

Dimensionality Reduction: a Probabilistic Perspective

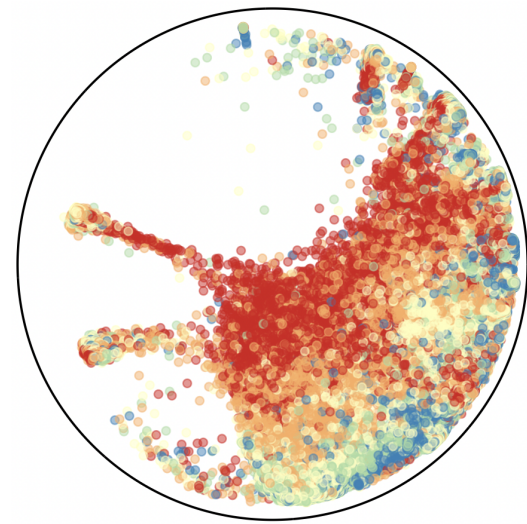
Aditya Ravuri Neil D. Lawrence



Background

Dimensionality Reduction (DR)

High dimensional data can sit on low dim. manifolds, which DR algos. estimate. We show that these correspond to probabilistic inference.



gene expression projected into a Poincaré disk showing cell differentiation

With DR methods, often a need for:

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

Existing Interpretations

PCA, FA, ICA, CMDS algorithms correspond to MLE in

$$\begin{aligned} \mathbf{Y} &= \mathbf{X}\mathbf{W} + \epsilon \text{ with } \mathbf{X}, \mathbf{W} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}) & (1) \\ \Leftrightarrow \mathbf{Y}|\mathbf{X} &\sim \mathcal{MN}(\mathbf{0}, \mathbf{X}\mathbf{X}^T + \sigma^2\mathbf{I}, \mathbf{I}) & (2) \\ \Leftrightarrow \mathbf{Y}|\mathbf{W} &\sim \mathcal{MN}(\mathbf{0}, \mathbf{I}, \mathbf{W}\mathbf{W}^T + \sigma^2\mathbf{I}), & (3) \end{aligned}$$

with different priors on \mathbf{X} (and hence model implications),

- isotropic Gaussian \leftrightarrow maximal variance
- non-isotropic Gaussian \leftrightarrow identifiability
- non-Gaussian \leftrightarrow causation

Eigencomponent based DR

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

$$\mathbf{S}|\cdot \sim \mathcal{W}((\cdot)(\cdot)^T + \sigma^2\mathbf{I}, d), \quad (4)$$

where $\mathbf{S} \equiv \mathbf{Y}\mathbf{Y}^T$ or $\mathbf{Y}^T\mathbf{Y}$ is the **sample** cov. and $(\cdot) = \mathbf{X}$ or \mathbf{W} .

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

MLE	linear in latent	Cov($\mathbf{S} \mathbf{X}$)	Cov($\mathbf{S} \mathbf{W}$)
Major eigs	data	$\mathbf{X}\mathbf{X}^T + \sigma^2\mathbf{I}$	$\mathbf{W}\mathbf{W}^T + \sigma^2\mathbf{I}$
Minor eigs	incidence	$(\mathbf{X}\mathbf{X}^T + \beta\mathbf{I})^{-1}$	$(\mathbf{W}\mathbf{W}^T + \beta\mathbf{I})^{-1}$

The minor eigs. result is due to Minor Component Analysis (**MCA**).

Our interpretation: DR algorithms outputting embeddings as eigen-components of PSD matrices are two step inference processes:

1. estimate a covariance $\hat{\mathbf{S}}$ or graph Laplacian(/precision) $\hat{\mathbf{L}}$,
2. obtain embedding through MLE using model in eq. (4) (MLE is computed using minor eigencomponents if precision is specified, major eigs. if covariance is instead).

List of our interpretations:

- Laplacian Eigenmaps, Spectral Embeddings:

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

- Locally Linear Embedding:

$$\hat{\mathbf{L}}(\mathbf{Y}) = \text{MLE in a GMRF/Matérn-1 graph GP}$$

- Kernel PCA:

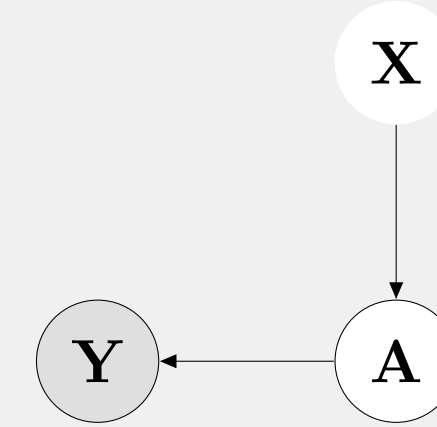
$$\hat{\mathbf{S}}(\mathbf{Y}) = k(\mathbf{Y}_i, \mathbf{Y}_j)$$

