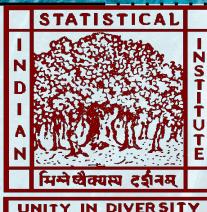


19 November, 2021

Ph.D. VIVA-VOCE SEMINAR



INTEGRATIVE CLUSTERING OF MULTI-VIEW DATA: Subspace Clustering, Graph Approximation to Manifold Learning

APARAJITA KHAN

SENIOR RESEARCH FELLOW

Supervisor: PROF. PRADIPTA MAJI

MACHINE INTELLIGENCE UNIT
INDIAN STATISTICAL INSTITUTE

MULTI-VIEW CLUSTERING

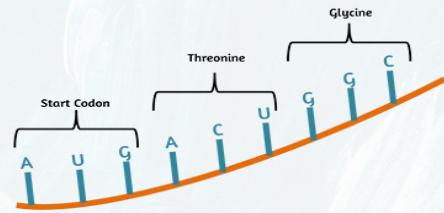
M Data sources



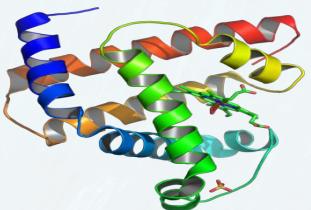
Epigenomic



Genomic



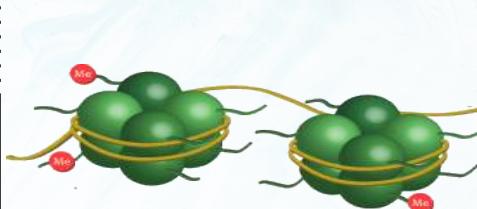
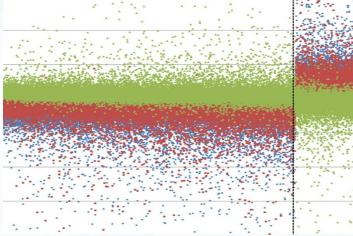
Transcriptomic



Proteomic

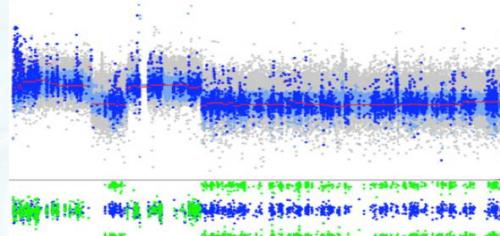
MULTI-VIEW CLUSTERING

DNA
Methylation



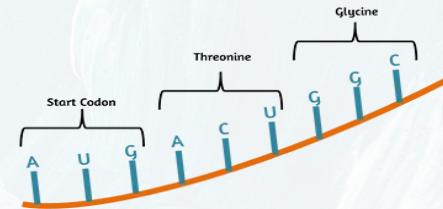
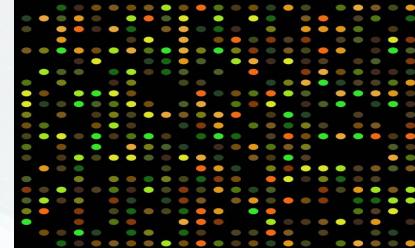
Epigenomic

Copy Number
Variation



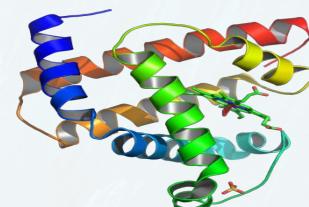
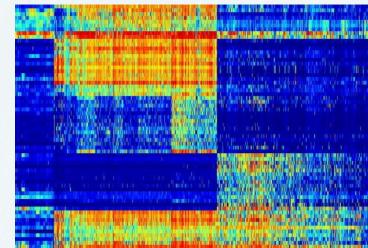
Genomic

Gene
Expression



Transcriptomic

Protein
Expression

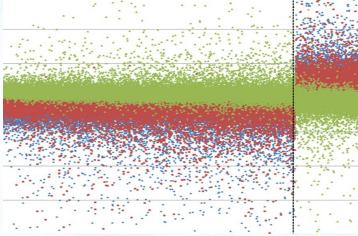


Proteomic

n
Samples/
Patients

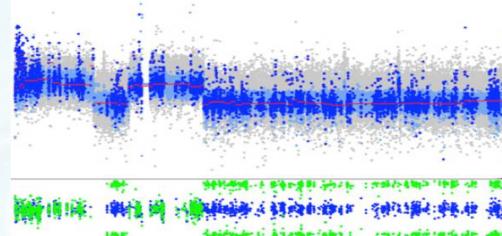
MULTI-VIEW CLUSTERING

DNA
Methylation



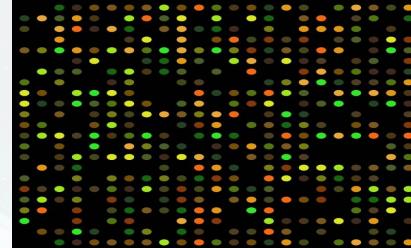
$$X_1 \in \mathbb{R}^{n \times d_1}$$

Copy Number
Variation



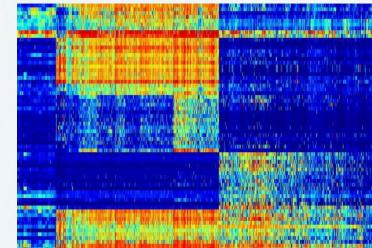
$$X_2 \in \mathbb{R}^{n \times d_2}$$

Gene
Expression



...

Protein
Expression



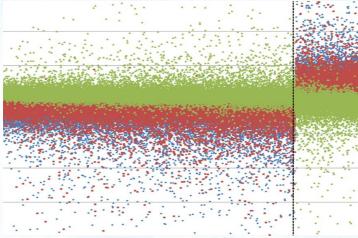
$$X_M \in \mathbb{R}^{n \times d_M}$$

n
Samples/
Patients

Modalities / Views

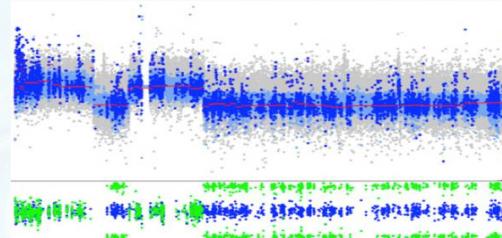
MULTI-VIEW CLUSTERING

DNA
Methylation



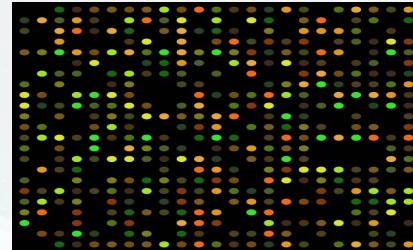
$$X_1 \in \mathbb{R}^{n \times d_1}$$

Copy Number
Variation

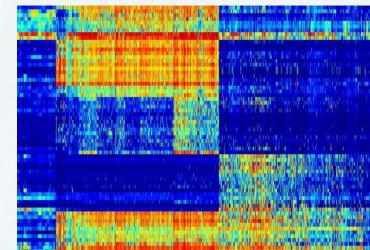


$$X_2 \in \mathbb{R}^{n \times d_2}$$

Gene
Expression



Protein
Expression



n
Samples/
Patients

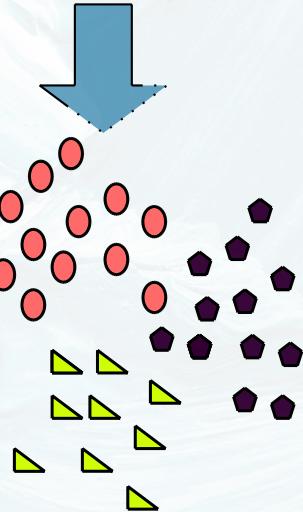
$$X_M \in \mathbb{R}^{n \times d_M}$$

...

Clustering n samples

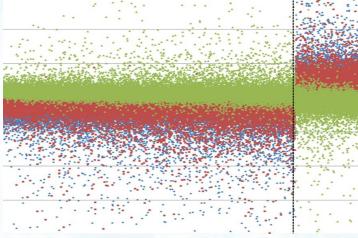
- Comprehensive View of the System
- Complementary Information

- Resilience to Noise
- Cross-Platform Analysis



MULTI-VIEW CLUSTERING

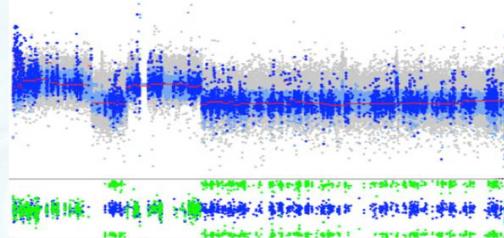
DNA
Methylation



$$X_1 \in \mathbb{R}^{n \times d_1}$$

β -values $\in [0, 1]$

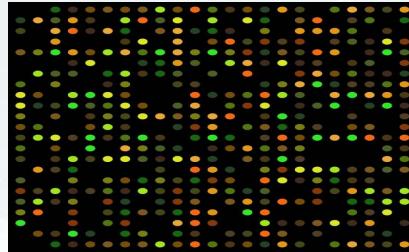
Copy Number
Variation



$$X_2 \in \mathbb{R}^{n \times d_2}$$

segment mean in \log_2

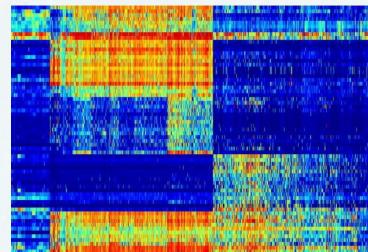
Gene
Expression



...

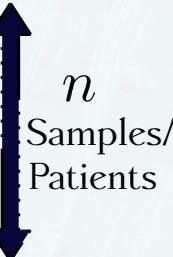
Reads per million $\sim 10^6$

Protein
Expression



$$X_M \in \mathbb{R}^{n \times d_M}$$

fold change in \log_2

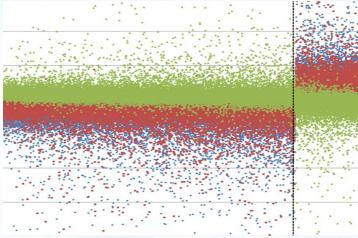


CHALLENGES

- DATA HETEROGENITY IN TERMS OF UNIT, VARIANCE, AND SCALE

MULTI-VIEW CLUSTERING

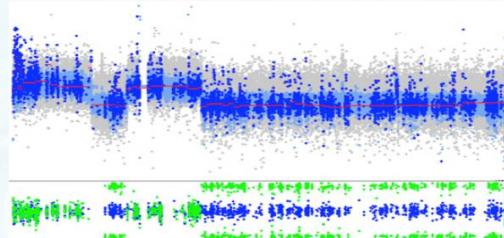
DNA
Methylation



$$X_1 \in \mathbb{R}^{n \times d_1}$$

~450K CpG sites

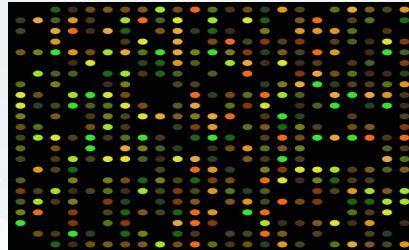
Copy Number
Variation



$$X_2 \in \mathbb{R}^{n \times d_2}$$

~2K CNV segments

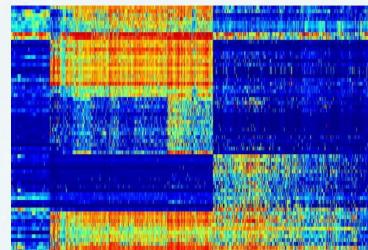
Gene
Expression



...

