Rank Priors for Non-Linear Dimensionality Reduction

Andreas Geiger [geiger@mrt.uka.de]
Department of Measurement and Control
Karlsruhe Institute of Technology

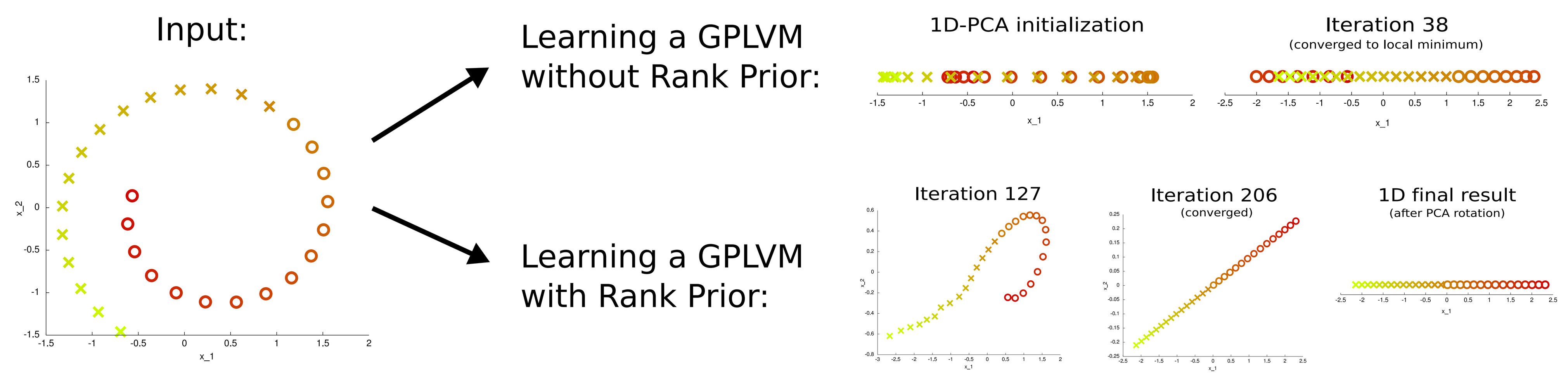
Raquel Urtasun [rurtasun@icsi.berkeley.edu]
UC Berkeley, EECS & ICSI - MIT CSAIL

Trevor Darrell [trevor@eecs.berkeley.edu]
UC Berkeley, EECS & ICSI - MIT CSAIL

Abstract

Iterative non-linear dimensionality reduction methods are often susceptible to local minima and perform poorly when initialized far from the global optimum. In this work we introduce a prior over the dimensionality of the latent space that penalizes high dimensional spaces, and simultaneously optimize both the latent space and its intrinsic dimensionality in a continuous fashion. We initialize the latent space to the observation space and automatically infer the latent dimensionality. We report results applying our prior to various probabilistic non-linear dimensionality reduction tasks, and show that our method can outperform other techniques as well as previously suggested initialization strategies.

Illustration:
2D swiss roll



Problem

- High dimensional datasets are challenging to analyze
- It is desirable to reduce the dimensionality of the data
- This allows for more efficient learning and inference

Existing approaches

- Linear dimensionality reduction techniques (e.g., PCA, PPCA)
  - \+ Simple
  - \+ Closed form solution
  - \+ Computationally efficient
  - \- Can not deal with complex datasets
  - \- Bad results in non-linear cases (human motion)
- Graph-based methods (e.g., LLE, Isomap)
  - \+ Often convex
  - \+ Computationally efficient
  - \+ Can handle non-linearities
  - \- Suffer in presence of noisy or sparse data
  - \- Data must be sampled homogeneously
  - \- Correct neighborhood size k is crucial
- Non-Linear probabilistic models (e.g., GPLVM)
  - \+ Can handle non-linearities
  - \+ Probabilistic solution
  - \- Local minima due to non-convexity
  - \- Bad initialization leads to bad results
  - \- Only applied to small databases of single activities

