

# Session 8: SVD, linear algebra & graphs

Optimization and Computational Linear Algebra for Data Science

# Contents

1. Singular Value Decomposition
2. Graphs
3. Graph Laplacian
4. Spectral clustering

# Midterm next week

- ❖ Thu. Oct. 29, the questions have to be downloaded from Gradescope between 00:01 AM and 9:59 PM.
- ❖ **Duration:** 1 hour and 40 minutes to work on the problems + 20 minutes to scan and upload your work.
- ❖ Upload your work **as a single PDF**.
- ❖ In case the upload does not work for you, **email me your work**.

# Singular Value Decomposition

# Singular Value decomposition

## Theorem

Let  $A \in \mathbb{R}^{n \times m}$ . Then there exists two orthogonal matrices  $U \in \mathbb{R}^{n \times n}$  and  $V \in \mathbb{R}^{m \times m}$  and a matrix  $\Sigma \in \mathbb{R}^{n \times m}$  such that  $\Sigma_{1,1} \geq \Sigma_{2,2} \geq \dots \geq 0$  and  $\Sigma_{i,j} = 0$  for  $i \neq j$ , that verify

$$A = U\Sigma V^T.$$

# Comments

# Low-rank approximation

# Graphs



# Graphs, degree

# Graph Laplacian

# Graph Laplacian

## Definition

The Laplacian matrix of  $G$  is defined as

$$L = D - A.$$

# Graph Laplacian

## Definition

The Laplacian matrix of  $G$  is defined as

$$L = D - A.$$

For all  $x \in \mathbb{R}^n$ ,

$$x^\top Lx = \sum_{i \sim j} (x_i - x_j)^2.$$

# Properties of the Laplacian

For all  $x \in \mathbb{R}^n$ , 
$$x^\top Lx = \sum_{i \sim j} (x_i - x_j)^2.$$

# Properties of the Laplacian

For all  $x \in \mathbb{R}^n$ , 
$$x^\top Lx = \sum_{i \sim j} (x_i - x_j)^2.$$

# Properties of the Laplacian

For all  $x \in \mathbb{R}^n$ , 
$$x^\top Lx = \sum_{i \sim j} (x_i - x_j)^2.$$

# Algebraic connectivity

## Proposition

The graph  $G$  is connected if and only if  $\lambda_2 > 0$ .

More generally, the multiplicity of the eigenvalue 0 of  $L$  (i.e. the number of  $i$  such that  $\lambda_i = 0$ ) is equal to the number of connected components of  $G$ .

$\lambda_2$  is sometimes called the «algebraic connectivity» of  $G$  and measure (heuristically) how well  $G$  is connected.

**Exercise:** show that  $\lambda_2$  increases when one adds edges to  $G$ .



# Spectral clustering with the Laplacian

# Algorithm

**Input:** Graph Laplacian  $L$ , number of clusters  $k$

1. Compute the first  $k$  orthonormal eigenvectors  $v_1, \dots, v_k$  of the Laplacian matrix  $L$ .
2. Associate to each node  $i$  the vector  $x_i = (v_1(i), \dots, v_k(i))$ .
3. Cluster the points  $x_1, \dots, x_n$  with (for instance) the  $k$ -means algorithm.
4. Deduce a clustering of the nodes of the graph.

# The case of two groups

**For  $k = 2$  groups:**

1. Compute the second eigenvector  $v_2$  of the Laplacian matrix  $L$ .
2. Associate to each node  $i$  the number  $x_i = v_2(i)$ .
3. Cluster the nodes in:

$$S = \{i \mid v_2(i) \geq \delta\} \quad \text{and} \quad S^c = \{i \mid v_2(i) < \delta\},$$

for some  $\delta \in \mathbb{R}$ .

# Cut of a partition

# Minimal cut problem

# Spectral clustering as a «relaxation»

# Questions?

# Questions?