Session 1: Vector spaces

Optimization and Computational Linear Algebra for Data Science

Marylou Gabrié (based on material by Léo Miolane)

Contents

1. Vector spaces

So far, « Vectors = arrows »

Two fundamental operations:

1. Add two vectors $\vec{u}$ and $\vec{v}$ to obtain another vector $\vec{u} + \vec{v}$

2. Multiply a vector $\vec{u}$ by a «scalar» (= a real number) λ to get another vector $\lambda \cdot \vec{u}$

Coordinate representation

- One can represent vectors using coordinates
- 2D vectors in the plane $\vec{u} = (u_1, u_2) \in \mathbb{R}^2$
- 3D vectors in space $\vec{u} = (u_1, u_2, u_3) \in \mathbb{R}^3$
- n-dimensional vectors $\vec{u} = (u_1, u_2, \ldots, u_n) \in \mathbb{R}^n$

- $\vec{u} + \vec{v} = (u_1 + v_1, u_2 + v_2, \ldots, u_n + v_n)$
- $\lambda \cdot \vec{u} = (\lambda u_1, \lambda u_2, \ldots, \lambda u_n)$

1.1 Vector spaces: Definition

Definition (simplified - see notes)

A vector space consists of a set V (whose elements are called vectors) and two operations $+$ and $\cdot$ such that

- The sum of two vectors is a vector: for $\vec{x}, \vec{y} \in V$, the sum $\vec{x} + \vec{y}$ is a vector, i.e. $\vec{x} + \vec{y} \in V$.
- Multiplying a vector $\vec{x} \in V$ by a scalar $\lambda \in \mathbb{R}$ gives a vector $\lambda \cdot \vec{x} \in V$.
- The operations $+$ and $\cdot$ are "nice and compatible".

« Nice and compatible » ?

1. The vector sum is commutative and associative. For all $\vec{x}, \vec{y}, \vec{z} \in V$:

$$\vec{x} + \vec{y} = \vec{y} + \vec{x} \qquad \text{and} \qquad \vec{x} + (\vec{y} + \vec{z}) = (\vec{x} + \vec{y}) + \vec{z}.$$

2. There exists a zero vector $\vec{0} \in V$ that verifies $\vec{x} + \vec{0} = \vec{x}$ for all $\vec{x} \in V$.
3. For all $\vec{x} \in V$, there exists $\vec{y} \in V$ such that $\vec{x} + \vec{y} = \vec{0}$. Such $\vec{y}$ is called the additive inverse of $\vec{x}$ and is written $-\vec{x}$.
4. Identity element for scalar multiplication: $1 \cdot \vec{x} = \vec{x}$ for all $\vec{x} \in V$.
5. Distributivity: for all $\alpha, \beta \in \mathbb{R}$ and all $\vec{x}, \vec{y} \in V$,

$$(\alpha + \beta) \cdot \vec{x} = \alpha \cdot \vec{x} + \beta \cdot \vec{y} \qquad \text{and} \qquad \alpha \cdot (\vec{x} + \vec{y}) = \alpha \cdot \vec{x} + \alpha \cdot \vec{y}.$$

6. Compatibility between scalar multiplication and the usual multiplication: for all $\alpha, \beta \in \mathbb{R}$ and all $\vec{x} \in V$, we have

$$\alpha \cdot (\beta \cdot \vec{x}) = (\alpha\beta) \cdot \vec{x}.$$

Example 1: $\mathbb{R}^n$

The set $V = \mathbb{R}^n$ endowed with the usual vector addition $+$

$$(x_1, \ldots, x_n) + (y_1, \ldots, y_n) = (x_1 + y_1, \ldots, x_n + y_n)$$

and the usual scalar multiplication $\cdot$

$$\alpha \cdot (x_1, \ldots, x_n) = (\alpha x_1, \ldots, \alpha x_n)$$

is a vector space.

We will work in $\mathbb{R}^n$ 99% ot the time !