~20K Genes

Protein
Expression



$$X_M \in \mathbb{R}^{n \times d_M}$$

~300 Proteins

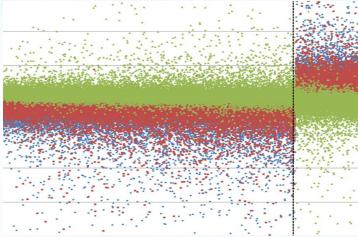
n
Samples/
Patients

CHALLENGES

- DATA HETEROGENITY IN TERMS OF UNIT, VARIANCE, AND SCALE
- HIGH DIMENSION LOW SAMPLE SIZE NATURE OF VIEWS

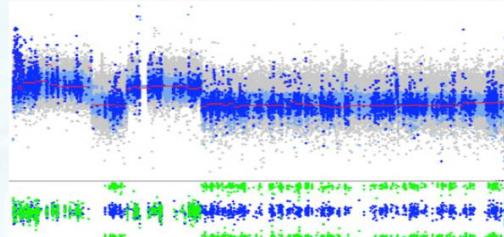
MULTI-VIEW CLUSTERING

DNA
Methylation



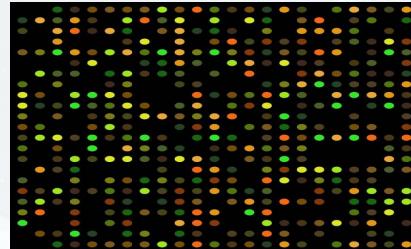
$$X_1 \in \mathbb{R}^{n \times d_1}$$

Copy Number
Variation



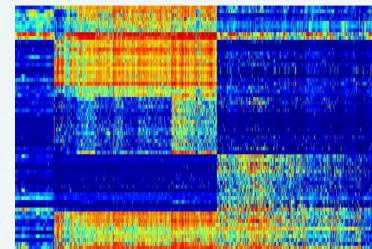
$$X_2 \in \mathbb{R}^{n \times d_2}$$

Gene
Expression



...

Protein
Expression



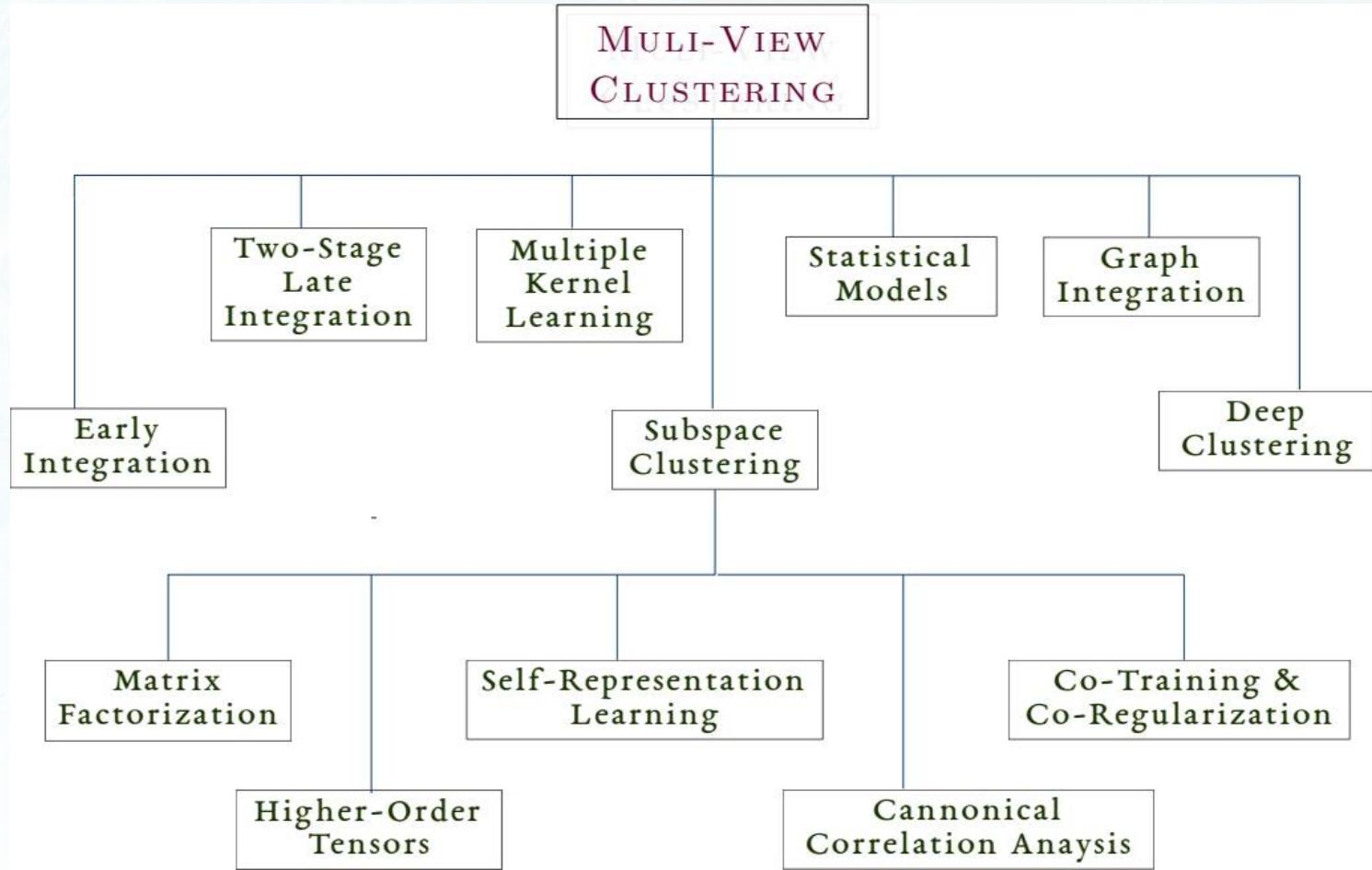
$$X_M \in \mathbb{R}^{n \times d_M}$$

\uparrow
 n
Samples/
Patients
 \downarrow

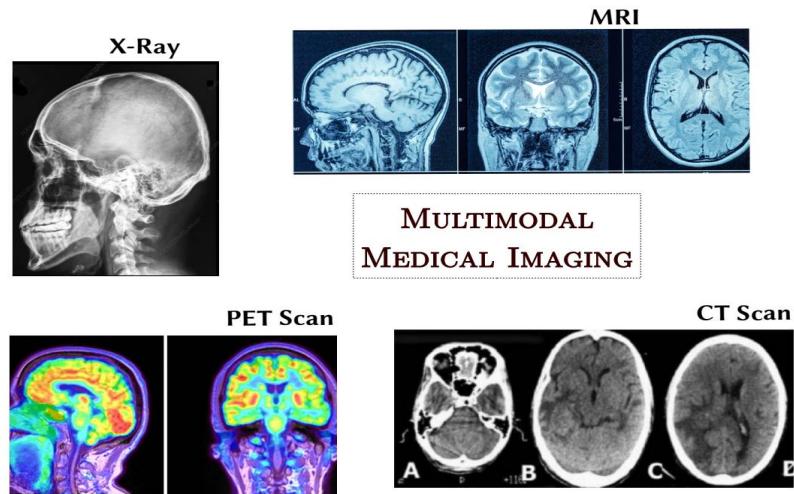
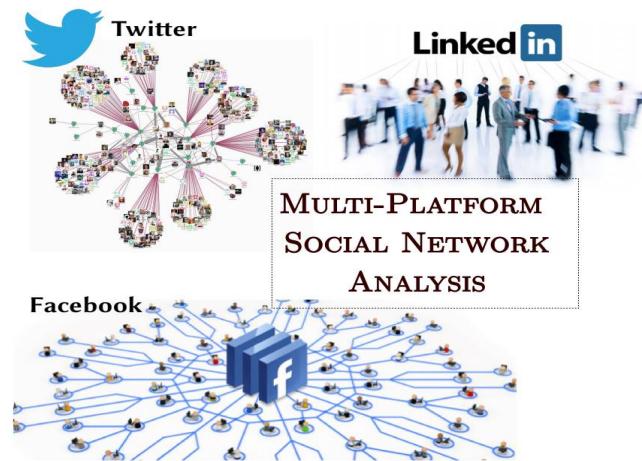
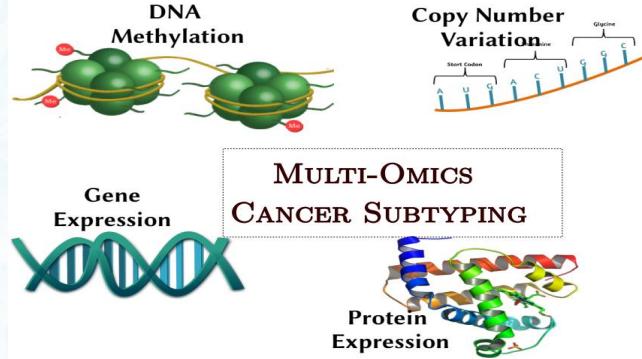
CHALLENGES

- DATA HETEROGENITY IN TERMS OF UNIT, VARIANCE, AND SCALE
- HIGH DIMENSION LOW SAMPLE SIZE NATURE OF VIEWS
- NOISY VIEWS, ALL VIEWS ARE NOT EQUIALLY INFORMATIVE
- LOW-RANK NON-LINEAR GEOMETRY OF VIEWS

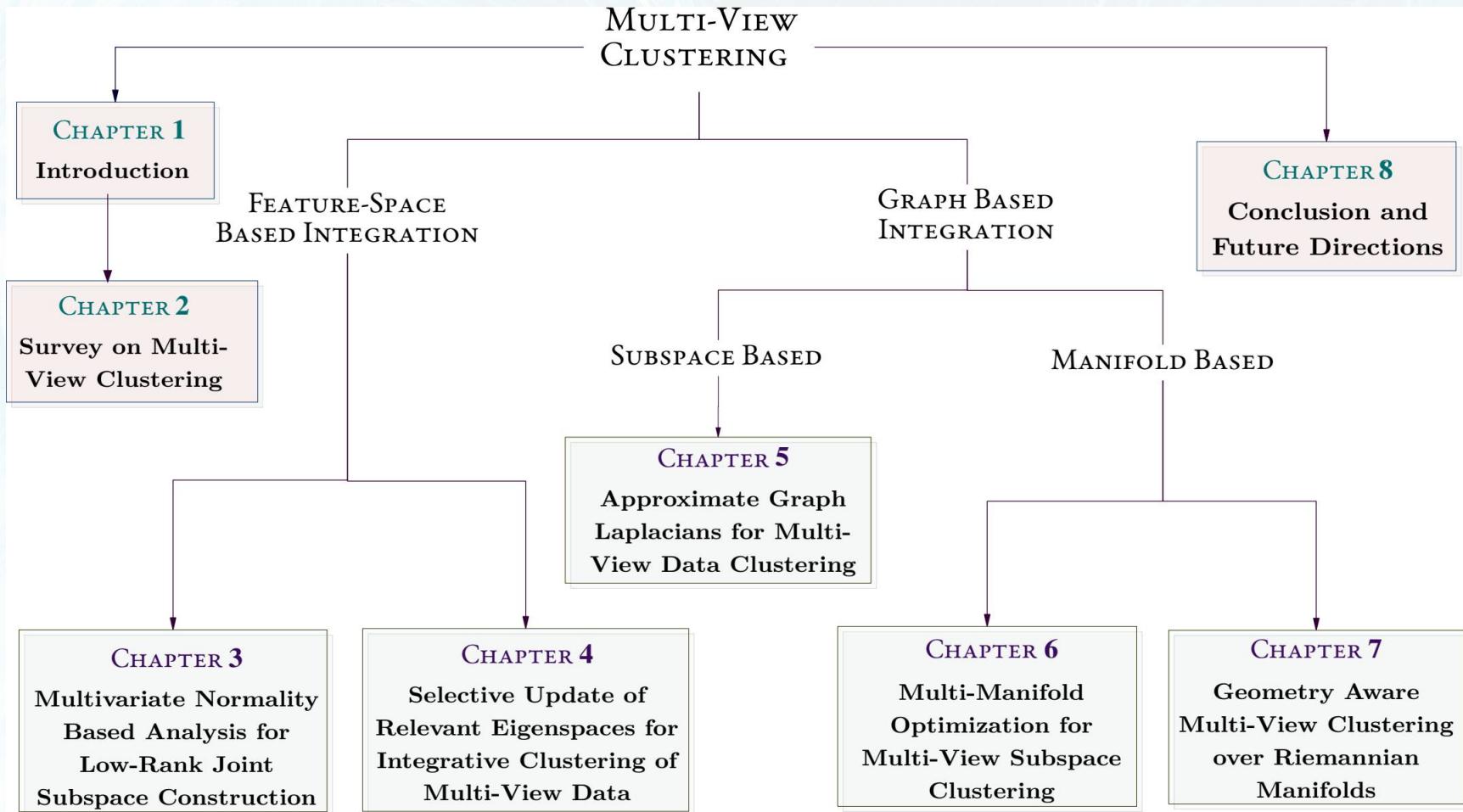
MULTI-VIEW CLUSTERING APPROACHES



MULTI-VIEW CLUSTERING: Application Areas



OUTLINE OF THE THESIS



CHAPTER 3

Multivariate Normality Based Analysis for Low-Rank Joint Subspace Construction

A. Khan and P. Maji, “Low-Rank Joint Subspace Construction for Cancer Subtype Discovery,” **IEEE/ACM Transactions on Computational Biology and Bioinformatics (TCBB)**, vol. 17, no. 4, pp. 1290–1302, 2020. DOI: 10.1109/TCBB.2019.2894635.

