Multi-view Spectral Clustering and Generation from a Shared Latent Space

Francesco Tonin

with works of L. Houthuys, A. Pandey, Q. Tao, P. Patrinos, J. A.K. Suykens, and others

Department of Electrical Engineering ESAT-STADIUS
KU Leuven. Kasteelpark Arenberg 10 B-3001 Leuven, Belgium
francesco.tonin@esat.kuleuven.be

Bioinformatics and AI Seminar
May 11, 2022

Table of Contents

Unsupervised Multi-view Learning

Unsupervised Multi-view Learning

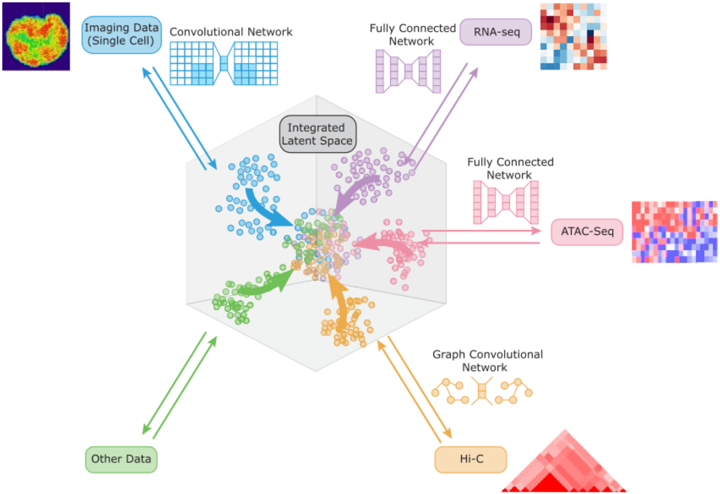
- Real-world data can be collected from different sources, e.g.,
  - Newspaper articles in multiple languages,
  - Multi-omic data: genomics, transcriptomics, methylomics...
- **Aim**: fuse information from multiple views to gain greater insights compared to considering a single view.

Unsupervised Multi-view Learning

- Real-world data can be collected from different sources, e.g.,
  - Newspaper articles in multiple languages,
  - Multi-omic data: genomics, transcriptomics, methylomics...
- **Aim**: fuse information from multiple views to gain greater insights compared to considering a single view.
- **Multi-view dimensionality reduction**: the multi-view dataset is reduced to a lower-dimensional space to compactly represent the heterogeneous data. Applications:
  - **Multi-view generation and domain translation**: generation of new samples in multiple views simultaneously or generation of missing views.
  - **Multi-view clustering**: using several views can reveal structures not seen with a single data source (e.g., cancer subtypes can be defined based on gene expression and DNA methylation together).

Generative Multi-view Kernel PCA

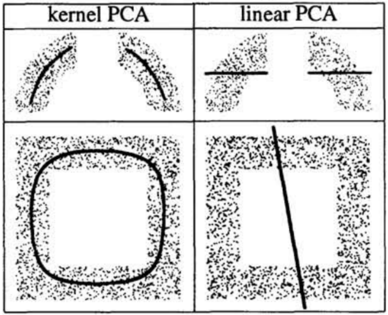
Motivation



(reproduced from [Yang et al., 2021])

Multi-view data integration and domain translation: each view represents a different modality of the same population of cells.

Kernel Principal Component Analysis (KPCA)



(reproduced from [Mika et al., 1999])

- Nonlinear extension of PCA.
- Linear PCA is performed in the feature space induced by the feature map ϕ.