Continuous Dimensionality Reduction

Gaussian Process Latent Variable Likelihood

$$p(\mathbf{Y}|\mathbf{X}) = \prod_{d=1}^{D} \mathcal{N}(\mathbf{Y}^{(d)}|\mathbf{0},\mathbf{K})$$

Latent variables; Conditional independence assumption; Observations (e.g. angular positions); Covariance matrix

⊕

Rank Prior over latent space

$$p(\mathbf{X}) = \frac{1}{Z}\exp\left(-\alpha\sum_{i=1}^{D}\phi(s_i)\right)$$

Regularization factor; i'th singular value of X; Normalization constant; Penalty function (e.g., logarithmic, exponential, power family, L1, L2)

⊜

MAP optimization (1)

$$\mathcal{L} = \frac{D}{2}\ln|\mathbf{K}| + \frac{D}{2}tr(\mathbf{K}^{-1}\mathbf{Y}\mathbf{Y}^T) + \alpha\sum_{i=1}^{D}\phi(s_i)$$

$$\min\mathcal{L} \quad s.t. \ \ \forall i\ s_i \geq 0, \quad \Delta E = 0$$

Gaussian Process LVM

Latent Variable models (LVMs) assume that the observed data Y is generated by some latent variables X:

$$y^{(d)} = f(\mathbf{x}) + \eta \qquad \text{with} \qquad \eta \sim \mathcal{N}(0,\theta_3)$$

The GPLVM places a GP prior over the space of mapping functions f. Assuming conditional independence and marginalizing over f yields

$$p(\mathbf{Y}|\mathbf{X}) = \prod_{d=1}^{D}\mathcal{N}(\mathbf{Y}^{(d)}|\mathbf{0},\mathbf{K})$$

Our approach

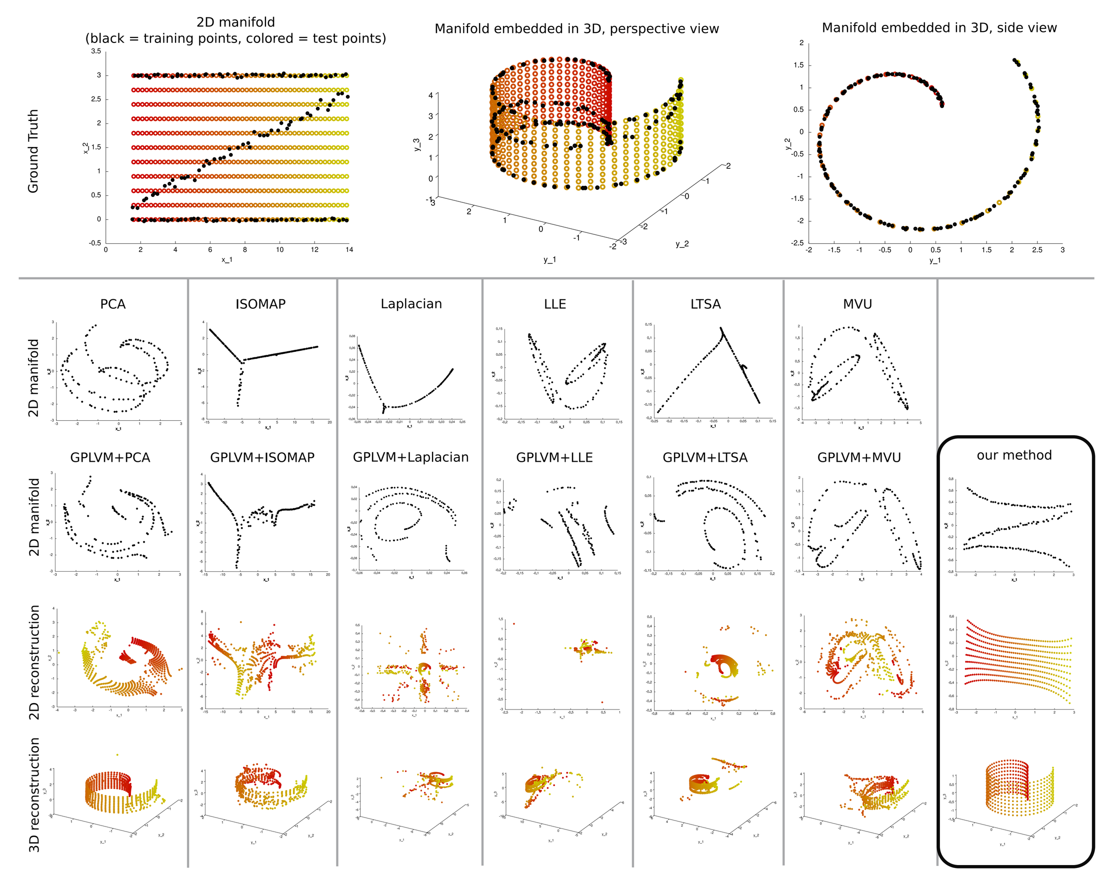
- Reduce the problem of local minima by performing continuous dimensionality reduction
- No distortion is introduced since we initialize the latent coordinates to the observations
- By introducing a prior over the dimensionality estimate simultaneously the latent space and its dimensionality.

