$

- In addition to the range and kernel spaces of a matrix, two more important vector subspaces for a given matrix $\mathbf{A}$ are the:
    - **Row space** or **coimage** of a matrix is the column (image) space of its transpose, $\operatorname{im} \mathbf{A}^T$.
      Its dimension is also equal to the the rank.
    - **Left nullspace** or **cokernel** of a matrix is the nullspace or kernel of its transpose, $\ker \mathbf{A}^T$.

Part II of Fundamental Theorem

- The **orthogonal complement** $\mathcal{V}^{\perp}$ or orthogonal subspace of a subspace $\mathcal{V}$ is the set of all vectors that are orthogonal to every vector in $\mathcal{V}$.
- Let $\mathcal{V}$ be the set of vectors in $x \in \mathbb{R}^3$ such that $x_3 = 0$. Then $\mathcal{V}^{\perp}$ is the set of all vectors with $x_1 = x_2 = 0$.
- Second fundamental theorem in linear algebra:

$$\operatorname{im} \mathbf{A}^T = (\ker \mathbf{A})^{\perp}$$

$$\ker \mathbf{A}^T = (\operatorname{im} \mathbf{A})^{\perp}$$

Outline

Linear Transformation

- A function $L : \mathcal{V} \rightarrow \mathcal{W}$ mapping from a vector space $\mathcal{V}$ to a vector space $\mathcal{W}$ is a **linear function** or a **linear transformation** if

$$L(\alpha \mathbf{v}) = \alpha L(\mathbf{v}) \text{ and } L(\mathbf{v}_1 + \mathbf{v}_2) = L(\mathbf{v}_1) + L(\mathbf{v}_2).$$

- Any linear transformation L can be represented as a multiplication by a matrix $\mathbf{L}$

$$L(\mathbf{v}) = \mathbf{L}\mathbf{v}.$$

- For the common bases of $\mathcal{V} = \mathbb{R}^n$ and $\mathcal{W} = \mathbb{R}^m$, the product $\mathbf{w} = \mathbf{L}\mathbf{v}$ is simply the usual **matix-vector product**,

$$w_i = \sum_{k=1}^{n} L_{ik} v_k,$$

which is simply the dot-product between the i-th row of the matrix and the vector $\mathbf{v}$.

Matrix algebra

$$w_i = (\mathbf{Lv})_i = \sum_{k=1}^{n} L_{ik} v_k$$

- The composition of two linear transformations $\mathbf{A} = [m, p]$ and $\mathbf{B} = [p, n]$ is a **matrix-matrix product** $\mathbf{C} = \mathbf{AB} = [m, n]$:

$$\mathbf{z} = \mathbf{A}(\mathbf{Bx}) = \mathbf{Ay} = (\mathbf{AB})\,\mathbf{x}$$

$$z_i = \sum_{k=1}^{n} A_{ik} y_k = \sum_{k=1}^{p} A_{ik} \sum_{j=1}^{n} B_{kj} x_j = \sum_{j=1}^{n} \left(\sum_{k=1}^{p} A_{ik} B_{kj} \right) x_j = \sum_{j=1}^{n} C_{ij} x_j$$

$$C_{ij} = \sum_{k=1}^{p} A_{Ik} B_{kj}$$

- Matrix-matrix multiplication is **not commutative**, $\mathbf{AB} \neq \mathbf{BA}$ in general.

The Matrix Inverse

- A square matrix $\mathbf{A} = [n, n]$ is **invertible or nonsingular** if there exists a **matrix inverse** $\mathbf{A}^{-1} = \mathbf{B} = [n, n]$ such that:

$$\mathbf{AB} = \mathbf{BA} = \mathbf{I},$$

  where $\mathbf{I}$ is the identity matrix (ones along diagonal, all the rest zeros).
- The following statements are equivalent for $\mathbf{A} \in \mathbb{R}^{n,n}$:
  - $\mathbf{A}$ is **invertible**.
  - $\mathbf{A}$ is **full-rank**, $\operatorname{rank} \mathbf{A} = n$.
  - The columns and also the rows are linearly independent and form a **basis** for $\mathbb{R}^n$.
  - The **determinant** is nonzero, $\det \mathbf{A} \neq 0$.
  - Zero is not an eigenvalue of $\mathbf{A}$.

Matrix Algebra

- Matrix-vector multiplication is just a special case of matrix-matrix multiplication. Note $\mathbf{x}^T\mathbf{y}$ is a scalar (dot product).

$$\mathbf{C}\left(\mathbf{A}+\mathbf{B}\right)=\mathbf{CA}+\mathbf{CB} \text{ and } \mathbf{ABC}=\left(\mathbf{AB}\right)\mathbf{C}=\mathbf{A}\left(\mathbf{BC}\right)$$

$$\left(\mathbf{A}^T\right)^T=\mathbf{A} \text{ and } \left(\mathbf{AB}\right)^T=\mathbf{B}^T\mathbf{A}^T$$

$$\left(\mathbf{A}^{-1}\right)^{-1}=\mathbf{A} \text{ and } \left(\mathbf{AB}\right)^{-1}=\mathbf{B}^{-1}\mathbf{A}^{-1} \text{ and } \left(\mathbf{A}^T\right)^{-1}=\left(\mathbf{A}^{-1}\right)^T$$

- Instead of **matrix division**, think of multiplication by an inverse:

$$\mathbf{AB}=\mathbf{C} \quad \Rightarrow \quad \left(\mathbf{A}^{-1}\mathbf{A}\right)\mathbf{B}=\mathbf{A}^{-1}\mathbf{C} \quad \Rightarrow \quad \begin{cases} \mathbf{B} & =\mathbf{A}^{-1}\mathbf{C} \\ \mathbf{A} & =\mathbf{CB}^{-1} \end{cases}$$