Primal problem [Suykens et al., 2002]:

$$\min_{w,e} \frac{1}{2}\|w\|^2 - \frac{1}{2\lambda}\sum_{i=1}^{N} e_i^2 \quad \text{s.t.} \quad e_i = w^T\phi(x_i)$$

Dualization of Kernel PCA

Obtain upper bound to primal objective using Fenchel–Young inequality $\frac{1}{2\lambda}e^T e + \frac{\lambda}{2}h^T h \geq e^T h$ [Suykens, 2017]; [Tonin, Patrinos, and Suykens, 2021] :

$$
\begin{aligned}
J_p &= \frac{\eta}{2} \operatorname{Tr} W^T W - \frac{1}{2\lambda}\sum_{i=1}^{N} e_i^T e_i \quad \text{s.t.} \quad e_i = W^T \phi(x_i) \\
&\leq -\sum_{i=1}^{N} e_i^T h_i + \frac{\lambda}{2}\sum_{i=1}^{N} h_i^T h_i + \frac{\eta}{2} \operatorname{Tr}\left(W^T W\right) \quad \text{s.t.} \quad e_i = W^T \phi(x_i) \\
&= -\sum_{i=1}^{N} \phi(x_i)^T W h_i \; + \; \boxed{\frac{\lambda}{2}\sum_{i=1}^{N} h_i^T h_i + \frac{\eta}{2} \operatorname{Tr}\left(W^T W\right)} \triangleq J
\end{aligned}
$$

Dualization of Kernel PCA

Obtain upper bound to primal objective using Fenchel–Young inequality $\frac{1}{2\lambda} e^T e + \frac{\lambda}{2} h^T h \geq e^T h$ [Suykens, 2017]; [Tonin, Patrinos, and Suykens, 2021] :

$$
\begin{aligned}
J_p &= \frac{\eta}{2} \operatorname{Tr} W^T W - \frac{1}{2\lambda} \sum_{i=1}^{N} e_i^T e_i \quad \text{s.t.} \quad e_i = W^T \phi(x_i) \\
&\leq - \sum_{i=1}^{N} e_i^T h_i + \frac{\lambda}{2} \sum_{i=1}^{N} h_i^T h_i + \frac{\eta}{2} \operatorname{Tr}\left(W^T W\right) \quad \text{s.t.} \quad e_i = W^T \phi(x_i) \\
&= - \boxed{\sum_{i=1}^{N} \phi(x_i)^T W h_i} + \frac{\lambda}{2} \sum_{i=1}^{N} h_i^T h_i + \frac{\eta}{2} \operatorname{Tr}\left(W^T W\right) \triangleq J
\end{aligned}
$$

RKM Dual Problem of Kernel PCA

Stationary points of J:

$$\begin{cases} \frac{\partial J}{\partial h_i} = 0 \implies W^T \phi(x_i) = \lambda h_i, \ \forall i = 1, \ldots, N, \\ \frac{\partial J}{\partial W} = 0 \implies W = \frac{1}{\eta} \sum_{i=1}^{N} \phi(x_i) h_i^T. \end{cases}$$

Substituting the expression for W in the first equation gives the eigenvalue problem:

$$\frac{1}{\eta} K H^T = H^T \Lambda,$$

where

- $H = [h_1, \ldots, h_N] \in \mathbb{R}^{s \times N}$,
  - $s \leq N$ is the number of selected principal components
- $\Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_s) \in \mathbb{R}^{s \times s}$,
- $K \in \mathbb{R}^{N \times N}$ is the kernel matrix: $K_{ij} = \phi(x_i)^T \phi(x_j)$.

Multi-view Kernel PCA

Consider two views and two corresponding feature maps $\phi_1 : \Omega_x \to \mathcal{H}_x$ and $\phi_2 : \Omega_y \to \mathcal{H}_y$. The multi-view KPCA objective is [Pandey, Schreurs, and Suykens, 2021]:

$$J_{\text{MV-KPCA}} = \sum_{i=1}^{N} - \phi_1(x_i)^T U h_i - \phi_2(y_i)^T V h_i + \frac{\lambda}{2} h_i^T h_i + \frac{\eta_1}{2} \operatorname{Tr}\left(U^T U\right) + \frac{\eta_2}{2} \operatorname{Tr}\left(V^T V\right),$$

first view: $\phi_1(x_i)^T U h_i$; second view: $\phi_2(y_i)^T V h_i$

where

- U, V are the unknown interconnection matrices,
- h_i is the latent variable of a common subspace $\mathcal{H} \subseteq \mathcal{H}_x \oplus \mathcal{H}_y$.