Algorithm overview

1. Initialize latent space to observation space (X=Y)
2. Optimize dimensionality and latent variables in a continuous fashion using equation (1)
3. After convergence apply PCA to remove dimensions associated with small singular values
4. Keep X fixed and optimize hyperparamters of K

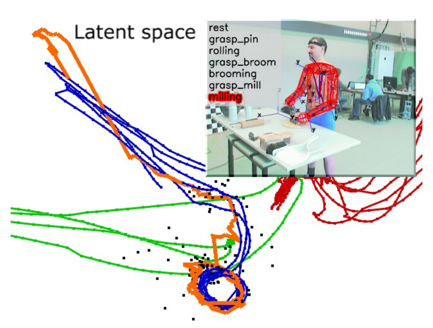
Experiment 1: Sparse 3D swiss roll

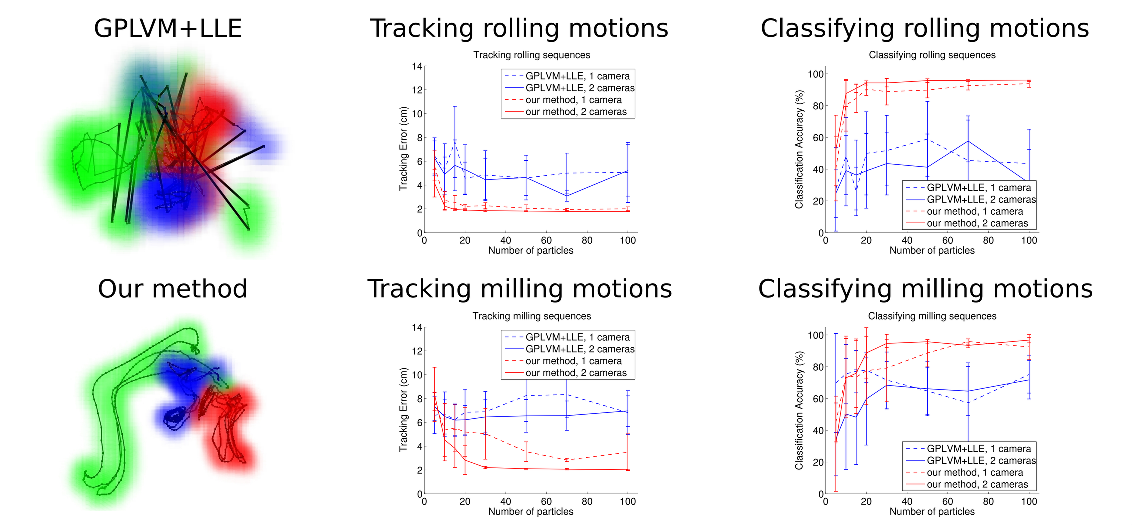
This experiment compares our method to the Gaussian Process Latent Variable Model (GPLVM) initialized with different methods (PCA, ISOMAP, Laplacian, LLE, LTSA, MVU) and parameters on the task of finding a 2D manifold on a sparsely sampled swiss roll, embedded in 3D.



