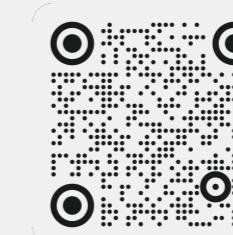


# Dimensionality Reduction: a Probabilistic Perspective

Aditya Ravuri

Neil D. Lawrence

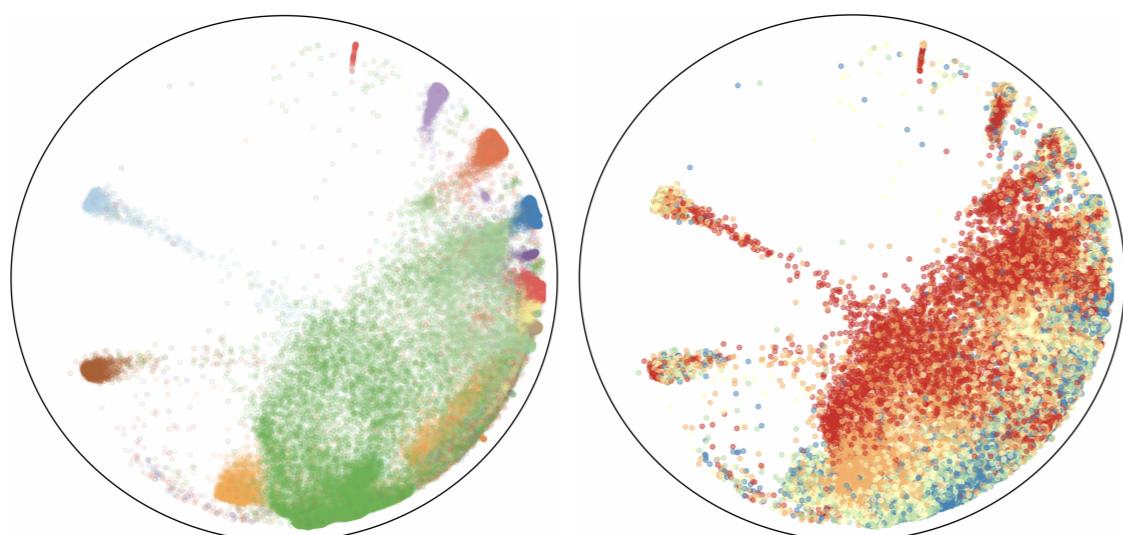


UNIVERSITY OF  
CAMBRIDGE

## Background

### Dimensionality Reduction (DR)

High dimensional data  $\mathbf{Y}$  can sit near low dimensional manifolds. DR algos estimate the low dimensional positions  $\mathbf{X}$ .



gene expression projected into a Poincaré disk showing cell differentiation  
**left:** colored by cell type, **right:** colored by pseudotime

With DR methods, often a need for:

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

**Contribution:** We show that DR algos perform probabilistic inference, enabling the applications above.

### Background

PCA, FA, GMM, NMF, LDA, ICA algorithms correspond to MLE in models of the form below [Murphy, 2023]. Different distributions impose different implications, e.g. identifiability.

$$\mathbf{x} \sim \text{prior} \quad (1)$$

$$\mathbf{y}|\mathbf{x} \sim \text{exp\_family\_dist}(\mathbf{x}) \quad (2)$$

### a: Two Step MLE (eigencomponent-based DR)

MLEs in certain models over PSD matrices occur at eigencomps

Due to [Lawrence 2005], [Tipping 1999] and [Williams 2002], we show that MLE in the following models results in major and minor eigencomponents of  $\mathbf{S}$  and  $\mathbf{\Gamma}$  respectively.

$$\text{major: } \mathbf{S}|\mathbf{X} \sim \mathcal{W}(\mathbf{XX}^T + \sigma^2 \mathbf{I}, d) \quad (3)$$

$$\text{minor: } \mathbf{\Gamma}|\mathbf{X} \sim \mathcal{W}((\mathbf{XX}^T + \beta \mathbf{I})^{-1}, d) \quad (4)$$

We prove consistency between these results.

Any algorithm that sets  $\mathbf{X} = \text{eigencomps}$  is hence a 2-step process:

1. that estimates a covariance  $\hat{\mathbf{S}}(\mathbf{Y})$  or precision  $\hat{\mathbf{\Gamma}}(\mathbf{Y})$ ,
2. and obtains an embedding through MLE using the set of models in eq. (3).

algorithm	$\mathbf{M}(\mathbf{Y})$	computation
PCA	$\mathbf{S}$	$\mathbf{YY}^T/d$
Lap. Eigenmaps, Spec. Emb.	$\mathbf{\Gamma}$	$-w(\mathbf{Y}, \mathbf{Y}')$
Locally Lin. Emb.	$\mathbf{\Gamma}$	MLE in a GMRF
kPCA	$\mathbf{S}$	$k(\mathbf{Y}, \mathbf{Y}')$
Diff. Maps	$\mathbf{S}$	$\exp[-t\mathbf{L}(\mathbf{Y})]$