Training corresponds to the following eigenvalue problem:

$$\left(\frac{1}{\eta_1} K_1 + \frac{1}{\eta_2} K_2\right) H^T = H^T \Lambda,$$

where K_1, K_2 are the kernel matrices of the first and second view.

Generative Multi-view Kernel PCA

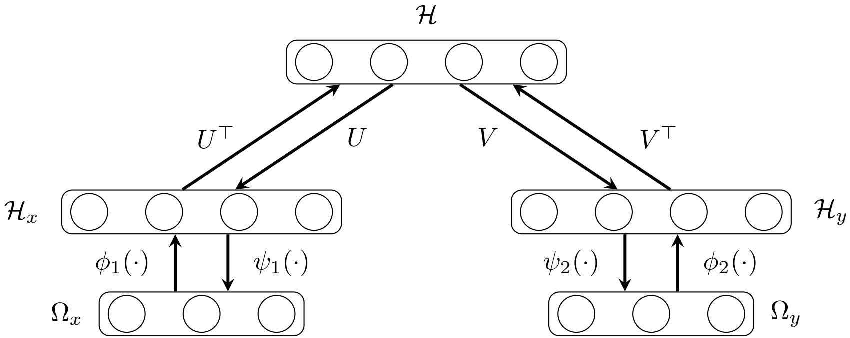
☑ **Task**: given a latent variable $h^\star$, generate data point $x^\star \in \Omega_x, y^\star \in \Omega_y$ in both views.

? **Problem**: compute the inverse images of the feature maps ϕ_1, ϕ_2 (*pre-image problem*).

💡 **Solution**: use parametrized feature maps and learn both the feature maps and the pre-image maps $\psi_1 : \mathcal{H}_x \to \Omega_x$, $\psi_2 : \mathcal{H}_y \to \Omega_y$. The parametrization depends on the data type of each view (e.g., CNN for images, GNN for graphs, LSTM for time series, etc.)

$$\min_{\substack{U,V,h_i,\\ \theta_1,\theta_2,\zeta_1,\zeta_2}} J_{\text{Gen-RKM}} = J^{\text{stab}}_{\text{MV-KPCA}} + \frac{\gamma}{2N}\left(\sum_{i=1}^{N} \left\| x_i - \psi_{1\zeta_1}(\phi_{1\theta_1}(x_i)) \right\|^2 + \sum_{i=1}^{N} \left\| y_i - \psi_{2\zeta_2}(\phi_{2\theta_2}(y_i)) \right\|^2 \right)$$

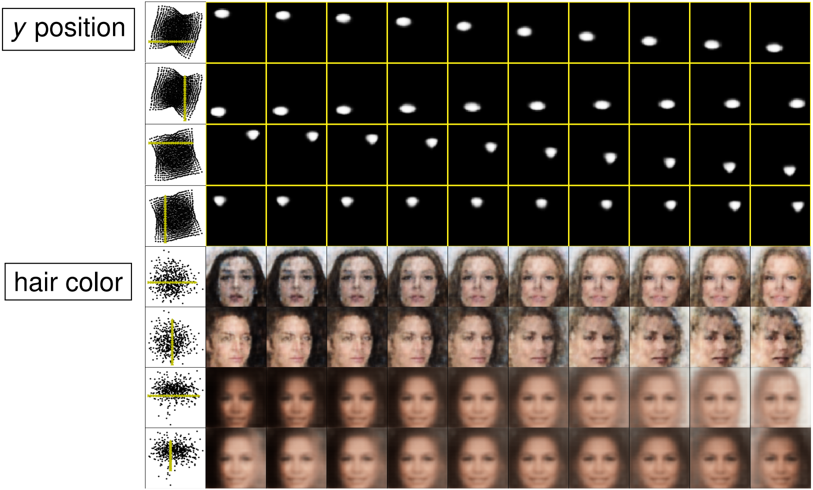
Generative Multi-view Kernel PCA



(reproduced from [Pandey, Schreurs, and Suykens, 2021])