Experiment 3: Upper body tracking

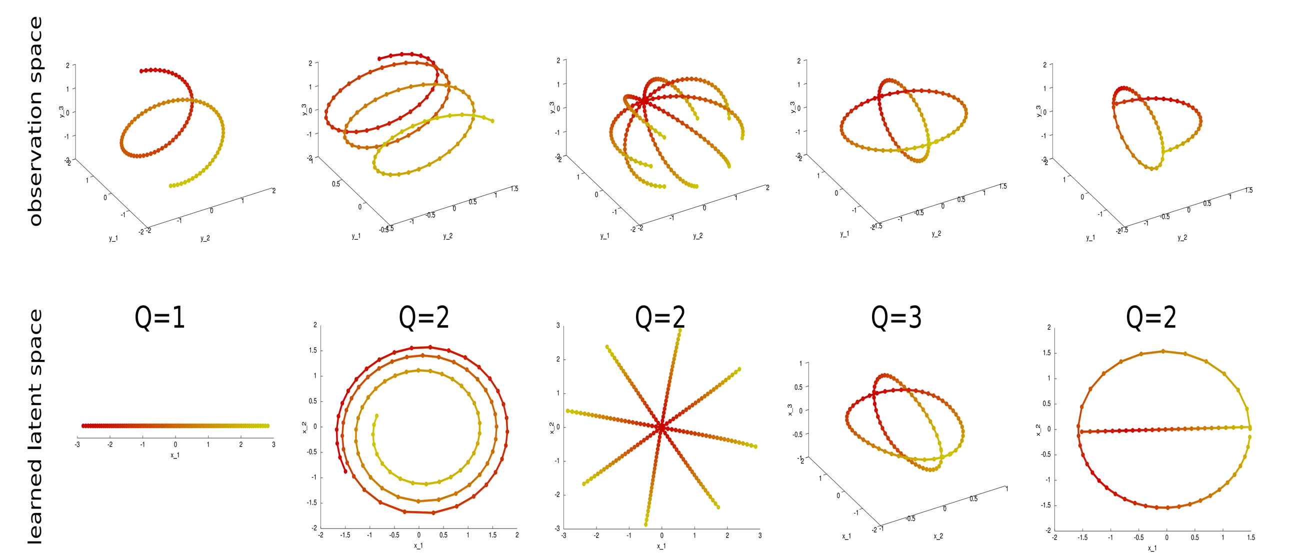
This figure depicts learning of different types of motion into one single 3D latent space using GPLVM +PCA (top-left) and our method (bottom-left). We used this model for tracking human body motion in a kitchen scenario with a particle filter.





