Learning the low-dimensional geometry of the wireless channel

Paul Ferrand, Alexis Decurninge, Luis Garcia Ordóñez and Maxime Guillaud
Mathematical and Algorithmic Sciences Lab, Paris Research Center, Huawei Technologies
Email: {paul.ferrand,alexis.decurninge,luis.ordonez,maxime.guillaud}@huawei.com

Introduction

Background on Channel State Information (CSI)

- CSI data as seen from the baseband signal processing perspective is a random process in a high dimensional space (large number of receiving antennas, large sampling in the frequency domain, etc.)
- Physical considerations (e.g. ray-tracing channel models) hint at an unknown low-dimensional manifold structure of the CSI - roughly speaking, the CSI is a continuous function of a limited number of geometric and RF propagation parameters

Objectives

- Learning the geometrical structure of CSI data *while involving as little expert knowledge as possible*
- *Data-driven* identification of the parameterizations or features adapted to estimation, compression, prediction of CSI

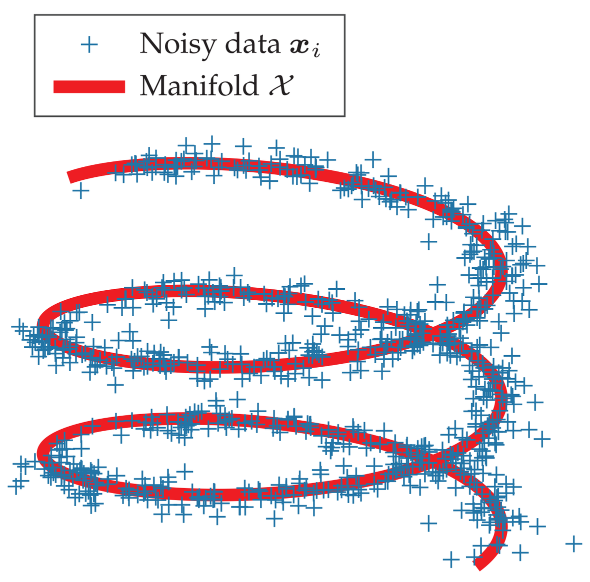
Dimensionality reduction of CSI

Assumptions on data manifold

- $\mathcal{X} \subset \mathbb{R}^D$ is a d-dimensional and continuous set in the Euclidean space $\mathbb{R}^D$ with $D \gg d$
- We suppose that a generative model exists through some unknown continuous mapping $g : \mathbb{R}^d \mapsto \mathcal{X}$ and seek for a mapping $f : \mathbb{R}^D \mapsto \mathbb{R}^d$ so that $f \approx g^{-1}$

$$\underbrace{\begin{matrix}\mathbb{R}^d & \rightarrow \\ \boldsymbol{w}_i & \overset{g}{\mapsto}\end{matrix}}_{\text{Generative model}} \quad \underbrace{\begin{matrix}\mathcal{X} \subset \mathbb{R}^D \\ \boldsymbol{x}_i\end{matrix}}_{\text{Noise-free observations}} \quad \underbrace{\begin{matrix}\rightarrow & \mathbb{R}^d \\ \overset{f}{\mapsto} & \boldsymbol{w}'_i\end{matrix}}_{\text{Dimensionality Reduction}}$$

- In practice, the observations $\boldsymbol{x}_i$ are noisy in the ambient space



Toy Example of a 1-dimensional manifold embedded in $\mathbb{R}^3$.

Dimensionality Reduction

Principle: use distance between samples in order to learn $f \approx g^{-1}$ up to an isometry.
General additional assumption on the sample indices $\mathcal{S} = \{1, \dots, T\}$:

$$\|\boldsymbol{w}'_i - \boldsymbol{w}'_j\| \approx d(\boldsymbol{x}_i, \boldsymbol{x}_j) \quad \text{for all } (i,j) \in \mathcal{T} \subset \mathcal{S} \times \mathcal{S} \tag{1}$$

Non-parametric dimensionality reduction

- $d(\boldsymbol{x}_i, \boldsymbol{x}_j) = \|\boldsymbol{x}_i - \boldsymbol{x}_j\|$, $\mathcal{T} = \mathcal{S} \times \mathcal{S}$: multi-dimensional scaling (MDS).
- $d(\boldsymbol{x}_i, \boldsymbol{x}_j) = \|\boldsymbol{x}_i - \boldsymbol{x}_j\|$, $\mathcal{T} = \{(i,j) \in \mathcal{S} \times \mathcal{S} : \|\boldsymbol{x}_i - \boldsymbol{x}_j\| \leq \eta\}$: Isomap, UMAP