Outline

Vector norms

- Norms are the abstraction for the notion of a length or **magnitude**.
- For a vector $\mathbf{x} \in \mathbb{R}^n$, the p-norm is

$$\|\mathbf{x}\|_p = \left(\sum_{i=1}^{n} |x_i|^p \right)^{1/p}$$

and special cases of interest are:

1. The 1-norm (L^1 norm or Manhattan distance), $\|\mathbf{x}\|_1 = \sum_{i=1}^{n} |x_i|$
2. The 2-norm (L^2 norm, **Euclidian distance**), $\|\mathbf{x}\|_2 = \sqrt{\mathbf{x} \cdot \mathbf{x}} = \sqrt{\sum_{i=1}^{n} |x_i|^2}$
3. The ∞-norm (L^∞ or maximum norm), $\|\mathbf{x}\|_\infty = \max_{1 \le i \le n} |x_i|$

1. Note that all of these norms are inter-related in a finite-dimensional setting.

Matrix norms

- Matrix norm **induced** by a given vector norm:

$$\|\mathbf{A}\| = \sup_{\mathbf{x}\neq\mathbf{0}} \frac{\|\mathbf{A}\mathbf{x}\|}{\|\mathbf{x}\|} \quad \Rightarrow \|\mathbf{A}\mathbf{x}\| \leq \|\mathbf{A}\| \, \|\mathbf{x}\|$$

- The last bound holds for matrices as well, $\|\mathbf{AB}\| \leq \|\mathbf{A}\| \, \|\mathbf{B}\|$.
- Special cases of interest are:
  1. The 1-norm or **column sum norm**, $\|\mathbf{A}\|_1 = \max_j \sum_{i=1}^n |a_{ij}|$
  2. The ∞-norm or **row sum norm**, $\|\mathbf{A}\|_\infty = \max_i \sum_{j=1}^n |a_{ij}|$
  3. The 2-norm or **spectral norm**, $\|\mathbf{A}\|_2 = \sigma_1$ (largest singular value)
  4. The Euclidian or **Frobenius norm**, $\|\mathbf{A}\|_F = \sqrt{\sum_{i,j} |a_{ij}|^2}$
     (note this is not an induced norm)

Outline

Conditioning

- Consider a function $\mathbf{f}:\mathbb{R}^n \rightarrow \mathbb{R}^m$, and perturb $\mathbf{x}$ to the **absolute condition number**

$$\text{Cond}_{\mathbf{x}}(f) = \sup_{\delta\mathbf{x}\neq\mathbf{0}} \frac{\|\mathbf{f}(\mathbf{x}+\delta\mathbf{x}) - \mathbf{f}(\mathbf{x})\|}{\|\delta\mathbf{x}\|}$$

  where $\|\delta\mathbf{x}\| \ll \|\mathbf{x}\|$ is a small perturbation (assume $\mathbf{x}\neq\mathbf{0}$).
- This measures how **sensitive** the value of the function is to small errors in the input (e.g., roundoff or measurement).
- For differentiable scalar functions $f(x\in\mathbb{R})\in\mathbb{R}$,

$$\text{Cond}_x(f) = \left|f'(x)\right|.$$

- More commonly used is the **relative condition number**

$$\text{cond}_{\mathbf{x}}(f) = \sup_{\delta\mathbf{x}\neq\mathbf{0}} \frac{\|\mathbf{f}(\mathbf{x}+\delta\mathbf{x}) - \mathbf{f}(\mathbf{x})\| / \|\mathbf{f}(\mathbf{x})\|}{\|\delta\mathbf{x}\| / \|\mathbf{x}\|}$$

  which measures the maximum relative change in the output for a given small relative change in the input.

Conditioning number

- Consider a linear mapping $\mathbf{f}(\mathbf{x}) = \mathbf{A}\mathbf{x}$. What is the relative conditioning number?

$$\begin{aligned}\operatorname{cond}_{\mathbf{x}}(f) &= \sup_{\delta\mathbf{x}\neq\mathbf{0}} \frac{\|\mathbf{A}(\mathbf{x}+\delta\mathbf{x}) - \mathbf{A}\mathbf{x}\| / \|\mathbf{A}\mathbf{x}\|}{\|\delta\mathbf{x}\| / \|\mathbf{x}\|} \\ &= \frac{\|\mathbf{x}\|}{\|\mathbf{A}\mathbf{x}\|} \sup_{\delta\mathbf{x}\neq\mathbf{0}} \frac{\|\mathbf{A}\delta\mathbf{x}\|}{\|\delta\mathbf{x}\|} = \\ &= \frac{\|\mathbf{x}\|}{\|\mathbf{A}\mathbf{x}\|} \|\mathbf{A}\| \geq 1.\end{aligned}$$

- To get an upper bound, consider an invertible square $\mathbf{A}$,

$$\operatorname{cond}_{\mathbf{x}}(f) = \frac{\left\|\mathbf{A}^{-1}(\mathbf{A}\mathbf{x})\right\|}{\|\mathbf{A}\mathbf{x}\|} \|\mathbf{A}\| \leq \left\|\mathbf{A}^{-1}\right\| \frac{\|\mathbf{A}\mathbf{x}\|}{\|\mathbf{A}\mathbf{x}\|} \|\mathbf{A}\|$$

which leads us to define a **matrix condition number**

$$\kappa(\mathbf{A}) = \left\|\mathbf{A}^{-1}\right\| \|\mathbf{A}\| > 1.$$