

# Dimensionality Reduction: A Probabilistic Perspective

Aditya Ravuri   Neil D. Lawrence



## Background

### Dimensionality Reduction (DR)

**Premise:** high dimensional data sitting on low dimensional manifold.

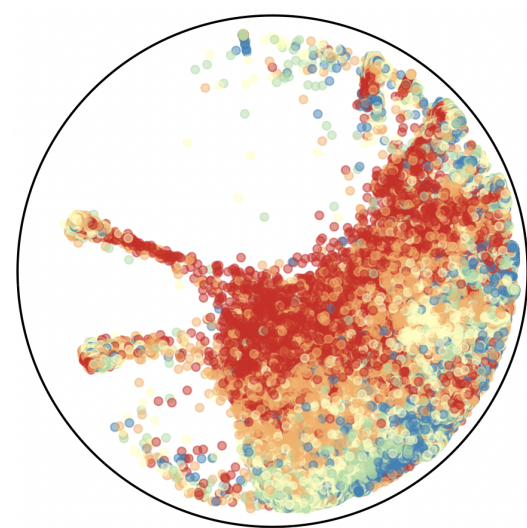


Figure 1. gene expression projected into a Poincaré disk showing cell differentiation.

Often a need for:

- uncertainty quantification
- incorporation of prior knowledge (e.g. confounders)
- identifiability and interpretability

### Existing Interpretations

It's known that **PCA, FA, ICA, CMDS** algorithms correspond to MLE in

$$\mathbf{Y}|\mathbf{X} \sim \mathcal{N}(\mathbf{0}, \mathbf{X}\mathbf{X}^T + \sigma^2\mathbf{I}), \quad (1)$$

with different priors (isotropic, non-isotropic and non-Gaussian) on  $\mathbf{X}$  each with significant implications (maximal variance, causal discovery).

Are there probabilistic interpretations to other DR methods? ... Yes!

## Case Study A: Eigencomponent based DR

Note that,

1. Non-constant likelihood terms in eq. (1) are equivalent to those of,

$$\mathbf{S}|\mathbf{X} \sim \mathcal{W}(\mathbf{X}\mathbf{X}^T + \sigma^2\mathbf{I}, d), \quad (2)$$

where  $\mathbf{S} \equiv \mathbf{Y}\mathbf{Y}^T$  is the sample covariance, **an estimate**.

2. Eigencomponents result in MLEs for latents in such models.

EC	interpretation	Cov( $\mathbf{S} \mathbf{X}$ )	Cov( $\mathbf{S}' \mathbf{W}$ )
Major	$\mathbf{Y} = \mathbf{X}\mathbf{W} + \epsilon$ ; cov. linear	$\mathbf{X}\mathbf{X}^T + \sigma^2\mathbf{I}$	$\mathbf{W}\mathbf{W}^T + \sigma^2\mathbf{I}$
Minor	precision linear in $\mathbf{X}/\mathbf{W}$	$(\mathbf{X}\mathbf{X}^T + \beta\mathbf{I})^{-1}$	$(\mathbf{W}\mathbf{W}^T + \beta\mathbf{I})^{-1}$

**Our interpretation:** Any DR method outputting embeddings as eigencomponents of PSD matrices can be interpreted as two step inference processes, where one:

1. constructs an estimator of a covariance matrix  $\hat{\mathbf{S}}$  or a graph Laplacian  $\hat{\mathbf{L}}$ .
2. obtains an emedding through MLE in the class of models above (minor eig.s if precision, major if covariance).

A list of our interpretations:

- Laplacian Eigenmaps, Spectral Embeddings:

$$\hat{\mathbf{L}}(\mathbf{Y})_{ij} = -k(\mathbf{Y}_i, \mathbf{Y}_j)$$

- LLE:

$$\hat{\mathbf{L}}(\mathbf{Y}) = \text{MLE in a locally linear GMRF}$$

- Kernel PCA:

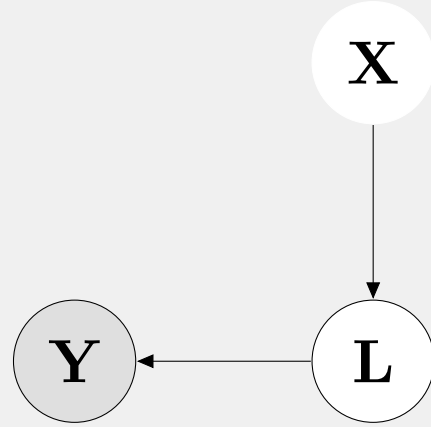
$$\hat{\mathbf{S}}(\mathbf{Y}) = k(\mathbf{Y}_i, \mathbf{Y}_j)$$

- Diffusion Maps ( $\alpha = 1$ ): (this is a Matérn- $\infty$  graph GP covariance)

$$\hat{\mathbf{S}}(\mathbf{Y}) = \exp[-t\Delta(\mathbf{Y})]$$

## Case Study B: VI based DR

Generative Model



Variational Approximation



Figure 2. a class of models whose ELBOs give rise to (t-)SNE and UMAP objectives

**(t-)SNE & UMAP** also correspond to models that estimate a graph Laplacian and tie it to a generative model, but correspond to variational inference (VI) algorithms.