Signal-plus-Noise Model of Data

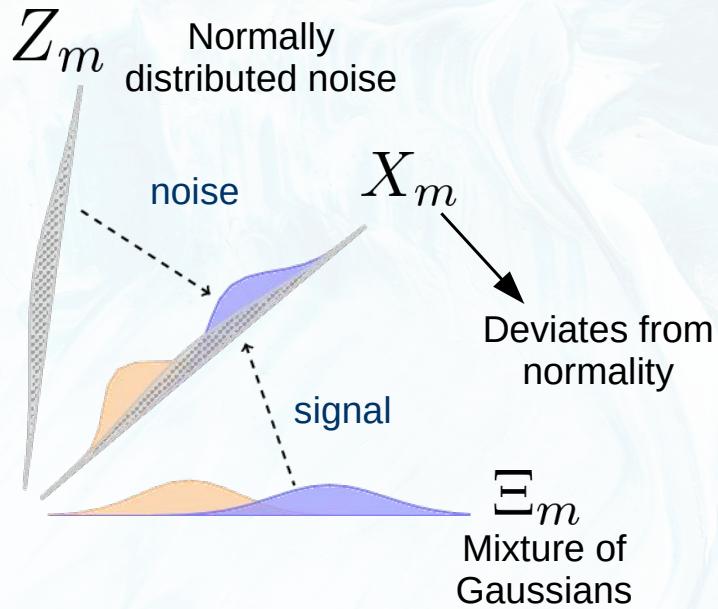
Feature-space representation of M views

$$X_1 \in \mathbb{R}^{n \times d_1}$$

$$X_2 \in \mathbb{R}^{n \times d_2}$$

...

$$X_M \in \mathbb{R}^{n \times d_M}$$



Signal-plus-Noise model of $X_m \in \mathbb{R}^{n \times d_m}$

$$X_m = \Xi_m + Z_m$$

Principal Subspace of a View

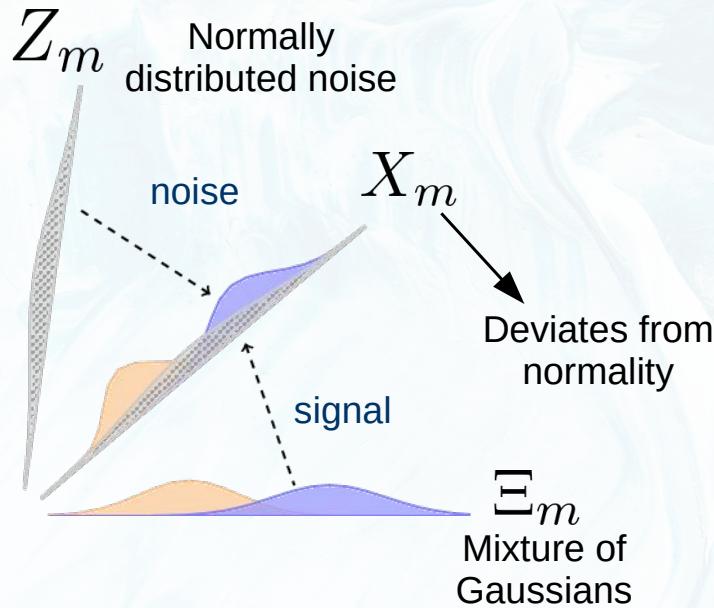
Feature-space representation of M views

$$X_1 \in \mathbb{R}^{n \times d_1}$$

$$X_2 \in \mathbb{R}^{n \times d_2}$$

...

$$X_M \in \mathbb{R}^{n \times d_M}$$



Signal-plus-Noise model of $X_m \in \mathbb{R}^{n \times d_m}$

$$X_m = \Xi_m + Z_m$$

r_m - Rank of latent signal component

Singular value decomposition

$$X_m - \mathbf{1}\mu(X_m)^T = U(X_m)\Sigma(X_m)V(X_m)^T$$

r_m -dimensional principal subspace of X_m

$$\Psi(X_m) = \langle U(X_m), \Sigma(X_m) \rangle$$

Summarizes meaningful variation of X_m

Rank Estimation

$$Z_m \sim \mathcal{N}(0, \Lambda)$$

r_m -dimensional principal subspace should be normally distributed if no cluster structure

Better cluster structure  Higher deviance from Normality

Hypothesis Test:

\mathcal{H}_0 : r -dimensional subspace does not deviate more from normality compared to the $(r - 1)$ -dimensional one.

\mathcal{H}_1 : r -dimensional subspace deviates more from normality compared to the $(r - 1)$ -dimensional one.

Test Statistic: Royston's H statistic for multivariate normality

$$H_r(U(X_m)) = \frac{e}{r} \sum_{i=1}^r R^i$$

 r -dimensional left singular subspace of X_m

$$\gamma_r = H_r - H_{r-1}$$

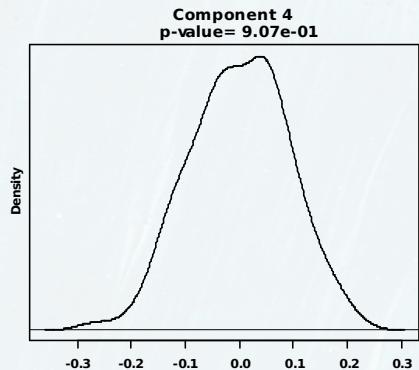
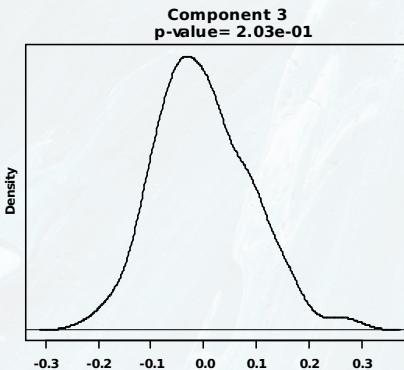
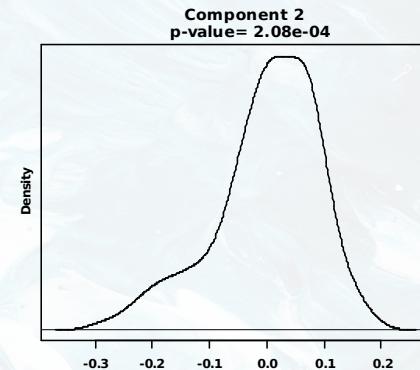
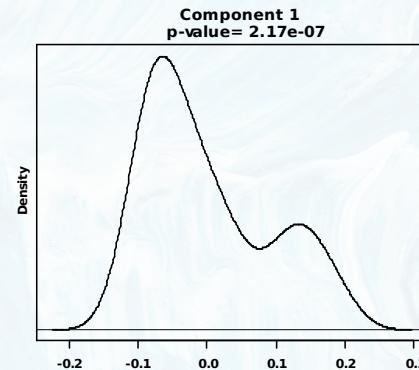
reduces to

Does the r -th singular vector reflect cluster structure and deviate from normality?

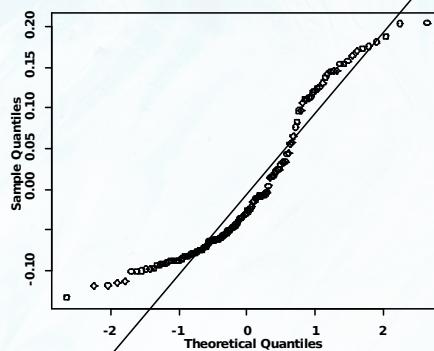
Rank Estimation

Signal components deviate from normality

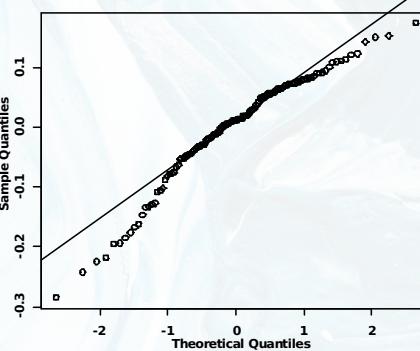
Normally distributed noise components



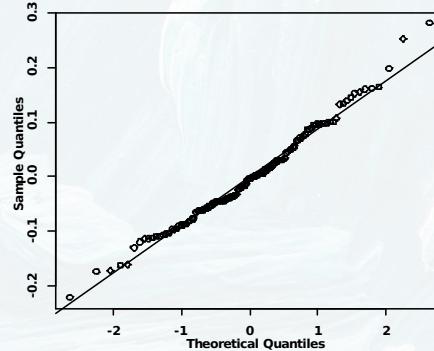
Normal Q-Q Plot



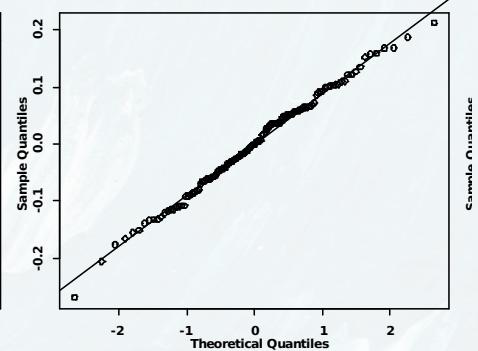
Normal Q-Q Plot



Normal Q-Q Plot



Normal Q-Q Plot

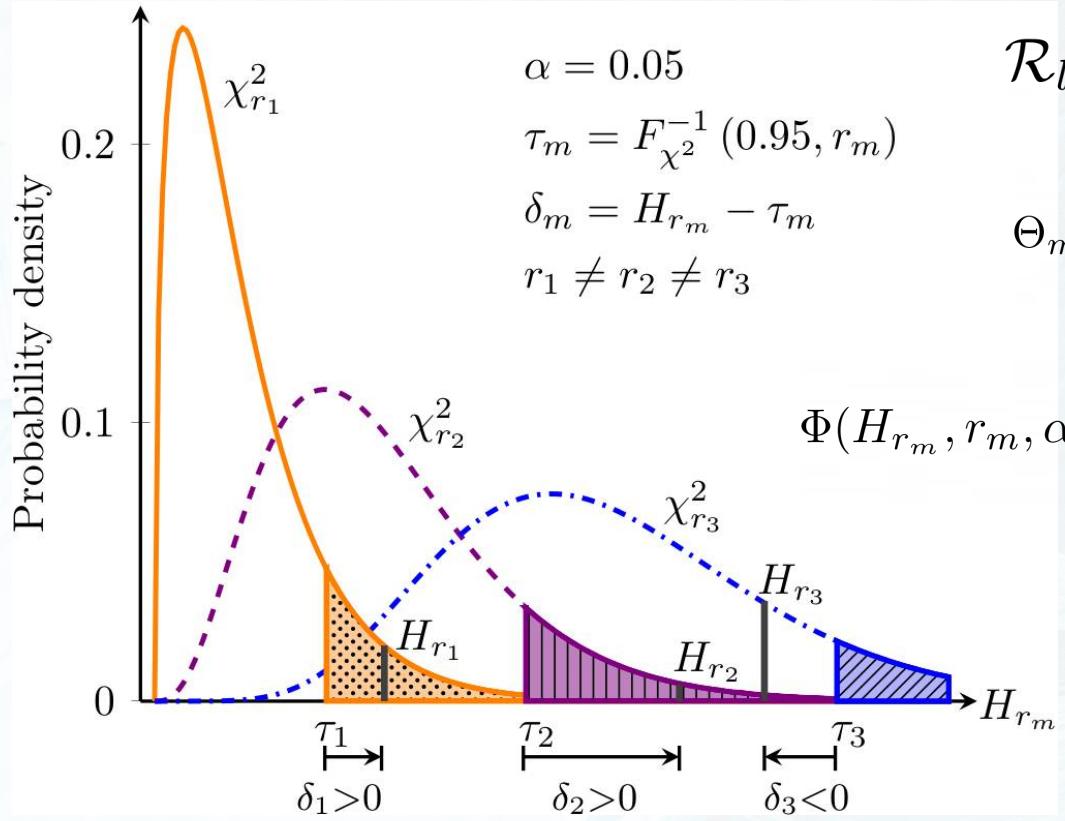


Density and Q-Q plots of top 4 singular vectors of RNA expression view of CESC data set