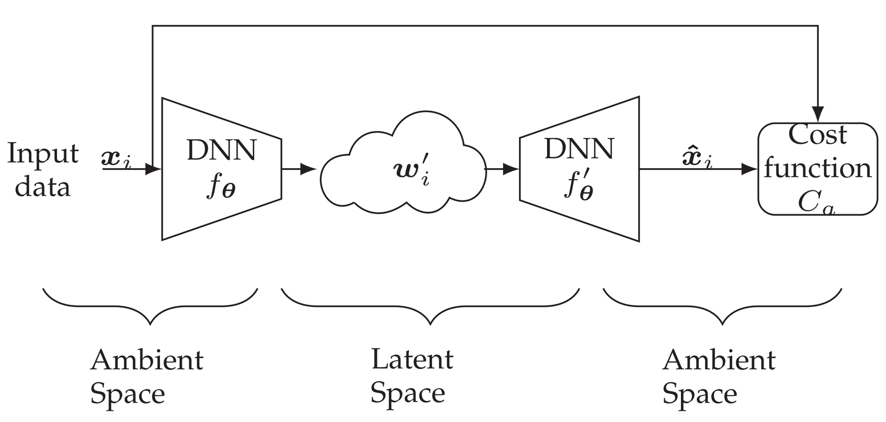
Main limitations

- Euclidean distance may not be adapted to the manifold $\mathcal{X}$
- No explicit estimation of the mapping f: difficult out-of-sample extrapolation

Parameteric dimensionality reduction: $f(\boldsymbol{x}_i) = f_{\boldsymbol{\theta}}(\boldsymbol{x}_i)$ with $\mathcal{T} = \mathcal{S} \times \mathcal{S}$

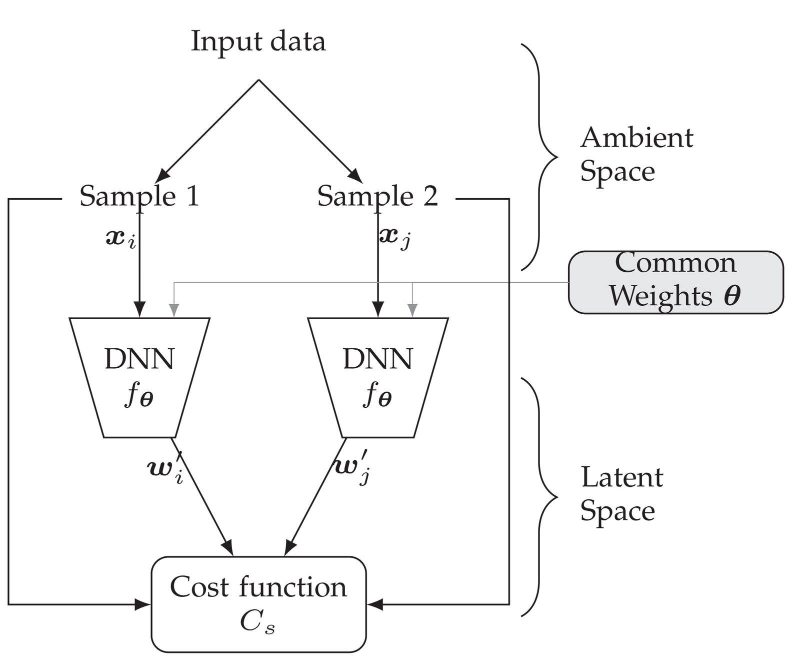
- Distance in the ambient space parameterized by $\boldsymbol{\theta}$

$$d_{\boldsymbol{\theta}}(\boldsymbol{x}_i, \boldsymbol{x}_j) = \|f_{\boldsymbol{\theta}}(\boldsymbol{x}_i) - f_{\boldsymbol{\theta}}(\boldsymbol{x}_j)\|$$



Autoencoder

$$\boldsymbol{\theta} = \arg\min_{\boldsymbol{\theta}} \sum_{i \in \mathcal{S}} \|\boldsymbol{x}_i - f_{\boldsymbol{\theta}'} \circ f_{\boldsymbol{\theta}}(\boldsymbol{x}_i)\|^2$$



Siamese network

$$\boldsymbol{\theta} = \arg\min_{\boldsymbol{\theta}} \sum_{(i,j) \in \mathcal{T}} \left(\|\boldsymbol{x}_i - \boldsymbol{x}_j\| - d_{\boldsymbol{\theta}}(\boldsymbol{x}_i, \boldsymbol{x}_j)\right)^2$$

Triplets network

- Triplets networks consider a weaker assumption wrt (1)