ELBOs here take the form `constant - cost_func_of_algo`:

$$\mathcal{L} = \mathbb{E}_{q(\mathbf{L}|\mathbf{Y})}[\log p(\mathbf{Y}|\mathbf{L})] - \text{KL}(q(\mathbf{L}|\mathbf{Y})||p(\mathbf{L}|\mathbf{X}))$$

The variational & model probabilities  $q(\mathbf{A}|\mathbf{Y}), p(\mathbf{A}|\mathbf{X})$  factorize as,

$$\begin{cases} \prod_{i \neq j} \text{Bernoulli}(\mathbf{A}_{ij}|\{\mathbf{v} \text{ or } \mathbf{w}\}_{ij}^U) & \text{if UMAP} \\ \prod_i \text{Categorical}(\mathbf{A}_i|\{\mathbf{v} \text{ or } \mathbf{w}\}_{ij}^S) & \text{if (t-)SNE} \end{cases}$$

Methods in literature approximating components above using neural networks are simple extensions.

Graph GP generative models on  $\mathbf{L} \rightarrow \mathbf{Y}$  can outperform GPLVMs, VAEs.

## Applications & Extensions

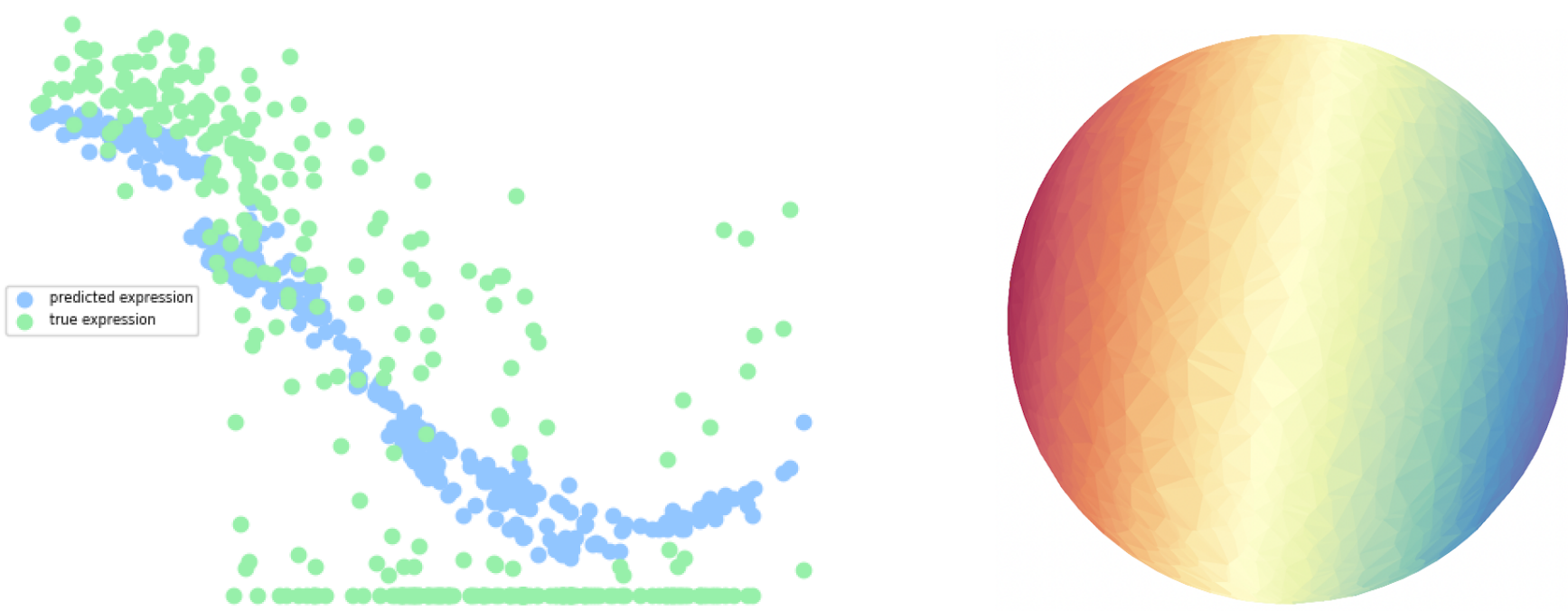


Figure 3. **left:** predicted gene expression given latents using a graph GP generative model. **right:** illustration of convergence of eigenvectors of  $\mathbf{L}$  to eigenfunctions of  $\Delta$  - thus hinting at a convergence of  $\mathbf{X} \rightarrow \mathbf{L} \rightarrow \mathbf{Y}$  to a Matérn GP on a manifold.