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

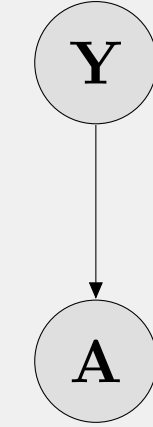
$$\hat{\mathbf{S}}(\mathbf{Y}) = \exp[-t\Delta(\mathbf{Y})]$$

Variational Inference based DR

Generative Model



Variational Approximation



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

(t-)SNE & UMAP can be interpret as variational inference (VI) algorithms in graph Laplacian estimation models with a generative model that contributes a constant (w.r.t. the latents) to the ELBO.

ELBOs here take the form **constant - cost_func_of_dr_algo**:

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

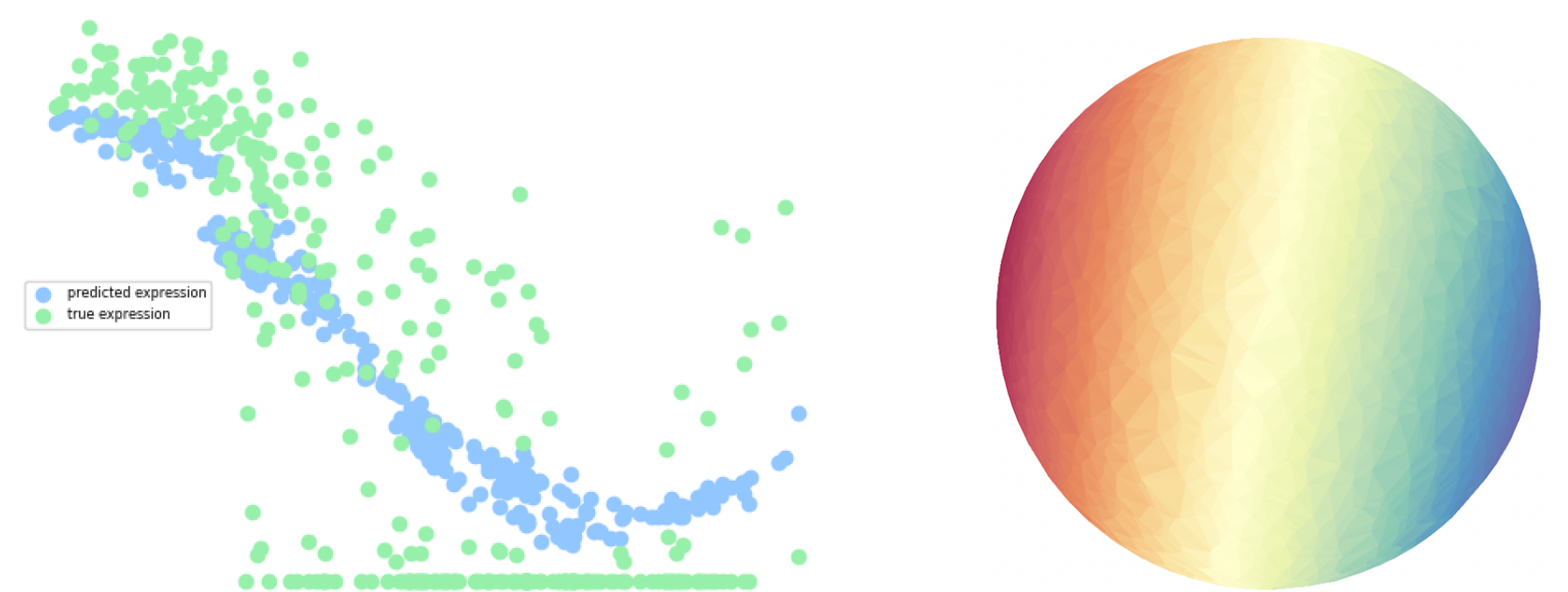
Variational & model probabilities $q(\mathbf{A}|\mathbf{Y}), p(\mathbf{A}|\mathbf{X})$ are specified as,

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

Methods in literature approximating the components above using neural networks are simple extensions of this framework.