Gen-RKM schematic. The single subspace $\mathcal{H} \subseteq \mathcal{H}_x \bigoplus \mathcal{H}_y$ is shared between the two views Ω_x, Ω_y. The ϕ_1, ϕ_2 are the view-specific feature maps, ψ_1, ψ_2 are the pre-image maps. The inter-connection matrices U, V capture the view-specific dependencies between the shared latent variables and the mapped data sources.

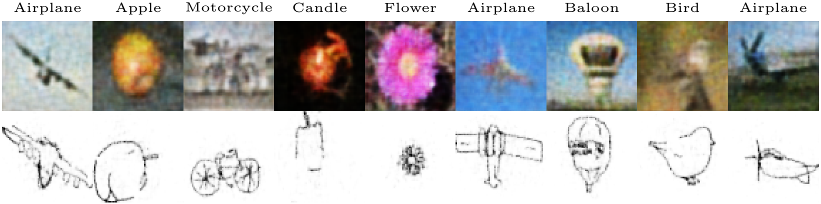
Generative Multi-View Kernel PCA: Disentanglement



Each row shows the generated images corresponding to the **traversal** in the latent space plotted in the first column. [Pandey, Schreurs, and Suykens, 2021]

Interpretability: changing one latent variable affects only one generative factor. Application: separate biological factors.

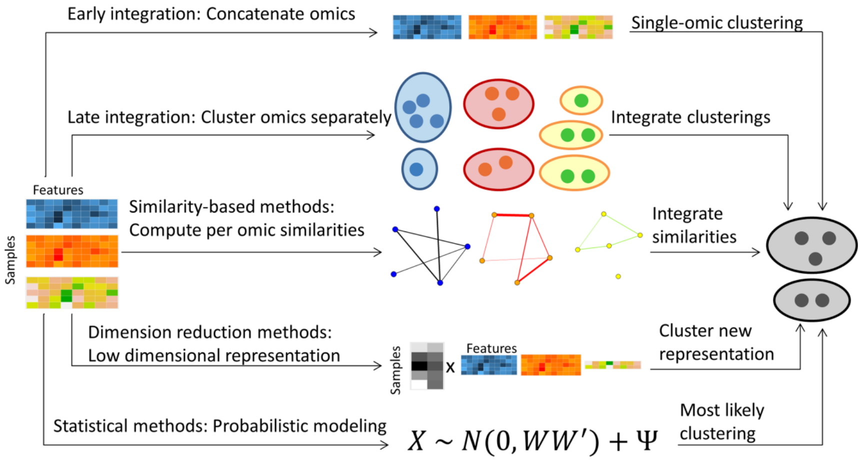
Generative Multi-View Kernel PCA: Multi-view Generation



Multi-view generation on the Sketchy dataset showing labels, images, and sketches generated together from the single latent space $\mathcal{H}$ **shared** among all views. [Pandey, Schreurs, and Suykens, 2021]

Multi-View Kernel Spectral Clustering

Multi-omic clustering approaches



(reproduced from [Rappoport and Shamir, 2018])

Overview of multi-omic clustering algorithms to reveal further insights into biomedical omics dataset.