## The Why & Connections to other projects

Probabilistic interpretations:

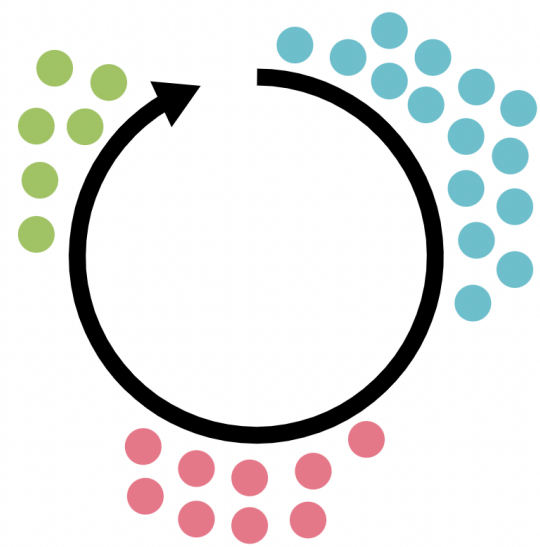


Figure 4. enable efficient reformulations  
enable prior specification  
e.g. RE models are GPs.

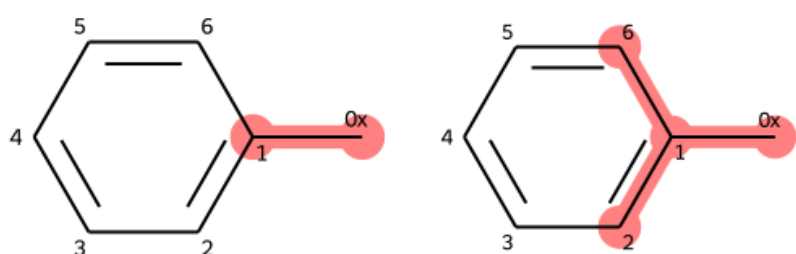


Figure 5. enable reasoning about uncertainties and methods like bayesopt  
e.g. graph kernels are deep (gnn) kernels

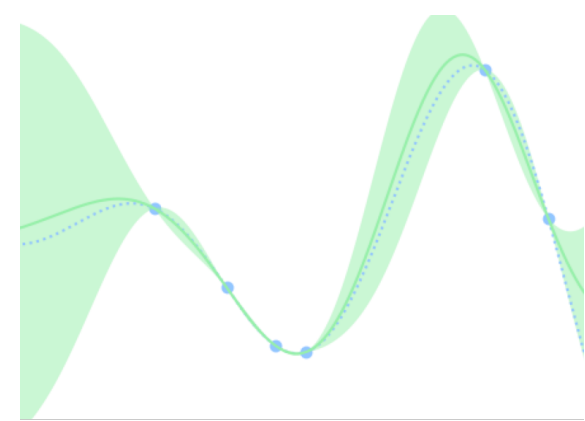


Figure 6. aid communication  
enable software cross-utilization  
e.g. GAMs are GPs

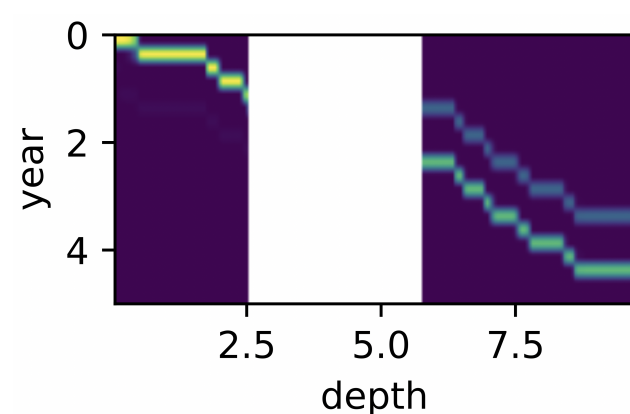


Figure 7. enable model extension  
enable alternative inference methods  
e.g. DTW is (cts-)HMM inference