Example 2: functions

The set $V \stackrel{\text{def}}{=} \{f \mid f : \mathbb{R} \to \mathbb{R}\}$ of all functions from $\mathbb{R}$ to itself endowed with the addition $+$ and the scalar multiplication $\cdot$ defined by

$$\begin{array}{rccc} f+g: & \mathbb{R} & \to & \mathbb{R} \\ & t & \mapsto & f(t)+g(t) \end{array} \qquad \text{and} \qquad \begin{array}{rccc} \alpha \cdot f: & \mathbb{R} & \to & \mathbb{R} \\ & t & \mapsto & \alpha f(t) \end{array}$$

is a vector space.

Useful in signal processing.

Example 3: random variables

The set of random variables on a given probability space Ω is a vector space:
If X and Y are two random variables and $\alpha \in \mathbb{R}$, $X + Y$ and αX are also random variables.

Important to have this in mind when doing stats/probabilities!

Why do we need all this?

- **Get geometric intuition.**

  We will see for instance that the notion of length in $\mathbb{R}^n$ is deeply connected to the notion of variance of random variables.

- **Save time.**

  A theorem that applies to vector spaces will in particular be true for all the examples we listed before.

1.2 Subspaces: Definition

Definition

We say that a non-empty subset S of a vector space V is a *subspace* if it is closed under addition and multiplication by a scalar, that is if

1. for all $x, y \in S$ we have $x + y \in S$,
2. for all $x \in S$ and all $\alpha \in \mathbb{R}$ we have $\alpha x \in S$.

Remark: a subspace is a also vector space.

Examples

- $\mathbb{R}^n$ is a subspace of $\mathbb{R}^n$.
- $\{0\}$ is a subspace of $\mathbb{R}^n$.
- Any line that contains the origin is subspace of $\mathbb{R}^2$.

Remarks, questions ?

Remarks, questions ?

2. Span & linear dependency

Linear combination

Let V be a vector space (think for instance $V = \mathbb{R}^n$).

Definition

We say that $y \in V$ is a *linear combination* of the vectors $x_1, \dots, x_k \in V$ if there exists $\alpha_1, \dots, \alpha_k \in \mathbb{R}$ such that

$$y = \sum_{i=1}^{k} \alpha_i x_i = \alpha_1 x_1 + \cdots + \alpha_k x_k.$$

Remarks

- A linear combination is always a finite sum.
- If S is a subspace of V, then any linear combination of vectors $x_1, \dots, x_k$ of S is also in S:

$$\alpha_1 x_1 + \cdots + \alpha_k x_k \in S, \qquad \text{for all } \alpha_1, \dots, \alpha_k \in \mathbb{R}.$$

« Subspaces are closed under linear combinations. »

Exercise: Prove it !

Span

Definition

Let $x_1, \ldots, x_k$ be vectors of V. We define the *linear span* of $x_1, \ldots, x_k$ as the set of all linear combinations of these vectors:

$$\mathrm{Span}(x_1, \ldots, x_k) \stackrel{\text{def}}{=} \left\{ \alpha_1 x_1 + \cdots + \alpha_k x_k \,\middle|\, \alpha_1, \ldots, \alpha_k \in \mathbb{R} \right\}.$$

2.2 Linear dependency

Definition

Vectors $x_1, \dots x_k \in V$ are *linearly dependent* is there exists $\alpha_1, \dots, \alpha_k \in \mathbb{R}$ **that are not all zero** such that

$$\alpha_1 x_1 + \cdots + \alpha_k x_k = 0.$$

They are said to be *linearly independent* otherwise. **Abuse of language:** Instead of saying «$x_1, \dots, x_k$ are linearly dependent», we should say «the family $(x_1, \dots, x_k)$ is linearly dependent».

Key observation: « $x_1, \dots, x_k$ are linearly dependent » *is equivalent to* « one of the vectors $x_1, \dots, x_k$ can be obtained as a linear combination of the others.»

Why ?

A useful lemma

Lemma

Let $v_1, \dots, v_n \in V$ and let $x_1, \dots, x_k \in \mathrm{Span}(v_1, \dots, v_n)$.
Then, if $k > n$, $x_1, \dots, x_k$ are linearly dependent.