View Relevance Estimation

$$H_{r_m} \sim \chi^2_{r_m}$$

Chi-squared distribution with r_m degrees of freedom



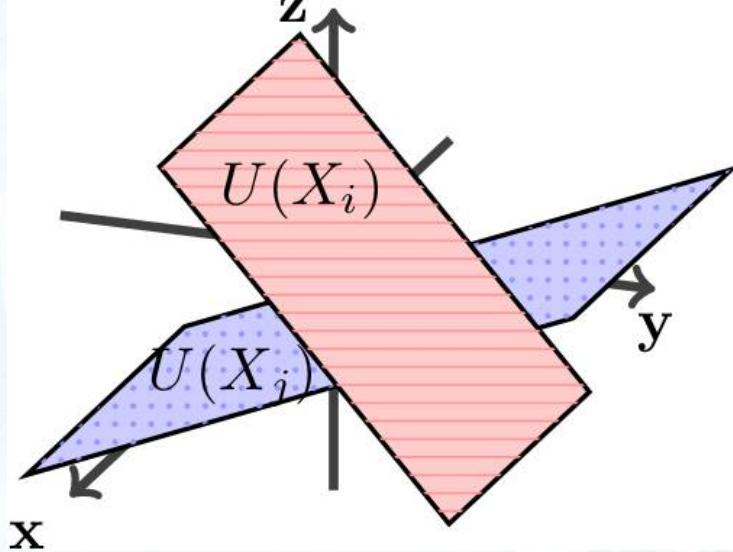
$$\mathcal{R}_l(X_m) = \Phi(H_{r_m}, r_m, \alpha) \times \Theta_m$$

$$\Theta_m = \text{tr}(\Sigma(X_m)) \Bigg/ \sum_{j=1}^M \text{tr}(\Sigma(X_j))$$

$$\Phi(H_{r_m}, r_m, \alpha) = \frac{1}{2} \left[1 + \frac{H_{r_m} - F_{\chi^2}^{-1}((1-\alpha), r_m)}{\max\{H_{r_m}, F_{\chi^2}^{-1}((1-\alpha), r_m)\}} \right]$$

Dependence Between Views

$$0 < \mathcal{D}(X_j|X_i) < 1$$

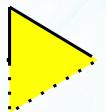


Projection

$$\mathcal{P} = U(X_i)U(X_i)^T U(X_j)$$

Dependence

$$\mathcal{D}(X_j|X_i) = \frac{\|\mathcal{P}\|_F^2}{\|U(X_j)\|_F^2}$$



Integrate most relevance view with high shared information (dependence)

NORMS: Proposed Algorithm

INPUT: $X_1, \dots, X_m, \dots, X_M, X_m \in \Re^{n \times d_m}$

OUTPUT: Joint subspace $\Psi(\mathbf{X})$

NORMS: Proposed Algorithm

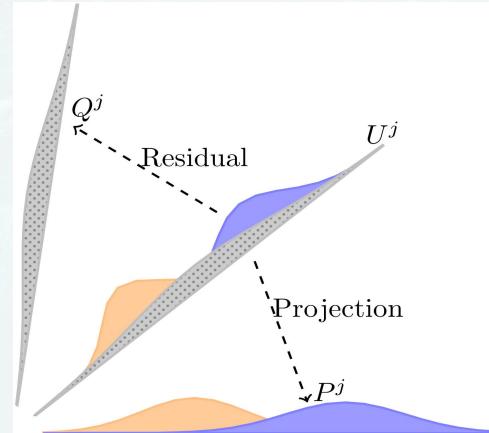
For each view:

Compute rank r_m ,
Principal subspace $\Psi(X_m)$,
Relevance $\mathcal{R}_l(X_m)$

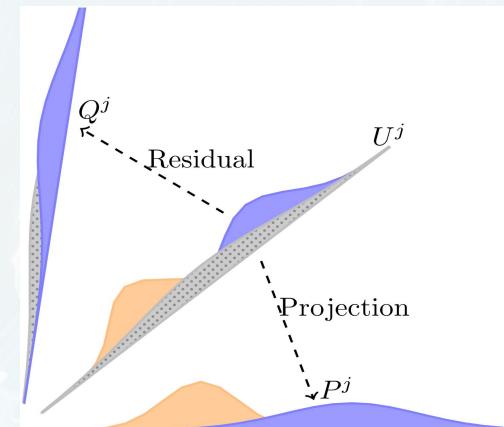
Views with
rank $r_m > 0$

Step 1:
View X_π with
 $\max \mathcal{R}_l(X_m) :$
 $\Psi(\mathbf{X}_1) = \Psi(X_\pi)$

Step $m \leftarrow 1$ to \mathcal{M} :
 $\mathcal{D}(\mathbf{X}_m | X_i)$ of remaining X_i
 Select X_ω with max dependence
 Projection $P : X_\omega$ on \mathbf{X}_m
 Residual $Q = X_\omega - P$
 $S \leftarrow U^j$'s deviating from normality



Normally distributed residuals



Residuals deviate from normality

NORMS: Proposed Algorithm

For each view:

Compute rank r_m ,
 Principal subspace $\Psi(X_m)$,
 Relevance $\mathcal{R}_l(X_m)$

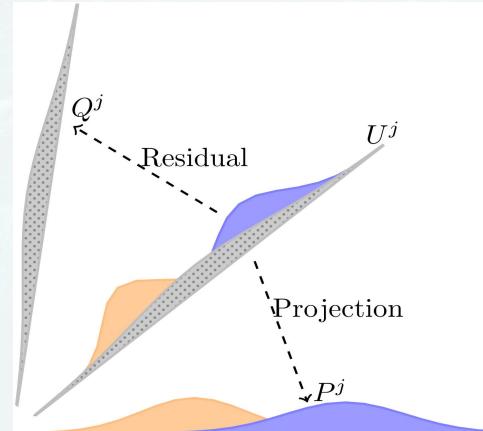
Views with
 rank $r_m > 0$

Step 1:
 View X_π with
 $\max \mathcal{R}_l(X_m) :$
 $\Psi(\mathbf{X}_1) = \Psi(X_\pi)$

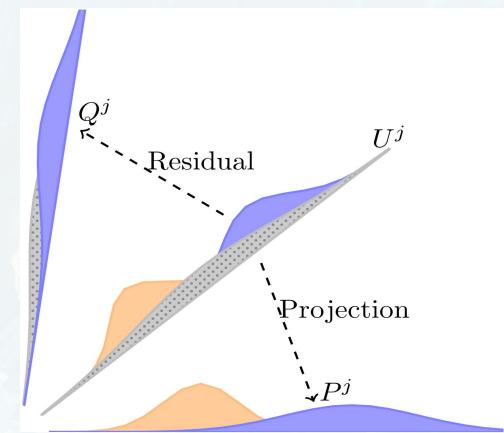
Step $m \leftarrow 1$ to \mathcal{M} :

$$\begin{aligned} U(\mathbf{X}_{m+1}) &= [U(\mathbf{X}_m) \quad S] \\ \Sigma(\mathbf{X}_{m+1}) &= \text{diag}(\Sigma(\mathbf{X}_m), \Sigma(S)) \\ \Psi(\mathbf{X}_{m+1}) &= \langle U(\mathbf{X}_{m+1}), \Sigma(\mathbf{X}_{m+1}) \rangle \end{aligned}$$

$$\begin{aligned} Y(\mathbf{X}_\mathcal{M}) &= U(\mathbf{X}_\mathcal{M}) \Sigma(\mathbf{X}_\mathcal{M}) \\ k\text{-means clustering on } Y(\mathbf{X}_\mathcal{M}) \end{aligned}$$



Normally distributed residuals



Residuals deviate from normality

Experimental Results

4 Multi-Omics Cancer Data Sets:

Ovarian carcinoma (OV)

[334 samples, 4 clusters]

Breast Carcinoma (BRCA)

[398 samples, 4 clusters]

Lower Grade Dlioma (LGG)

[267 samples, 3 clusters]

Cervical carcinoma (CESC)

[124 samples, 3 clusters]

Modalities/ Views:

DNA methylation (mDNA)

Gene expression (RNA)

Micro-RNA expression (miRNA)

Protein expression (RPPA)

The Cancer Genome Atlas (TCGA): <https://cancergenome.nih.gov/>

Relevance & Rank Estimation

Relevance and rank of each view and views selected by NormS algorithm

Modality	Relevance	Rank	Selected	Modality	Relevance	Rank	Selected
mDNA	CESC	0.1884817	3	RNA,	LGG	0.4320317	10
RNA		0.2921399	2	mDNA,		0.0289518	0
miRNA		0.1990886	5	miRNA,		0.0056958	0
RPPA		0.2006048	4	RPPA		0.2428867	6
mDNA	OV	0.0230986	0	RNA,	BRCA	0.2373227	5
RNA		0.4936741	3	miRNA,		0.2947759	3
miRNA		0.2474369	5	RPPA		0.1602746	4
RPPA		0.0579902	2			0.2464338	6

- Views with no signal component (rank 0) are filtered out

Importance of Relevance

Data Set	Different Measures	2nd Most Relevant	3rd Most Relevant	4th Most Relevant	Proposed Algorithm
LGG	Starting Modality	RPPA	RNA	miRNA	mDNA
	Rank of Modality	6	0	0	10
	Relevance	0.2428867	0.0289518	0.0056958	0.4320317
	Accuracy	0.7228464	-	-	0.7940075
	NMI	0.4739360	-	-	0.5325030
	ARI	0.3557794	-	-	0.4649223
	F-measure	0.7236789	-	-	0.7916535
	Rand	0.6978401	-	-	0.7465292
OV	Purity	0.7228464	-	-	0.7940075
	Starting Modality	miRNA	RPPA	mDNA	RNA
	Rank of Modality	5	2	0	3
	Relevance	0.2474369	0.0579902	0.0230986	0.4936741
	Accuracy	0.5568862	0.5568862	-	0.6976048
	NMI	0.2717504	0.2717504	-	0.4504552
	ARI	0.2015805	0.2015805	-	0.4142200
	F-measure	0.5552212	0.5552212	-	0.6910392
Lung	Rand	0.6949524	0.6949524	-	0.7766269
	Purity	0.5568862	0.5568862	-	0.6976048

- Best performance when starting with most relevant view

Ablation Study

Data Set	Different Measures	Fixed Rank ($k - 1$)	Without Dependency	Taking All Residuals	Proposed Algorithm
CESC	Accuracy	0.8387097	0.8790323	0.8387097	0.8870968
	NMI	0.6579328	0.6707970	0.6579328	0.6854921
	ARI	0.6040195	0.6875915	0.6040195	0.7004411
	F-measure	0.8181978	0.8706213	0.8181978	0.8801172
	Rand	0.8080252	0.8526095	0.8080252	0.8587726
	Purity	0.8387097	0.8790323	0.8387097	0.8870968
BRCA	Accuracy	0.7638191	0.7638191	0.7613065	0.7688442
	NMI	0.5556963	0.5231492	0.5517217	0.5437267
	ARI	0.5082782	0.4958426	0.5052503	0.5090183
	F-measure	0.7651760	0.7642758	0.7626970	0.7699789
	Rand	0.8002101	0.7928053	0.7989950	0.7999063
	Purity	0.7638191	0.7638191	0.7613065	0.7688442