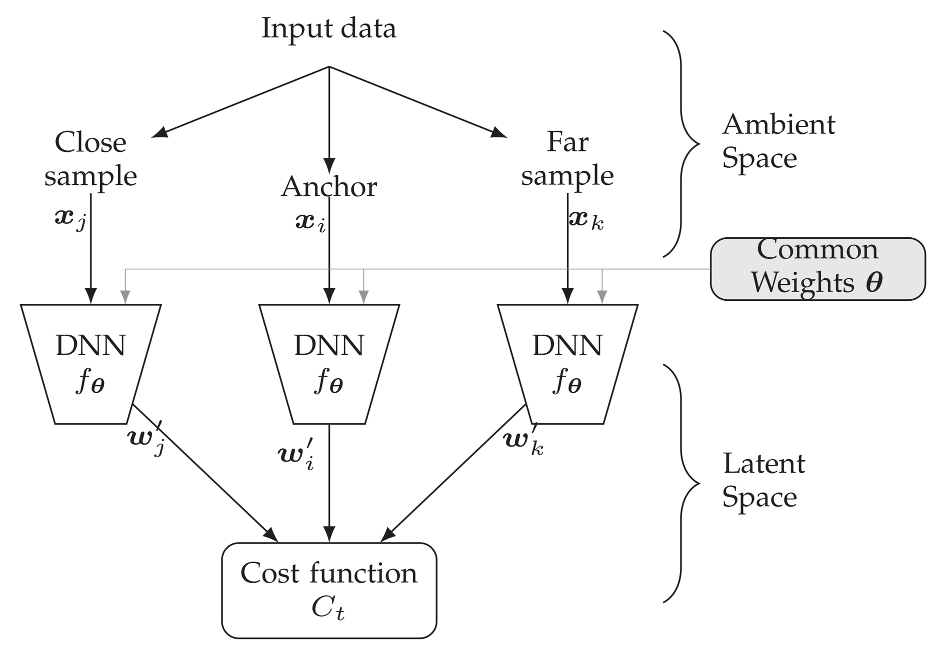
$$d(\boldsymbol{x}_j, \boldsymbol{x}_i) \leq d(\boldsymbol{x}_k, \boldsymbol{x}_i) \Rightarrow \|\boldsymbol{w}'_j - \boldsymbol{w}'_i\| \leq \|\boldsymbol{w}'_k - \boldsymbol{w}'_i\| \quad \text{for all } (i,j,k) \in \mathcal{T} \subset \mathcal{S}^3 \tag{2}$$

- In practice, we use side-information (timestamps) to define $\mathcal{T}$, i.e. triplets (i,j,k) so that

$$\left.\begin{matrix} 0 < |t_j - t_i| \leq T_c \\ T_c < |t_k - t_i| \leq T_f \end{matrix}\right\} \Rightarrow d(\boldsymbol{x}_j, \boldsymbol{x}_i) \leq d(\boldsymbol{x}_k, \boldsymbol{x}_i) \Rightarrow \|\boldsymbol{w}'_j - \boldsymbol{w}'_i\| \leq \|\boldsymbol{w}'_k - \boldsymbol{w}'_i\| \tag{3}$$

- Triplets network does not rely on the Euclidean distance on the ambient space minimizing

$$\boldsymbol{\theta} = \arg\min_{\boldsymbol{\theta}} \sum_{(i,j,k) \in \mathcal{T}} \left(d_{\boldsymbol{\theta}}(\boldsymbol{x}_i, \boldsymbol{x}_j) - d_{\boldsymbol{\theta}}(\boldsymbol{x}_i, \boldsymbol{x}_k) + M\right)^+$$



Experimental Results

Experimental Dataset

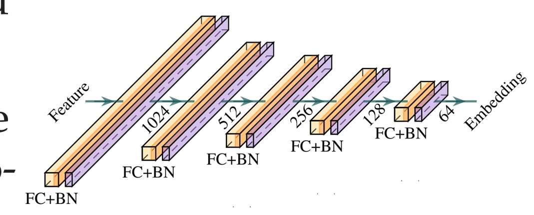
Several hours of joint CSI/position measurements with a commercial user equipment (UE)

- In total, of 3.5×10^6 CSI samples (sampling every 10ms)
- $\boldsymbol{h}_t \in \mathbb{C}^{64 \times 288}$ (64 receive antennas, 288 frequency subcarriers, $64 \times 288 = 18432$)