Lastly, we prove that Two Step MLE is equivalent to ProbDR VI  
We prove that,

$$\arg \max_{\mathbf{X}} \log \mathcal{W}(\hat{\mathbf{S}}/d | \mathbf{M}(\mathbf{X}), d)$$

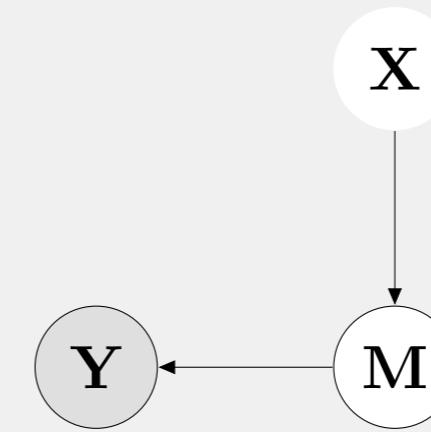
is equivalent to,

$$\arg \min_{\mathbf{X}} \text{KL}(\mathcal{W}(\mathbf{S} | \hat{\mathbf{S}}, d) \parallel \mathcal{W}(\mathbf{S} | \mathbf{M}(\mathbf{X}), d))$$

which is the KL divergence found in the ProbDR objective.

### b: The ProbDR Interpretation & WeirdVI

Generative Model



Variational Approximation



the ProbDR framework that gives rise to many DR algos

ELBOs here take the form `constant - cost_func_of_dr_algo`:

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

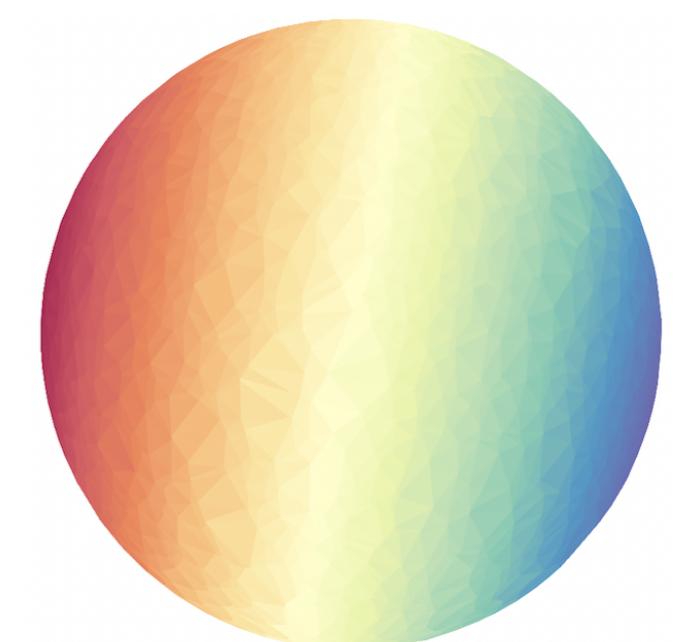
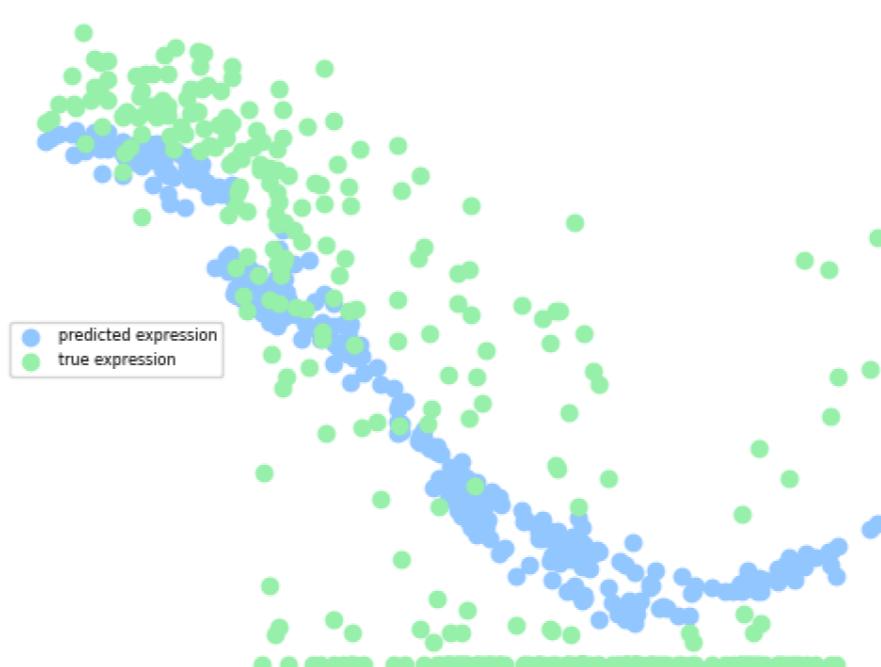
(t-)SNE & UMAP can be interpret as VI algorithms applied in the ProbDR framework, with  $\mathbf{M}$  being a graph Laplacian.

For t-(SNE) & UMAP, variational & model probs are specified as,

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

Experiments show that graph GP generative models on  $\mathbf{A} \rightarrow \mathbf{Y}$  can outperform GPLVMs & VAEs.