- Best performance when considering dependency with residual noise filtering

Comparative Performance Analysis

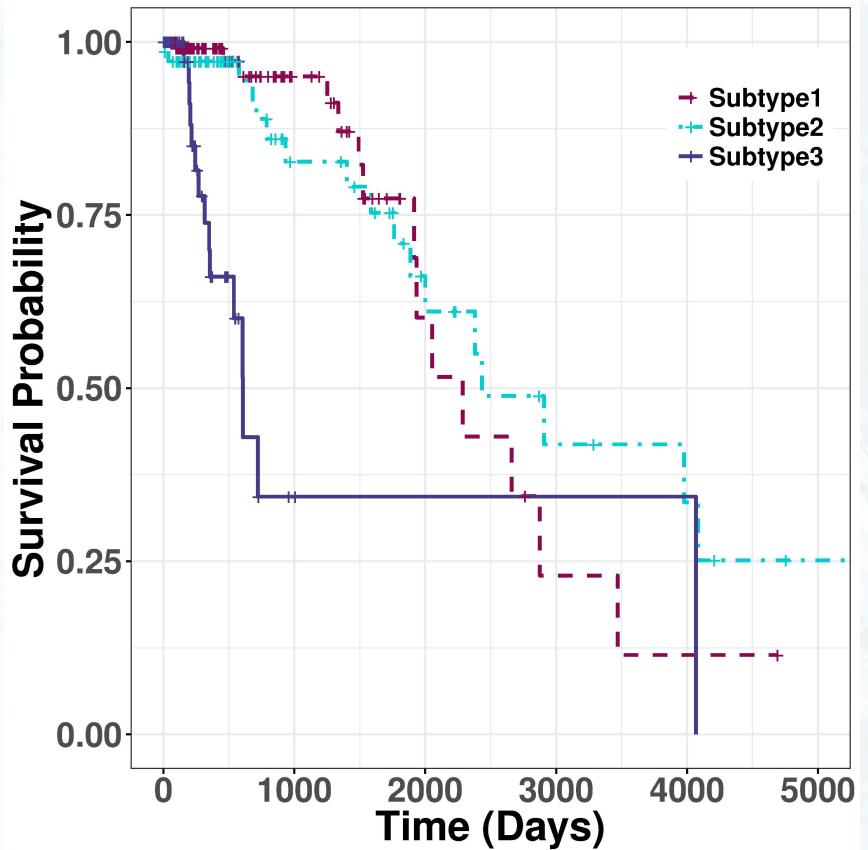
Data Set	Different Algorithms	Rank of Subspace	External Evaluation Index					
			Accuracy	NMI	ARI	F-measure	Rand	Purity
CESC	COCA	-	0.6693548	0.4172592	0.3677157	0.6870510	0.6971282	0.6774194
	BCC	-	0.6895161	0.2854917	0.3144526	0.6795619	0.6687779	0.6935484
	JIVE-Perm	24	0.7177419	0.4425848	0.3860367	0.7097880	0.7164962	0.7177419
	JIVE-BIC	4	0.8064516	0.5296325	0.5229385	0.8011385	0.7791765	0.8064516
	A-JIVE	48	0.6500000	0.3700238	0.3355826	0.6511586	0.6857724	0.6814516
	iCluster	2	0.5483871	0.1737526	0.1017765	0.5568753	0.5731707	0.5645161
	LRAcluster	1	0.8145161	0.5176602	0.5384740	0.8123256	0.7867821	0.8145161
	PCA-con	3	0.8548387	0.6750978	0.6333073	0.8390298	0.8237608	0.8548387
LGG	NormS	6	0.8870968	0.6854921	0.7004411	0.8801172	0.8587726	0.8870968
	COCA	-	0.6591760	0.2772248	0.2533847	0.6608123	0.6454901	0.6591760
	BCC	-	0.6340824	0.2737596	0.248606	0.63111660	0.6382755	0.6355805
	JIVE-Perm	8	0.5617978	0.2299551	0.1606599	0.5757978	0.6056715	0.5730337
	JIVE-BIC	8	0.6741573	0.3441747	0.3050874	0.6679019	0.6642730	0.6741573
	A-JIVE	48	0.7168539	0.4267241	0.3376560	0.7172792	0.6869055	0.7168539
	iCluster	2	0.4382022	0.1379678	0.0996867	0.5187438	0.5821858	0.5355805
	LRAcluster	2	0.4719101	0.1240057	0.1030798	0.5137382	0.5831714	0.5280899
	PCA-con	3	0.6666667	0.3438738	0.3031312	0.6574834	0.6616823	0.6666667
	NormS	14	0.7940075	0.5325030	0.4649223	0.7916535	0.7465292	0.7940075

Comparative Performance Analysis

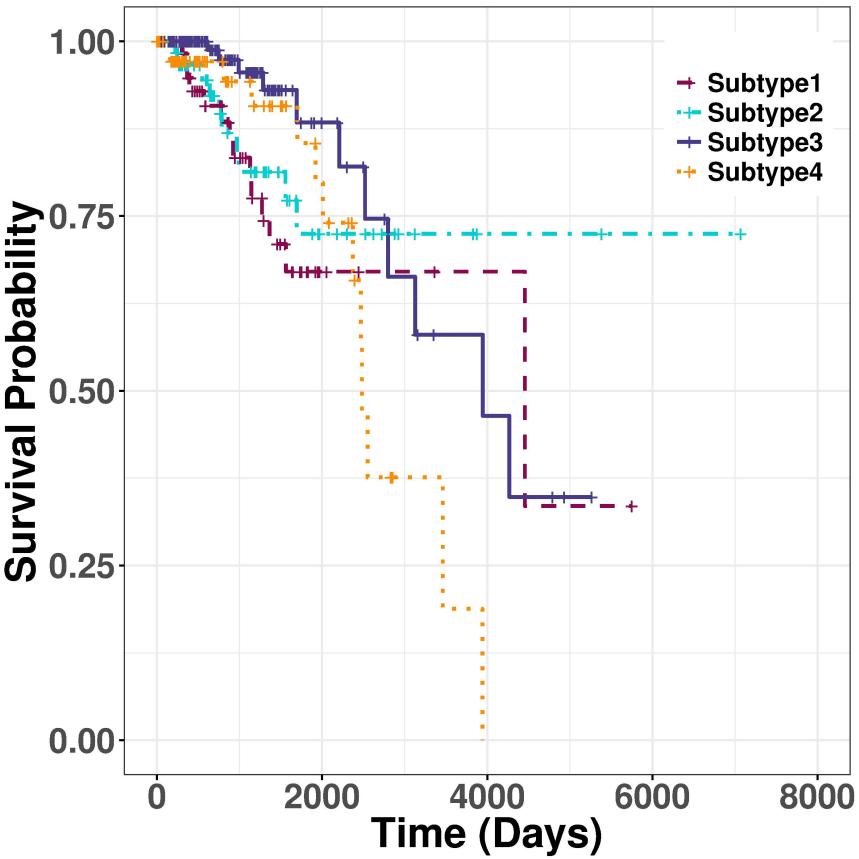
	Data Set	Different Algorithms	Rank of Subspace	External Evaluation Index				
				Accuracy	NMI	ARI	F-measure	Rand
OV	COCA	-	0.5943114	0.3131466	0.2810761	0.6068513	0.7039183	0.5943114
	BCC	-	0.4610778	0.1567582	0.1254690	0.4755846	0.6268706	0.4622754
	JIVE-Perm	32	0.5718563	0.2629523	0.2027605	0.5653910	0.6885005	0.5718563
	A-JIVE	64	0.5191617	0.2124862	0.1981556	0.5111353	0.6942997	0.5221557
	iCluster	3	0.5089820	0.2249889	0.2005886	0.4808256	0.6916078	0.5119760
	LRAcluster	2	0.6287425	0.3745173	0.2999204	0.6384046	0.7322472	0.6287425
	PCA-con	4	0.6946108	0.4424701	0.4068449	0.6868295	0.7734621	0.6946108
	Proposed	10	0.6976048	0.4504552	0.4142200	0.6910392	0.7766269	0.6976048
BRCA	COCA	-	0.7434673	0.5002408	0.4864778	0.7457304	0.7905295	0.7434673
	BCC	-	0.6251256	0.3169187	0.3049874	0.6242493	0.7055783	0.6334171
	JIVE-Perm	12	0.6859296	0.4287142	0.3772649	0.6889363	0.7464906	0.6859296
	JIVE-BIC	4	0.6608040	0.4372675	0.3603942	0.6678438	0.7286432	0.6608040
	A-JIVE	64	0.6140704	0.4482479	0.3710317	0.6707575	0.7363682	0.6841709
	iCluster	3	0.7638191	0.5176193	0.4745746	0.7658865	0.7842867	0.7638191
	LRAcluster	2	0.7110553	0.4368520	0.4035040	0.7101385	0.7521740	0.7110553
	PCA-con	4	0.7587940	0.5506612	0.5038795	0.7601317	0.7984380	0.7587940
	Proposed	11	0.7688442	0.5437267	0.5090183	0.7699789	0.7999063	0.7688442

- NormS has best performance in all four data sets: CESC, LGG, OV, BRCA accross all indices except NMI on BRCA data set

Kaplan-Meir Survival Analysis



LGG Data Set $p=5.901e-09$



BRCA Data Set $p=3.442e-02$

- NormS has identified cancer subtypes with statistically significant difference between survival profiles

Highlights

- Feature-space integration
- Statistical hypothesis testing: estimate rank and signal component of view
- View evaluation measures: Relevance & Dependence
- Select only relevant and non-redundant components for integration
- Clustering on joint subspace
- Experiments: TCGA multi-omics cancer data sets

CHAPTER 4

Selective Update of Relevant Eigenspaces for Integrative Clustering of Multi-View Data

A. Khan and P. Maji, “Selective Update of Relevant Eigenspaces for Integrative Clustering of Multimodal Data,” **IEEE Transactions on Cybernetics (TCYB)**, pp. 1-13, 2020. DOI: [10.1109/TCYB.2020.2990112](https://doi.org/10.1109/TCYB.2020.2990112).

Problem Statement

NormS issues:

- Column-wise concatenation of singular subspaces
- Variable rank joint subspace for different data sets
- Not the principal subspace

Can we construct the rank k principal subspace in less time compared to Principal Component Analysis (PCA) on integrated data?