3. Basis & Dimension

3.1 Basis definition

Definition

A family $(x_1, \dots, x_n)$ of vectors of V is a basis of V if

1. $x_1, \dots, x_n$ are linearly independent,
2. $\mathrm{Span}(x_1, \dots, x_n) = V$.

This means that $(x_1, \dots, x_n)$ is a basis of V if

1. None of the x_i is a linear combination of the others $(x_j)_{j \neq i}$.
2. Any vector of V can be expressed as a linear combination of $(x_1, \dots, x_n)$.

Example: the canonical basis of $\mathbb{R}^n$

Let us define the vectors $e_1, \dots, e_n \in \mathbb{R}^n$ by

$$\begin{aligned} e_1 &= (1, 0, 0, \dots, 0) \\ e_2 &= (0, 1, 0, \dots, 0) \\ &\vdots \\ e_n &= (0, 0, 0, \dots, 1). \end{aligned}$$

One can verify (homework!) that the family $(e_1, \dots, e_n)$ is a basis of $\mathbb{R}^n$. This basis is called the “canonical basis” of $\mathbb{R}^n$.

3.2 Dimension

Theorem

Let V be a vector space.

- If V admits a basis $(v_1, \ldots, v_n)$, then every basis of V has also n vectors. We say that V has dimension n and write $\dim(V) = n$.
- Otherwise, we say that V has infinite dimension: $\dim(V) = +\infty$.

Example:

- $\mathbb{R}^2$ has dimension 2, because the canonical basis (e_1, e_2) is a basis of $\mathbb{R}^2$ with 2 vectors.
- $\{f \mid f : \mathbb{R} \to \mathbb{R}\}$ has infinite dimension.

The dimension is well defined!

Theorem

If V admits a basis $(v_1, \dots, v_n)$, then every basis of V has also n vectors.

Proof.

$\square$

Properties of the dimension

Proposition

Let V be a vector space that has dimension $\dim(V) = n$. Then

1. Any family of vectors of V that spans V contains at least n vectors.

   i.e. if $x_1, \ldots, x_k \in V$ are such that $\mathrm{Span}(x_1, \ldots, x_k) = V$, then $k \geq n$.

2. Any family of vectors of V that are linearly independent contains at most n vectors.

   i.e. if $x_1, \ldots, x_k \in V$ are linearly independent, then $k \leq n$.

Proof.

Properties of the dimension

Proposition

Let V be a vector space that has dimension $\dim(V) = n$. Then

1. Any family of vectors of V that spans V contains at least n vectors.

   i.e. if $x_1, \dots, x_k \in V$ are such that $\mathrm{Span}(x_1, \dots, x_k) = V$, then $k \geq n$.

2. Any family of vectors of V that are linearly independent contains at most n vectors.

   i.e. if $x_1, \dots, x_k \in V$ are linearly independent, then $k \leq n$.

Proof.

Properties of the dimension

Proposition

Let V be a vector space of dimension n and let $x_1, \dots, x_n \in V$.

1. If $x_1, \dots, x_n$ are linearly independent, then $(x_1, \dots, x_n)$ is a basis of V.
2. If $\mathrm{Span}(x_1, \dots, x_n) = V$, then $(x_1, \dots, x_n)$ is a basis of V.

Very useful to show that a family of vector forms a basis:

Example: $x_1 = (12, 37)$ and $x_2 = (-9, 17)$ form a basis of $\mathbb{R}^2$.

An inequality

Proposition

Let U and V be two subspaces of $\mathbb{R}^n$. Assume that $U \subset V$. Then

$$\dim(U) \leq \dim(V) \leq n.$$

If **moreover** $\dim(U) = \dim(V)$, then $U = V$.

Subspaces dimensions: Vocabulary

Definition

Let S be a subspace of $\mathbb{R}^n$.

- We call S a *line* if $\dim(S) = 1$.
- We call S an *hyperplane* if $\dim(S) = n - 1$.

3.3 Coordinates of a vector in a basis

Definition & Theorem

If $(v_1, \dots, v_n)$ is a basis of V, then for every $x \in V$ there exists a unique vector $(\alpha_1, \dots, \alpha_n) \in \mathbb{R}^n$ such that

$$x = \alpha_1 v_1 + \cdots + \alpha_n v_n.$$

We say that $(\alpha_1, \dots, \alpha_n)$ are the coordinates of x in the basis $(v_1, \dots, v_n)$.