Kernel Spectral Clustering (KSC)

Weighted KPCA with k clusters [Alzate and Suykens, 2008]:

$$\min_{w^{(l)},b^{(l)},e^{(l)}} \quad \frac{1}{2}\sum_{l=1}^{k-1} {w^{(l)}}^T w^{(l)} - \frac{1}{2N}\sum_{l=1}^{k-1} \gamma^{(l)} {e^{(l)}}^T D^{-1} e^{(l)}$$

$$\text{s.t.} \quad e^{(l)} = \Phi w^{(l)} + b^{(l)} \vec{1}_N,\ l = 1, \ldots, k-1,$$

where

- $\Phi = [\phi(x_1)^T; \ldots; \phi(x_N)^T]$ is the feature matrix,
- $e^{(l)} \in \mathbb{R}^N$ is the l-th clustering score with clustering indicator $\mathrm{sign}(e^{(l)})$,
- D is the degree matrix, motivated by the random walks model, inducing the clustering: $D_{ii} = \sum_j \phi(x_i)^T \phi(x_j)$.

Multi-view Kernel Spectral Clustering

Extend KSC to V views with pairwise coupling terms [Houthuys, Langone, and Suykens, 2018]:

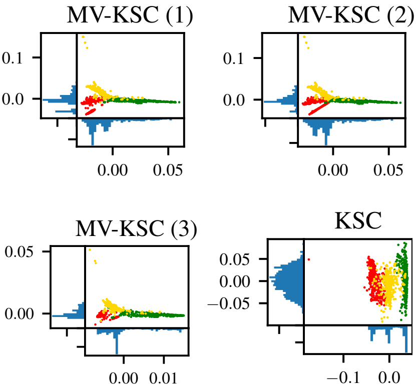
$$
\begin{aligned}
\min_{w^{[v]^{(l)}}, e^{[v]^{(l)}}} \quad & \frac{1}{2} \sum_{v=1}^{V} \sum_{l=1}^{k-1} {w^{[v]^{(l)}}}^T w^{[v]^{(l)}} \\
& - \frac{1}{2N} \sum_{v=1}^{V} \sum_{l=1}^{k-1} \gamma^{[v]^{(l)}} {e^{[v]^{(l)}}}^T {D^{[v]}}^{-1} e^{[v]^{(l)}} \\
& - \frac{1}{2} \sum_{v,u=1, v \neq u}^{V} \sum_{l=1}^{k-1} \rho^{(l)} {e^{[v]^{(l)}}}^T S^{[v,u]} e^{[u]^{(l)}} \\
\text{s.t.} \quad & e^{[v]^{(l)}} = \left(\Phi^{[v]} - \vec{1}_N {\hat{\mu}^{[v]}}^T \right) w^{[v]^{(l)}},
\end{aligned}
\tag{1}
$$

where

- $S^{[v,u]} = {D^{[v]}}^{-\frac{1}{2}} {D^{[u]}}^{-\frac{1}{2}}$ couples view v and u.

The **coupling term** describes the correlation between the clustering variables of two different views. Problem (1) is optimized with Lagrangian duality $\rightarrow$ solve a $VN \times VN$ eigenvalue problem.

Multi-view KSC - Latent Space



Dataset with $V = 3$ views with image and text data

- MV-KSC achieves better separation than KSC on concatenated views
- MV-KSC has one latent variable for each view $\rightarrow$ less interpretability with many views

MV-KSC with Shared Latent Space: MV-KSC-RKM

Obtain upper bound to weighted KPCA using the weighted Fenchel–Young inequality $\frac{1}{2\lambda} e^T D^{-1} e + \frac{\lambda}{2} h^T D h \geq e^T h$:

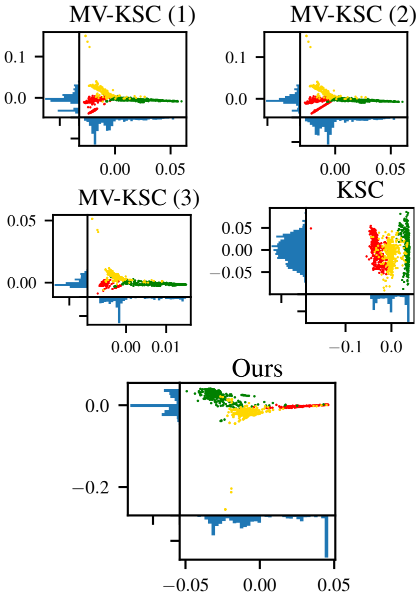
$$J_{\text{KSC}} = \frac{1}{2}\sum_{l=1}^{k-1} w^{(l)}{}^T w^{(l)} - \frac{1}{2N}\sum_{l=1}^{k-1} \gamma^{(l)} e^{(l)}{}^T D^{-1} e^{(l)} \quad \text{s.t.} \quad e^{(l)} = \Phi_c w^{(l)}$$