SVD Eigenspace Model

$$X_m - \mathbf{1}\mu(X_m)^T = U(X_m)\Sigma(X_m)V(X_m)^T$$

Rank k eigenspace of m^{th} view X_m :

$$\Psi(X_m) = \langle \mu(X_m), U(X_m), \Sigma(X_m), V(X_m) \rangle$$

$\xleftarrow{\quad\quad\quad}$
k largest singular triplets of individual view X_m

Feature-wise concatenation of m views

$$\tilde{\mathbf{X}}_m = [X_1 \quad X_2 \quad \dots \quad X_m]$$

Rank k eigenspace of $\tilde{\mathbf{X}}_m$:

$$\Psi(\tilde{\mathbf{X}}_m) = \langle \mu(\tilde{\mathbf{X}}_m), U(\tilde{\mathbf{X}}_m), \Sigma(\tilde{\mathbf{X}}_m), V(\tilde{\mathbf{X}}_m) \rangle$$

$\xleftarrow{\quad\quad\quad}$
k largest singular triplets of joint view at step m

Incremental Update of SVD Eigenspace

$$X_m - \mathbf{1}\mu(X_m)^T = U(X_m)\Sigma(X_m)V(X_m)^T$$

Rank k eigenspace of m^{th} view X_m :

$$\Psi(X_m) = \langle \mu(X_m), U(X_m), \Sigma(X_m), V(X_m) \rangle$$

Feature-wise concatenation of m views

$$\tilde{\mathbf{X}}_m = [X_1 \quad X_2 \quad \dots \quad X_m]$$

Rank k eigenspace of $\tilde{\mathbf{X}}_m$:

$$\Psi(\tilde{\mathbf{X}}_m) = \langle \mu(\tilde{\mathbf{X}}_m), U(\tilde{\mathbf{X}}_m), \Sigma(\tilde{\mathbf{X}}_m), V(\tilde{\mathbf{X}}_m) \rangle$$

Step m+1:

$$\tilde{\mathbf{X}}_{m+1} = \begin{bmatrix} \tilde{\mathbf{X}}_m & X_{m+1} \end{bmatrix}$$

$$\Psi(\tilde{\mathbf{X}}_{m+1}) = \Psi(\tilde{\mathbf{X}}_m) \oplus \Psi(X_{m+1})$$

instead of SVD on $\tilde{\mathbf{X}}_{m+1}$

Incremental Update of SVD Eigenspace

$$\text{Step } m+1: \tilde{\mathbf{X}}_{m+1} = \begin{bmatrix} \tilde{\mathbf{X}}_m & X_{m+1} \end{bmatrix}$$

$$U(\tilde{\mathbf{X}}_m)\Sigma(\tilde{\mathbf{X}}_m)V(\tilde{\mathbf{X}}_m)^T$$

$$U(X_{m+1})\Sigma(X_{m+1})V(X_{m+1})^T$$

$$U(\tilde{\mathbf{X}}_{m+1})\Sigma(\tilde{\mathbf{X}}_{m+1})V(\tilde{\mathbf{X}}_{m+1})^T$$

Sufficient basis that spans both $U(\tilde{\mathbf{X}}_m)$ and $U(X_{m+1})$

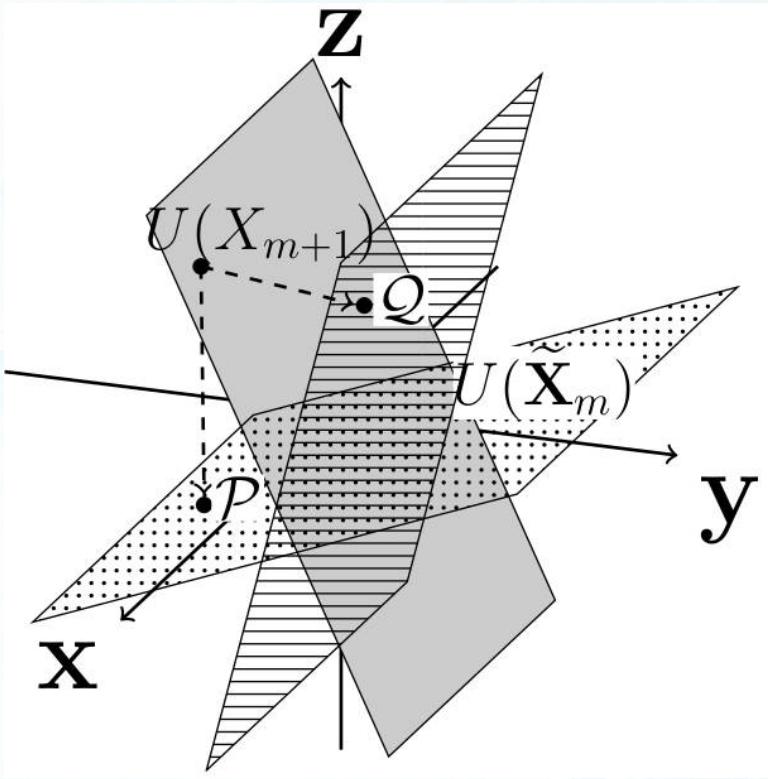
- Right subspaces in V correspond to separate sets of variables
- Left subspaces in U contain singular vectors corresponding to common set of n samples

Incremental Update of SVD Eigenspace

Step $m+1$: $\tilde{\mathbf{X}}_{m+1} = \begin{bmatrix} \tilde{\mathbf{X}}_m & X_{m+1} \end{bmatrix}$

$$U(\tilde{\mathbf{X}}_m)\Sigma(\tilde{\mathbf{X}}_m)V(\tilde{\mathbf{X}}_m)^T$$

$$U(X_{m+1})\Sigma(X_{m+1})V(X_{m+1})^T$$



1. Intersection $\mathcal{I} = U(\tilde{\mathbf{X}}_m)^T U(X_{m+1})$
2. Projection $\mathcal{P} = U(\tilde{\mathbf{X}}_m)\mathcal{I}$
3. Residual $\mathcal{Q} = U(X_{m+1}) - \mathcal{P}$
4. Residual Basis $\Gamma = \text{Gram_Schmidt}(Q)$
5. $U(\tilde{\mathbf{X}}_{m+1}) = \begin{bmatrix} U(\tilde{\mathbf{X}}_m) & \Gamma \end{bmatrix} \mathcal{R}(\tilde{\mathbf{X}}_{m+1})$

Incremental Update of SVD Eigenspace

$$\tilde{\mathbf{X}}_{m+1} = \begin{bmatrix} \tilde{\mathbf{X}}_m & X_{m+1} \end{bmatrix}$$

$$U(\tilde{\mathbf{X}}_{m+1}) = \begin{bmatrix} U(\tilde{\mathbf{X}}_m) & \Gamma \end{bmatrix} \mathcal{R}(\tilde{\mathbf{X}}_{m+1})$$

Substitute

SVD relation:

$$\tilde{\mathbf{X}}_{m+1} - \mathbf{1}\mu(\tilde{\mathbf{X}}_{m+1})^T = U(\tilde{\mathbf{X}}_{m+1})\Sigma(\tilde{\mathbf{X}}_{m+1})V(\tilde{\mathbf{X}}_{m+1})^T$$

Reduces to solving a smaller SVD
problem of size $2(k \times d_{\max})$

$$\mathcal{R}(\tilde{\mathbf{X}}_{m+1})\Sigma(\tilde{\mathbf{X}}_{m+1})V(\tilde{\mathbf{X}}_{m+1})^T = \begin{bmatrix} \mathbf{I}_k\Sigma(\tilde{\mathbf{X}}_m)V(\tilde{\mathbf{X}}_m)^T & \mathcal{I}\Sigma(X_{m+1})V(X_{m+1})^T \\ \mathbf{0} & \Gamma^T U(X_{m+1})\Sigma(X_{m+1})V(X_{m+1})^T \end{bmatrix}.$$

View Evaluation

Relevance

$\text{Rel}(X_m)$ = Between cluster variance
of a k partition of $\text{U}(X_m)$

Concorance

$C(X_i, X_j) = \text{NMI}(\mathcal{C}^i, \mathcal{C}^j)$ Normalized mutual
informationn between k
partition of $\text{U}(X_i)$ and $\text{U}(X_j)$

SURE: Proposed Algorithm

INPUT: $X_1, \dots, X_m, \dots, X_M, X_m \in \Re^{n \times d_m}$

OUTPUT: Joint eigenspace $\Psi(\tilde{\mathbf{X}}_M)$

SURE: Proposed Algorithm

For each view:

Compute SVD eigenspace $\Psi(X_m)$

k -means on left subspace $U(X_m)$

Relevance $\text{Rel}(X_m)$



Step 1:

X_π with max $\text{Rel}(X_m) :$

$$\tilde{\mathbf{X}}_1 = X_\pi, \mathcal{S} = \{X_\pi\}$$

$$\Psi(\tilde{\mathbf{X}}_1) \leftarrow \Psi(X_\pi)$$



Step $m \leftarrow 1$ to $M - 1$:

1. Average concordance:

$$\bar{\mathbb{C}}(X_j) = 1/|\mathcal{S}| \sum_{\omega \in \mathcal{S}} \mathbb{C}(X_\omega, X_j)$$

2. $X_l \leftarrow X_j$ with max $\bar{\mathbb{C}}$

3. if $\bar{\mathbb{C}}(X_l) \geq \tau$

$$\mathcal{S} = \mathcal{S} \cup \{X_l\}$$

$$\Psi(\tilde{\mathbf{X}}_{m+1}) = \Psi(\tilde{\mathbf{X}}_m) \oplus \Psi(X_l)$$

$$= \langle \mu(\tilde{\mathbf{X}}_{m+1}), U(\tilde{\mathbf{X}}_{m+1}), \Sigma(\tilde{\mathbf{X}}_{m+1}), V(\tilde{\mathbf{X}}_{m+1}) \rangle$$



Joint eigenspace: $\Psi(\tilde{\mathbf{X}}_M)$

$$\mathbf{Y} = U(\tilde{\mathbf{X}}_M) \Sigma(\tilde{\mathbf{X}}_M)$$

k -means clustering on \mathbf{Y}

Accuracy of Eigenspace Construction

Concatenation of all views:

$$\mathbf{X} = [X_1, X_2, \dots, X_M]$$

↓
SVD

$$\Psi(\mathbf{X}) = \langle \mu(\mathbf{X}), U(\mathbf{X}^r), \Sigma(\mathbf{X}^r), V(\mathbf{X}^r) \rangle$$

Full Eigenspace of \mathbf{X}

Accuracy of Eigenspace Construction

Concatenation of all views:

$$\mathbf{X} = [X_1, X_2, \dots, X_M]$$

↓
SVD

$$\Psi(\mathbf{X}) = \langle \mu(\mathbf{X}), U(\mathbf{X}^r), \Sigma(\mathbf{X}^r), V(\mathbf{X}^r) \rangle$$

Full Eigenspace of \mathbf{X}

$$\mathbf{X} = [X_1, X_2, \dots, X_M]$$

↓
SVD

$$\Psi(X_1) \quad \Psi(X_2) \quad \Psi(X_M)$$

$$\Psi(\tilde{\mathbf{X}}) = \bigoplus_{m=1}^M \Psi(X_m)$$

$$\Psi(\tilde{\mathbf{X}}) = \langle \mu(\mathbf{X}), U(\tilde{\mathbf{X}}^r), \Sigma(\tilde{\mathbf{X}}^r), V(\tilde{\mathbf{X}}^r) \rangle$$

Approximate Eigenspace of \mathbf{X}

Accuracy of Eigenspace Construction

Concatenation of all views:

$$\mathbf{X} = [X_1, X_2, \dots, X_M]$$

↓
SVD

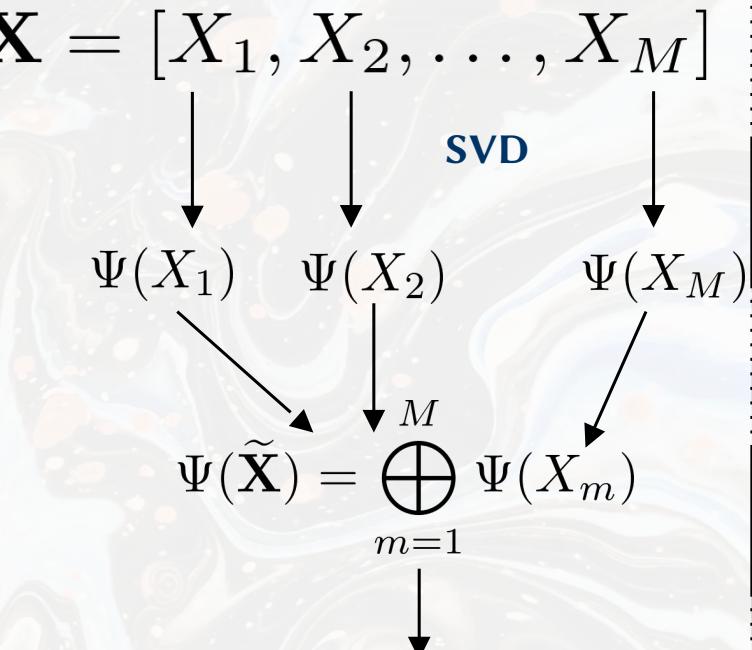
$$\Psi(\mathbf{X}) = \langle \mu(\mathbf{X}), U(\mathbf{X}^r), \Sigma(\mathbf{X}^r), V(\mathbf{X}^r) \rangle$$

Full Eigenspace of \mathbf{X}

How far does the rank r approximate eigenspace $\Psi(\tilde{\mathbf{X}})$ deviate from the full eigenspace $\Psi(\mathbf{X})$?

$$\Psi(\tilde{\mathbf{X}}) = \langle \mu(\mathbf{X}), U(\tilde{\mathbf{X}}^r), \Sigma(\tilde{\mathbf{X}}^r), V(\tilde{\mathbf{X}}^r) \rangle$$

Approximate Eigenspace of \mathbf{X}



Accuracy of Eigenspace Construction

Full Eigenspace of \mathbf{X}

$$\Psi(\mathbf{X}) = \langle \mu(\mathbf{X}), U(\mathbf{X}^r), \Sigma(\mathbf{X}^r), V(\mathbf{X}^r) \rangle$$

Approximate Eigenspace of \mathbf{X}

$$\Psi(\tilde{\mathbf{X}}) = \langle \mu(\mathbf{X}), U(\tilde{\mathbf{X}}^r), \Sigma(\tilde{\mathbf{X}}^r), V(\tilde{\mathbf{X}}^r) \rangle$$

How far does the rank r approximate eigenspace
 $\Psi(\tilde{\mathbf{X}})$ deviate from the full eigenspace $\Psi(\mathbf{X})$?