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

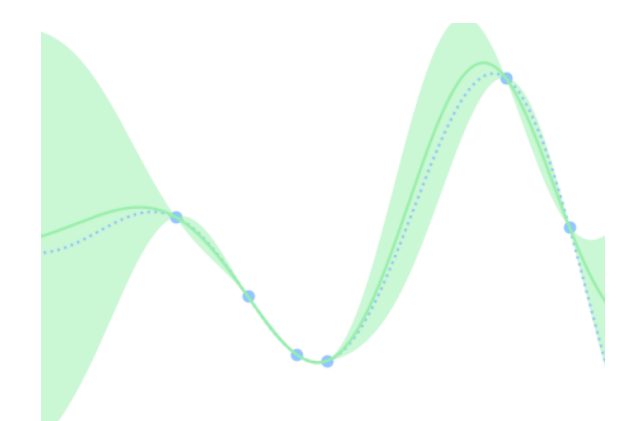
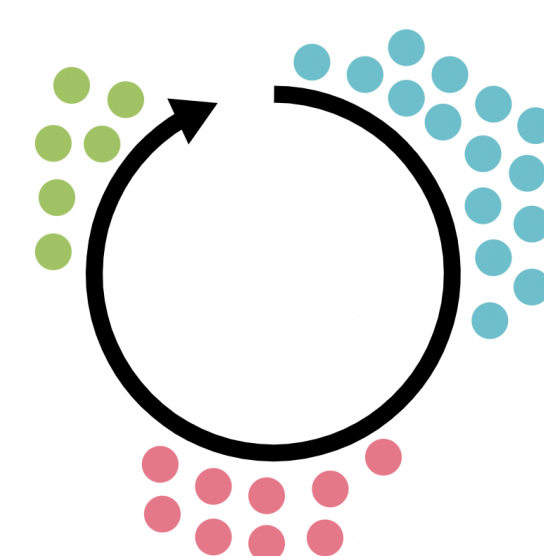
Applications & Extensions



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 Δ - thus hinting at a convergence of $\mathbf{X} \rightarrow \mathbf{L} \rightarrow \mathbf{Y}$ to a Matérn GP on a manifold, and thus GPLVMs.

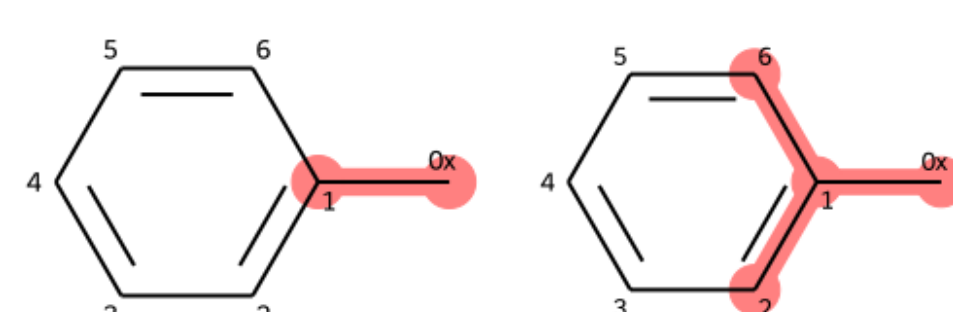
The Why & Connections to other projects

Probabilistic interpretations:

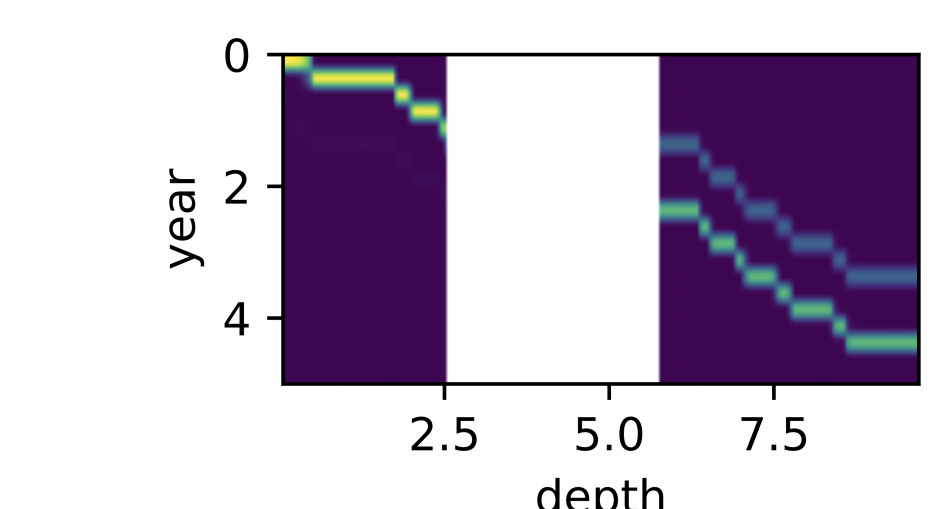


enable efficient reformulations
enable prior specification
e.g. RE models are GPs

aid communication
enable software cross-utilization
e.g. GAMs 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