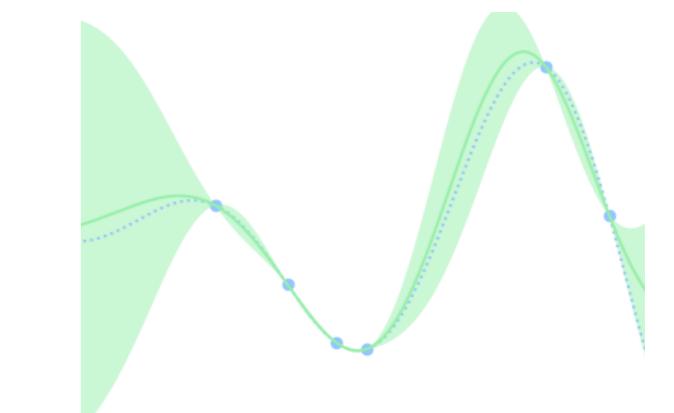
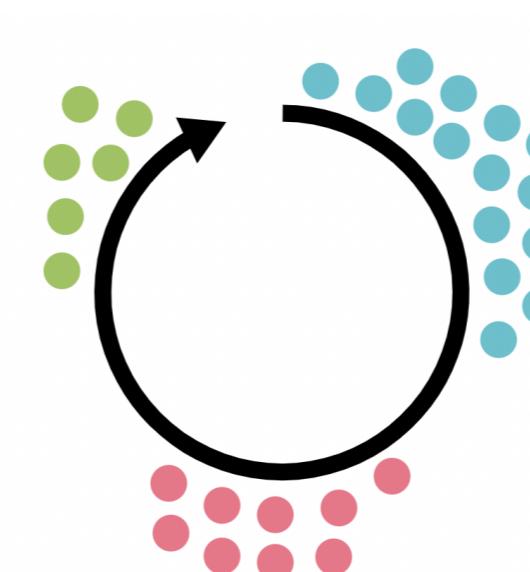
## Applications & Extensions



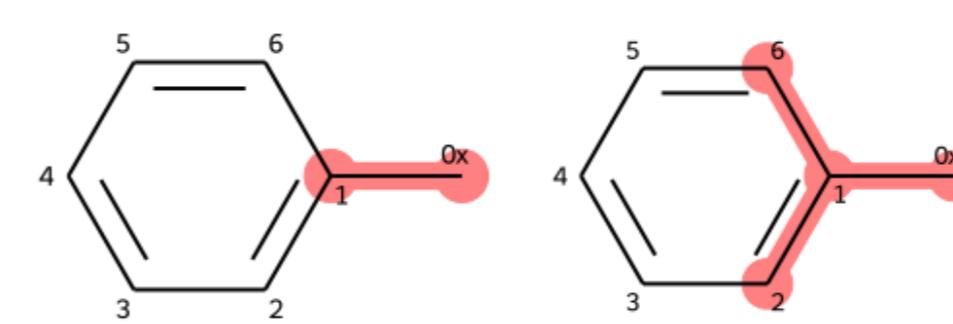
**left:** predicted gene expression given latents using a graph GP generative model, difficult to obtain without our generative model interp. **right:** illustration of convergence of eigenvectors of  $\mathbf{L}$  to eigenfunctions of  $\Delta$  - thus hinting at a convergence of  $\mathbf{X} \rightarrow \mathbf{L} \rightarrow \mathbf{Y}$  to a Matérn GP on a manifold, and thus GPLVMs.

### The Why & Connections to other projects

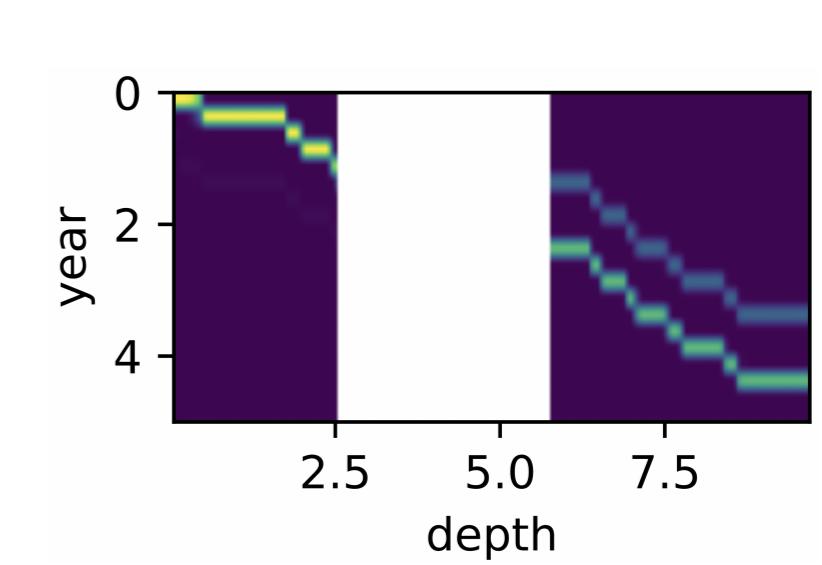
Probabilistic interpretations:



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



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



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