Cumulative
difference between

Left singular subspaces: $U(\mathbf{X}^r)$ and $U(\tilde{\mathbf{X}}^r)$

Right singular subspaces: $V(\mathbf{X}^r)$ and $V(\tilde{\mathbf{X}}^r)$

Accuracy of Eigenspace Construction

Full Eigenspace of \mathbf{X}

$$\Psi(\mathbf{X}) = \langle \mu(\mathbf{X}), U(\mathbf{X}^r), \Sigma(\mathbf{X}^r), V(\mathbf{X}^r) \rangle$$

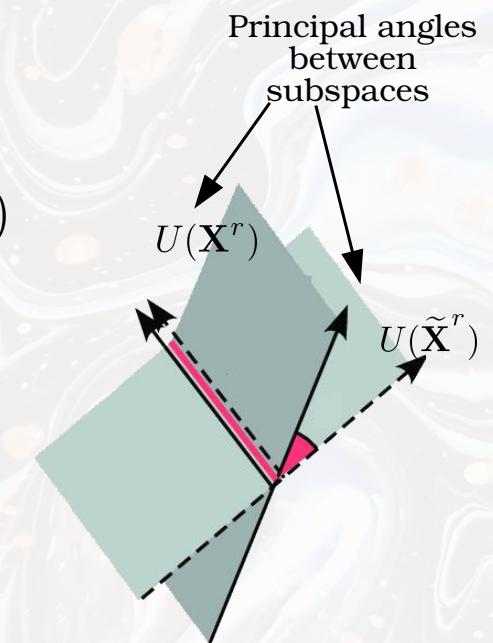
Approximate Eigenspace of \mathbf{X}

$$\Psi(\tilde{\mathbf{X}}) = \langle \mu(\mathbf{X}), U(\tilde{\mathbf{X}}^r), \Sigma(\tilde{\mathbf{X}}^r), V(\tilde{\mathbf{X}}^r) \rangle$$

How far does the rank r approximate eigenspace
 $\Psi(\tilde{\mathbf{X}})$ deviate from the full eigenspace $\Psi(\mathbf{X})$?

Left singular subspaces: $U(\mathbf{X}^r)$ and $U(\tilde{\mathbf{X}}^r)$

$$\| \sin \Theta(U(\mathbf{X}^r), U(\tilde{\mathbf{X}}^r)) \|_F^2 = \sum_{i=1}^r \sin^2 (\theta_i)$$



Accuracy of Eigenspace Construction

Full Eigenspace of \mathbf{X}

$$\Psi(\mathbf{X}) = \langle \mu(\mathbf{X}), U(\mathbf{X}^r), \Sigma(\mathbf{X}^r), V(\mathbf{X}^r) \rangle$$

Approximate Eigenspace of \mathbf{X}

$$\Psi(\tilde{\mathbf{X}}) = \langle \mu(\mathbf{X}), U(\tilde{\mathbf{X}}^r), \Sigma(\tilde{\mathbf{X}}^r), V(\tilde{\mathbf{X}}^r) \rangle$$

MATRIX PERTURBATION THEORY

$$\begin{aligned}\mathbf{X} &= [X_1 \quad \dots \quad X_m \quad \dots \quad X_M] \\ &= [(X_1^r + X_1^{r\perp}) \quad \dots \quad (X_m^r + X_m^{r\perp}) \quad \dots \quad (X_M^r + X_M^{r\perp})] \\ &= \mathbf{X}^r + \mathbf{X}^{r\perp} \\ \Psi(\mathbf{X}) &\quad \downarrow \text{Corresponding eigenspace} \\ \Psi(\tilde{\mathbf{X}}) &\end{aligned}$$

$\Psi(\mathbf{X})$ is a perturbation of $\Psi(\tilde{\mathbf{X}})$

Accuracy of Eigenspace Construction

Full Eigenspace of \mathbf{X}

$$\Psi(\mathbf{X}) = \langle \mu(\mathbf{X}), U(\mathbf{X}^r), \Sigma(\mathbf{X}^r), V(\mathbf{X}^r) \rangle$$

Approximate Eigenspace of \mathbf{X}

$$\Psi(\tilde{\mathbf{X}}) = \langle \mu(\mathbf{X}), U(\tilde{\mathbf{X}}^r), \Sigma(\tilde{\mathbf{X}}^r), V(\tilde{\mathbf{X}}^r) \rangle$$

MATRIX PERTURBATION THEORY

$$Gap\Theta\left(\mathbf{X}^r, \tilde{\mathbf{X}}^r\right) = \left[\frac{1}{2r} \left\{ \| \sin \Theta(U(\mathbf{X}^r), U(\tilde{\mathbf{X}}^r)) \|_F^2 + \| \sin \Theta(V(\mathbf{X}^r), V(\tilde{\mathbf{X}}^r)) \|_F^2 \right\} \right]^{\frac{1}{2}}$$

root mean squared principal sines between left and right singular subspaces

Accuracy of Eigenspace Construction

Full Eigenspace of \mathbf{X}

$$\Psi(\mathbf{X}) = \langle \mu(\mathbf{X}), U(\mathbf{X}^r), \Sigma(\mathbf{X}^r), V(\mathbf{X}^r) \rangle$$

Approximate Eigenspace of \mathbf{X}

$$\Psi(\tilde{\mathbf{X}}) = \langle \mu(\mathbf{X}), U(\tilde{\mathbf{X}}^r), \Sigma(\tilde{\mathbf{X}}^r), V(\tilde{\mathbf{X}}^r) \rangle$$

WEDIN'S SIN THETA THEOREM

$$Gap\Theta(\mathbf{X}^r, \tilde{\mathbf{X}}^r) \leq \frac{\sqrt{\|\mathbb{R}_L\|_F^2 + \|\mathbb{R}_R\|_F^2}}{\sqrt{2r}\delta}$$

Residual components of $\mathbb{R}_L = \mathbf{X}^r V(\mathbf{X}^r) - U(\mathbf{X}^r) \Sigma(\mathbf{X}^r)$;

left and right subspaces

$$\mathbb{R}_R = (\mathbf{X}^r)^T U(\mathbf{X}^r) - V(\mathbf{X}^r) \Sigma(\mathbf{X}^r)$$

Gap between eigenvalues $\delta = \min \left\{ \min_{1 \leq i \leq r, 1 \leq j \leq (n-r)} |\sigma_i - \tilde{\sigma}_{r+j}|, \min_{1 \leq i \leq r} \sigma_i \right\}$

Accuracy of Eigenspace Construction

Full Eigenspace of \mathbf{X}

$$\Psi(\mathbf{X}) = \langle \mu(\mathbf{X}), U(\mathbf{X}^r), \Sigma(\mathbf{X}^r), V(\mathbf{X}^r) \rangle$$

Approximate Eigenspace of \mathbf{X}

$$\Psi(\tilde{\mathbf{X}}) = \langle \mu(\mathbf{X}), U(\tilde{\mathbf{X}}^r), \Sigma(\tilde{\mathbf{X}}^r), V(\tilde{\mathbf{X}}^r) \rangle$$

WEDIN'S SIN THETA THEOREM

$$Gap\Theta(\mathbf{X}^r, \tilde{\mathbf{X}}^r) \leq \frac{\sqrt{\|\mathbb{R}_L\|_F^2 + \|\mathbb{R}_R\|_F^2}}{\sqrt{2r}\delta}$$

$$\lim_{r \rightarrow n} Gap\Theta(\mathbf{X}^r, \tilde{\mathbf{X}}^r) = 0$$

Full eigenspace converges to approximate eigenspace

Experimental Results

7 Multi-Omics Cancer Data Sets:

Glioblastoma Multiforme (GBM) [167 samples, 4 clusters]

Lung Carcinoma (LUNG) [671 samples, 2 clusters]

Kidney Carcinoma (KIDNEY) [737 samples, 3 clusters]

CESC, LGG, BRCA, OV

Modalities/ Views:

DNA methylation (mDNA)

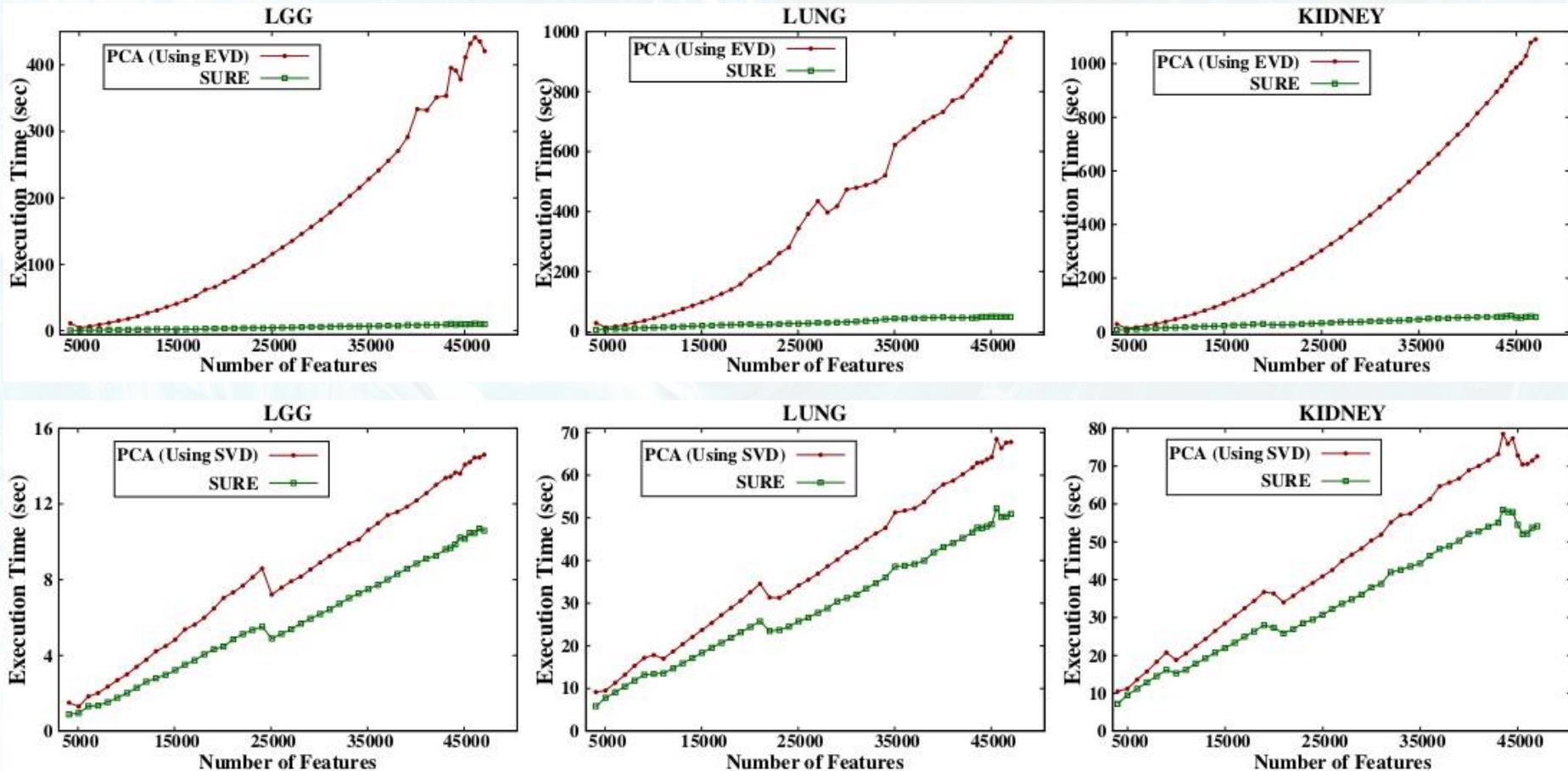
Gene expression (RNA)

