

# Understanding Wasserstein t-SNE

University of Tübingen

Fynn Bachmann

24.09.2021

# Contents

<b>1</b>	<b>Introduction: Visualizing Structure in Datasets</b>	<b>3</b>
1.1	Methods . . . . .	3
1.1.1	PCA . . . . .	3
1.1.2	t-SNE . . . . .	3
1.2	Distances . . . . .	3
1.2.1	Euclidean . . . . .	3
1.2.2	Wasserstein . . . . .	3
<b>2</b>	<b>Theory: Can there be structure in Covariances?</b>	<b>4</b>
2.1	Synthetic Data . . . . .	4
2.1.1	Hierarchical Gaussian Mixture . . . . .	4
2.1.2	Gaussian Wasserstein Distance . . . . .	4
2.1.3	Convex Interpolation Method . . . . .	4
<b>3</b>	<b>Analysis: Structure in Real-World-Dataset Covariances?</b>	<b>5</b>
3.1	German Election 2017/2021 . . . . .	5
3.1.1	Understanding the Covariance Embedding . . . . .	5
3.1.2	Results . . . . .	5
3.2	European Value Study 2017-2020 . . . . .	5
3.2.1	Discrete Data . . . . .	5
3.2.2	Logit Transformation . . . . .	5
3.3	Big Five Personality Traits Survey . . . . .	5
3.3.1	Results . . . . .	5
<b>4</b>	<b>Outlook: Computing Exact Wasserstein Distances</b>	<b>6</b>
4.1	Linear Programming . . . . .	6
4.1.1	Scalability in Participants . . . . .	6
4.1.2	Scalability in Features . . . . .	6

4.2	Exact Wasserstein Distance Embedding . . . . .	6
4.2.1	Complexity Analysis . . . . .	6
4.2.2	Comparision to Gaussian Approximation . . . . .	6

# Chapter 1

## Introduction: Visualizing Structure in Datasets

### 1.1 Methods

#### 1.1.1 PCA

#### 1.1.2 t-SNE

### 1.2 Distances

#### 1.2.1 Euclidean

#### 1.2.2 Wasserstein

# Chapter 2

## Theory: Can there be structure in Covariances?

### 2.1 Synthetic Data

#### 2.1.1 Hierarchical Gaussian Mixture

### 2.2 Gaussian Wasserstein Distance

#### 2.2.1 Convex Interpolation Method

# Chapter 3

## Analysis: Structure in Real-World-Dataset Covariances?

### 3.1 German Election 2017/2021

#### 3.1.1 Understanding the Covariance Embedding

#### 3.1.2 Results

### 3.2 European Value Study 2017-2020

#### 3.2.1 Discrete Data

#### 3.2.2 Logit Transformation

### 3.3 Big Five Personality Traits Survey

#### 3.3.1 Results

# Chapter 4

## Outlook: Computing Exact Wasserstein Distances

### 4.1 Linear Programming

#### 4.1.1 Scalability in Participants

#### 4.1.2 Scalability in Features

### 4.2 Exact Wasserstein Distance Embedding

#### 4.2.1 Complexity Analysis

#### 4.2.2 Comparision to Gaussian Approximation