Expert pre-processing

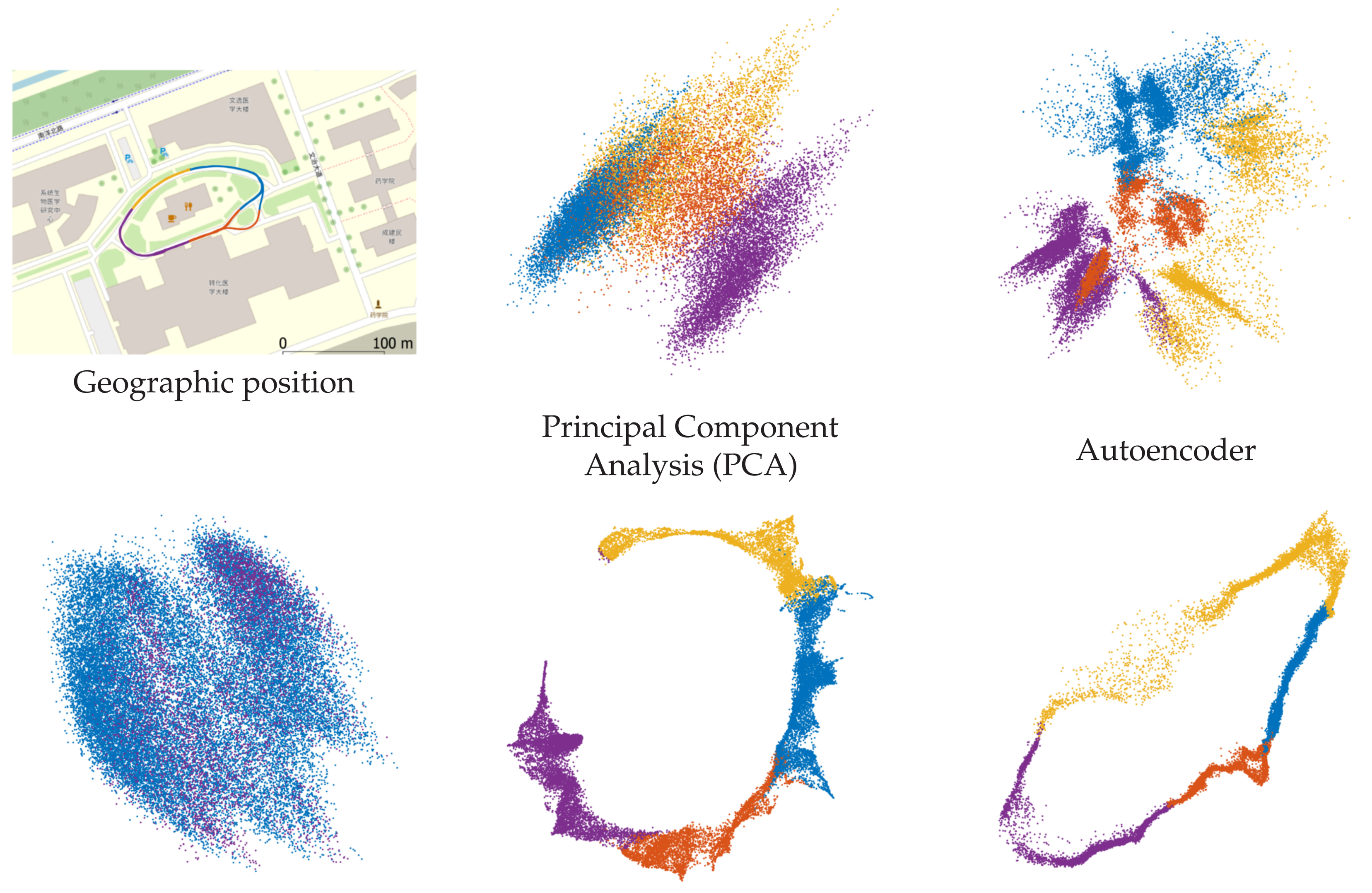
- Some CSI characteristics are not of interest: e.g. clock and frequency offset between the UE and the base station
- In order to circumvent these impairments and to reduce observation noise, we map the observations through a "projector" invariant wrt these characteristics

$$\boldsymbol{x}_t = p(\boldsymbol{h}_t) \in \mathbb{R}^{1024} \quad \text{(i.e. } D = 1024\text{)}$$



Used Neural Network

Dimensionality Reduction Results ($d = 2$)



Geographic position
Principal Component Analysis (PCA)
Autoencoder
Siamese network
UMAP
Triplets (margin cost)

Semi-supervised extension

We generalize the triplets network criterion with known positions $\boldsymbol{p}_i \in \mathbb{R}^2$

$$\theta = \arg\min_{\theta} \frac{1}{|\mathcal{T}|} \sum_{(i,j,k) \in \mathcal{T}} \left(d_{\boldsymbol{\theta}}(\boldsymbol{x}_i, \boldsymbol{x}_j) - d_{\boldsymbol{\theta}}(\boldsymbol{x}_i, \boldsymbol{x}_k) + M\right)^+ + \frac{\alpha}{|\mathcal{P}|} \sum_{i \in \mathcal{P}} \|f_{\boldsymbol{\theta}}(\boldsymbol{x}_i) - \boldsymbol{p}_i\|$$

Geographic position
Unsupervised ($\alpha = 0, M = 1$)
Semi-supervised ($\alpha = 1$, $M = 1, |\mathcal{P}| = 45$)