Proof.

3.3 Coordinates of a vector in a basis

Exercise

1. Show that the vectors $v_1 = (1, 1)$ and $v_2 = (1, -1)$ form a basis of $\mathbb{R}^2$.
2. Express the coordinates of $u = (x, y)$ in the basis (v_1, v_2) in terms of x and y.

Exercise

1. Show that the vectors $v_1 = (1, 1)$ and $v_2 = (1, -1)$ form a basis of $\mathbb{R}^2$.
2. Express the coordinates of $u = (x, y)$ in the basis (v_1, v_2) in terms of x and y.

4. Why do we care about this ?

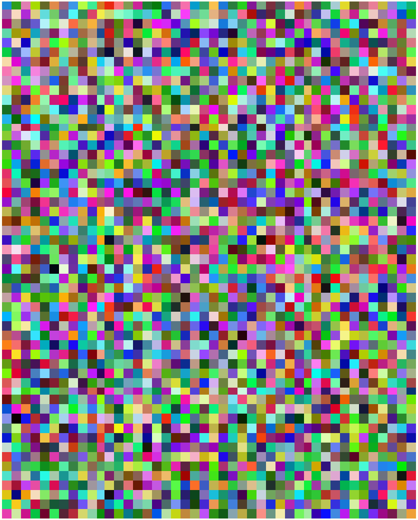
Application to image compression

- Image = Grid of pixels
- Represented as a vector $v \in \mathbb{R}^n$, for some large n.
- One needs to store n numbers.

$n = 44 \times 55 = 2420$

Can we do better?

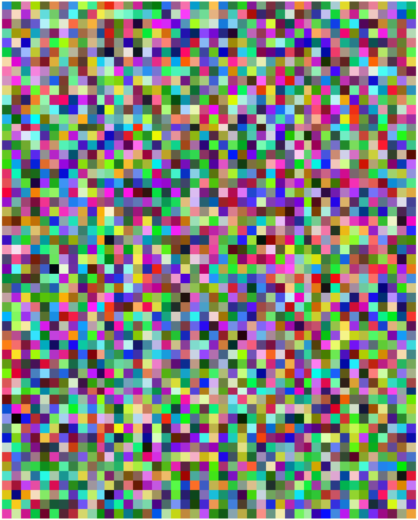
- If we want to store an arbitrary image, NO!



«Random» image

Can we do better?

- If we want to store an arbitrary image, NO!
- However, we are mainly storing images coming from the « real world »
- These images have some *structure*.



«Random» image

Can we do better?

- If we want to store an arbitrary image, NO!
- However, we are mainly storing images coming from the « real world »
- These images have some *structure*.

«Real» image

What do we mean by « structure » ?

Neighboring pixels are very likely to have similar colors.

- There exists a basis $(w_1, \ldots, w_n)$ of $\mathbb{R}^n$ in which «real» images $v \in \mathbb{R}^n$ are (approximately) **sparse**.
- This means that the coordinates $(\alpha_1, \ldots, \alpha_n)$ of v in the basis $(w_1, \ldots, w_n)$ contains a lot of zeros.

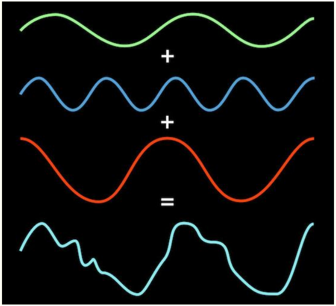
Store only the $k \ll n$ non-zero coordinates of v (in the w_i's basis') !

A toy example

Consider $n = 2$, that is images $v \in \mathbb{R}^2$ with only 2 pixels.
Take $v_1 = (1, 1)$ and $v_2 = (1, -1)$:

Examples of good bases

- **Fourier bases** (used in `.jpeg`, `.mp3`)



- `JPEG2000` uses **wavelet bases**, and achieves better performance than `JPEG`.
- The course **DS-GA 1013** deepens these concepts!

Questions?

Questions?

Questions?

Questions?