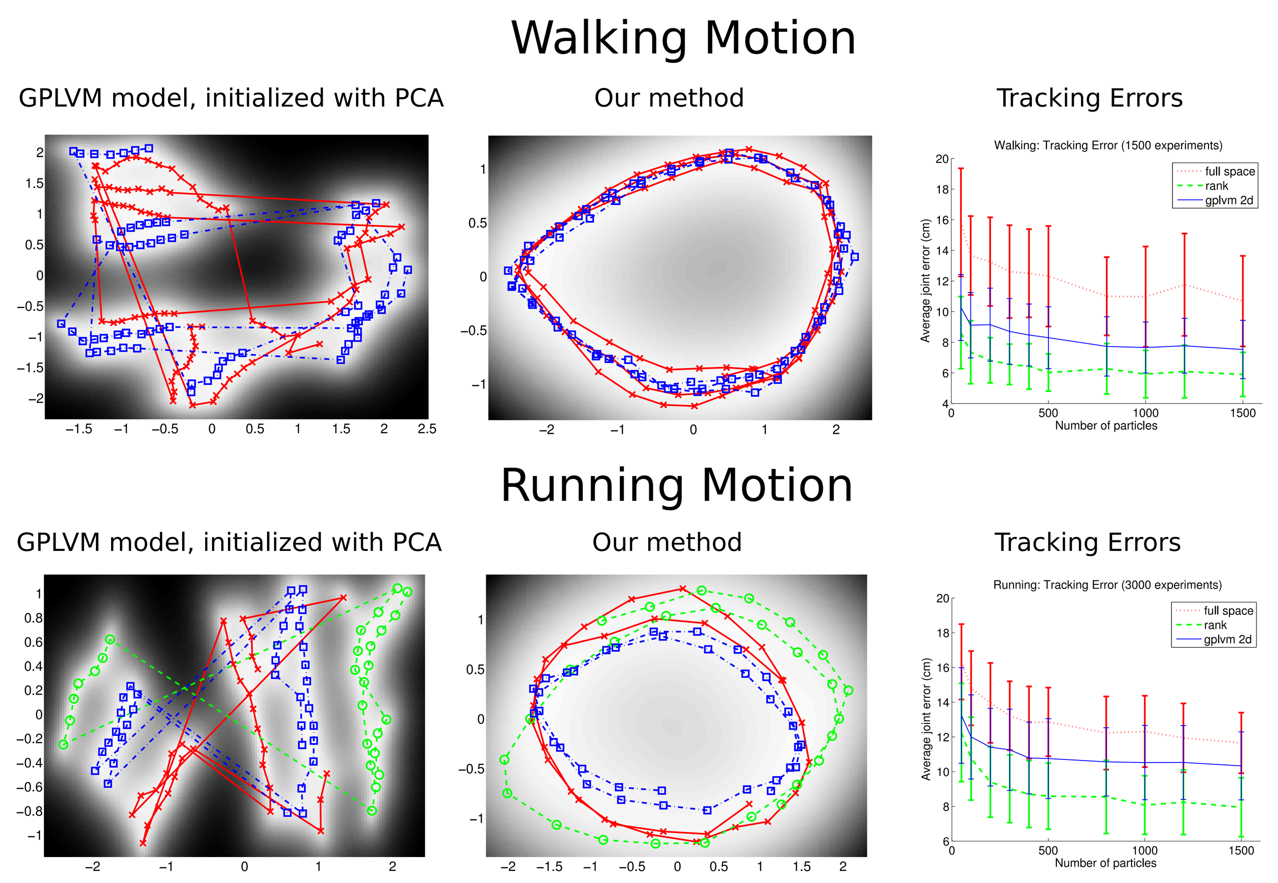
Experiment 2: Discover latent dimensionality

In this experiment we demonstrate the ability of our method to estimate the correct underlying latent dimensionality. Five artificial examples have been created and reduced using our method. The latent dimensionality Q has been picked according to the number of non-zero singular values of X after the optimization procedure.



Experiment 4: Tracking running and walking

This figure depicts the latent spaces of running (top) and walking (bottom) motions from 2D mocap data. The right column compares tracking errors (a particle filter was used) averaged over 10 splits. We compare tracking a short sequence in full parameter space (red) to the GPLVM initialized with PCA (blue) and our method (green).



In our experiments we used motion data from: http://mocap.cs.cmu.edu

University of California, Berkeley: Electrical Engineering & Computer Science (EECS), International Computer Science Institute (ICSI)
Karlsruhe Institute of Technology (KIT): Department of Measurement and Control (MRT), Karlsruhe School of Optics and Photonics (KSOP), Deutsche Forschungsgemeinschaft (DFG)