$$\leq -\sum_{l=1}^{k-1} \left(\Phi_c w^{(l)}\right)^T h^{(l)} + \sum_{l=1}^{k-1} \frac{\lambda^{(l)}}{2} h^{(l)}{}^T D h^{(l)} + \frac{\eta}{2}\sum_{l=1}^{k-1} w^{(l)}{}^T w^{(l)}$$

In the multi-view case, we incorporate the KSC view-specific objectives with different feature map for each view and couple the views by imposing that the latent variable *h* is common to all views. Consequences:

- Single latent space $\rightarrow$ improved **interpretability**
- Legendre–Fenchel transformation leads to a $N \times N$ eigenvalue problem $\rightarrow$ better **efficiency**, independent of the number of views and of the number of features

MV-KSC-RKM - Shared Latent Space



Dataset with $V = 3$ views with image and text data

- The clusters in the common latent space are better separated
- Improved data discovery thanks to a single latent space
- Outputs not only 2D plot, but also clustering labels (unlike t-SNE, UMAP)

Future Extensions

Future challenges

- Encode **prior knowledge** in the learned disentangled latent variables
  - Force the model to learn specific features such as the texture of the tissue
- Explicit feature maps for multi-view clustering, e.g., **Graph Neural Networks**
  - Spatial transcriptomics
- **Unpaired** multi-view learning
  - Different cells have different modalities

References

References I

Alzate, C. and J. Suykens (2008). “Multiway Spectral Clustering with Out-of-Sample Extensions through Weighted Kernel PCA”. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 32.2, pp. 335–347.

Houthuys, L., R. Langone, and J. Suykens (2018). “Multi-View Kernel Spectral Clustering”. *Information Fusion* 44, pp. 46–56.

Mika, S., B. Schölkopf, A. Smola, K.-R. Müller, M. Scholz, and G. Rätsch (1999). “Kernel PCA and De-Noising in Feature Spaces”. In: *Proceedings of the 1998 Conference on Advances in Neural Information Processing Systems*. MIT Press, pp. 536–542.

Pandey, A., J. Schreurs, and J. Suykens (2021). “Generative Restricted Kernel Machines: A Framework for Multi-View Generation and Disentangled Feature Learning”. *Neural Networks* 135, pp. 177–191.

References II

Rappoport, N. and R. Shamir (2018). "Multi-Omic and Multi-View Clustering Algorithms: Review and Cancer Benchmark". *Nucleic Acids Research* 46.20, pp. 10546–10562.

Suykens, J. (2017). "Deep Restricted Kernel Machines Using Conjugate Feature Duality". *Neural Computation* 29.8, pp. 2123–2163.

Suykens, J., T. Van Gestel, J. De Brabanter, B. De Moor, and J. Vandewalle (2002). *Least Squares Support Vector Machines*. World Scientific.

Tonin, F., P. Patrinos, and J. Suykens (2021). "Unsupervised Learning of Disentangled Representations in Deep Restricted Kernel Machines with Orthogonality Constraints". *Neural Networks* 142, pp. 661–679.

References III

Yang, K. D., A. Belyaeva, S. Venkatachalapathy, K. Damodaran, A. Katcoff, A. Radhakrishnan, G. V. Shivashankar, and C. Uhler (2021). "Multi-Domain Translation between Single-Cell Imaging and Sequencing Data Using Autoencoders". *Nature Communications* 12.1, p. 31.