Micro-RNA expression (miRNA)

Protein expression (RPPA)

The Cancer Genome Atlas (TCGA): <https://cancergenome.nih.gov/>

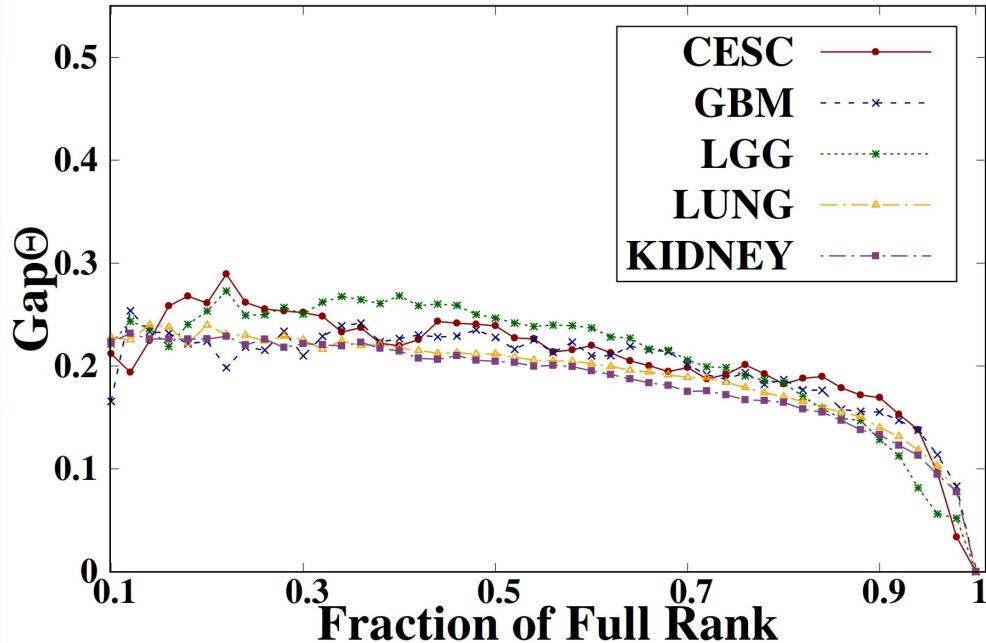
Execution Efficiency



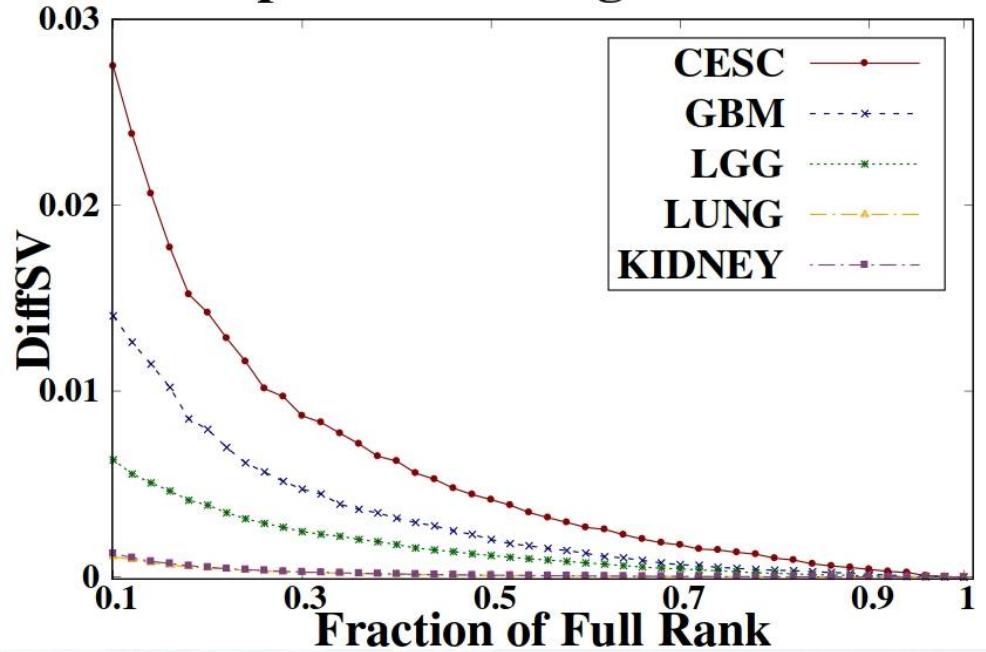
- SURE algorithm takes less time compared to PCA computed using both EVD and SVD

Eigenspace Accuracy

Gap Between Principal Angles



Gap Between Singular Values



- Approximate eigenspace converges to full eigenspace as rank r approaches the full rank

Importance of Data Integration

	Modality/ Algorithm	Accuracy	NMI	ARI	F-Measure	Rand	Purity
CESC	mDNA	0.5241935	0.2420431	0.1175554	0.5453798	0.5819565	0.5806452
	RNA	0.8467742	0.6242327	0.6168352	0.8310850	0.8164175	0.8467742
	miRNA	0.5564516	0.1589298	0.1512301	0.5697384	0.6087071	0.5887097
	RPPA	0.5000000	0.0803494	0.0917321	0.5166786	0.5847102	0.5322581
	PCA_subset	0.8145161	0.5868124	0.5579264	0.7882956	0.7844217	0.8145161
	SURE	0.8629032	0.6461946	0.6507274	0.8512028	0.833989	0.8629032
Best PCA subset:		RNA, miRNA, RPPA					
Subset selected by SURE:		RNA, miRNA					
GBM	RNA	0.7619048	0.5636125	0.4870354	0.7775749	0.8029655	0.7619048
	miRNA	0.6071429	0.3636915	0.3329748	0.6408620	0.7343171	0.6547619
	CNV	0.4166667	0.1207564	0.1061846	0.4678243	0.5688623	0.4464286
	PCA_subset	0.7916667	0.5729951	0.5441936	0.8072570	0.8244226	0.7916667
	SURE	0.797619	0.5815764	0.5588514	0.8120413	0.8300542	0.7976190
	Best PCA subset:		RNA, miRNA, CNV				
Subset selected by SURE:		RNA, miRNA, CNV					
LGG	mDNA	0.7940075	0.5335888	0.4668931	0.7904750	0.7465292	0.7940075
	RNA	0.659176	0.2782794	0.2558892	0.6600498	0.6461660	0.6591760
	miRNA	0.4007491	0.0318103	0.0251035	0.4425295	0.5499986	0.5018727
	RPPA	0.5767790	0.1808821	0.1435186	0.5820448	0.5910563	0.5767790
	PCA_subset	0.6554307	0.3414426	0.2968495	0.6576214	0.6572893	0.6554307
	SURE	0.7940075	0.5335888	0.4668931	0.7904750	0.7465292	0.7940075
Best PCA subset:		mDNA, RNA, miRNA					
Subset selected by SURE:		mDNA					

- **SURE algorithm has better performance compared to individual views**

Importance of Relevance Based Ordering

	Integration starts with	Starting view	Relevance of view	Selected views (in order)	External evaluation index			
					NMI	ARI	F-Measure	Purity
CESC	2 nd best	mDNA	0.47007	{mDNA}	0.24204	0.11755	0.54537	0.58064
	3 rd best	RPPA	0.45550	{RPPA, mDNA, miRNA, RNA}	0.67509	0.63330	0.83902	0.85483
	4 th best	miRNA	0.44951	{miRNA, RNA}	0.64619	0.65072	0.85120	0.86290
	SURE	RNA	0.47533	{RNA, miRNA}	0.64619	0.65072	0.85120	0.86290
GBM	2 nd best	CNV	0.48196	{CNV}	0.12075	0.10618	0.46782	0.44642
	3 rd best	miRNA	0.43332	{miRNA, RNA, CNV}	0.58157	0.55885	0.81204	0.79761
	SURE	RNA	0.50859	{RNA, miRNA, CNV}	0.58157	0.55885	0.81204	0.79761
LGG	2 nd best	RNA	0.44798	{RNA, RPPA, mDNA, miRNA}	0.34387	0.30313	0.65748	0.66666
	3 rd best	RPPA	0.43591	{RPPA, RNA, mDNA, miRNA}	0.34387	0.30313	0.65748	0.66666
	4 th best	miRNA	0.42871	{miRNA}	0.03181	0.02510	0.44252	0.50187
	SURE	mDNA	0.50396	{mDNA}	0.53358	0.46689	0.79047	0.79400

- Best performance when starting with most relevant view

Importance of Concordance

Data Set	Algorithm Settings	External Evaluation Index					
		Accuracy	NMI	ARI	F-Measure	Rand	Purity
CESC	Without C	0.8548387	0.6750978	0.6333073	0.8390298	0.8237608	0.8548387
	SURE	0.8629032	0.6461946	0.6507274	0.8512028	0.833989	0.8629032
GBM	Without C	0.797619	0.5815764	0.5588514	0.8120413	0.8300542	0.797619
	SURE	0.797619	0.5815764	0.5588514	0.8120413	0.8300542	0.797619
LGG	Without C	0.6666667	0.3438738	0.3031312	0.6574834	0.6616823	0.6666667
	SURE	0.7940075	0.5335888	0.4668931	0.7904750	0.7465292	0.7940075
LUNG	Without C	0.9418778	0.6878184	0.7806842	0.9417093	0.8903486	0.9418778
	SURE	0.9418778	0.6878184	0.7806842	0.9417093	0.8903486	0.9418778
KIDNEY	Without C	0.9525102	0.7726162	0.8534490	0.9530685	0.9269512	0.9525102
	SURE	0.9525102	0.7726162	0.8534490	0.9530685	0.9269512	0.9525102

- Clustering performance degrades for some data sets when not considering the concordance between views

Comparative Performance Analysis

	Different Algorithms	Rank of Subspace	External Evaluation Index					
			Accuracy	NMI	ARI	F-Measure	RAND	Purity
CESC	COCA	-	0.6693548	0.4172592	0.3677157	0.6870510	0.6971282	0.6774194
	BCC	-	0.6895161	0.2854917	0.3144526	0.6795619	0.6687779	0.6935484
	JIVE-Perm	24	0.7177419	0.4425848	0.3860367	0.7097880	0.7164962	0.7177419
	JIVE-BIC	4	0.8064516	0.5296325	0.5229385	0.8011385	0.7791765	0.8064516
	A-JIVE	48	0.6500000	0.3700238	0.3355826	0.6511586	0.6857724	0.6814516
	iCluster	2	0.5483871	0.1737526	0.1017765	0.5568753	0.5731707	0.5645161
	LRAcluster	1	0.8145161	0.5176602	0.5384740	0.8123256	0.7867821	0.8145161
	PCA-con	3	0.8548387	0.6750978	0.6333073	0.8390298	0.8237608	0.8548387
	NormS	6	0.8870968	0.6854921	0.7004411	0.8801172	0.8587726	0.8870968
	SURE	3	0.8629032	0.6461946	0.6507274	0.8512028	0.8339890	0.8629032
GBM	COCA	-	0.6863095	0.3682423	0.3367219	0.6771487	0.7251354	0.6863095
	BCC	-	0.4113095	0.1273042	0.0578081	0.4363617	0.5873253	0.4511905
	JIVE-Perm	12	0.6666667	0.3802445	0.3664034	0.6909252	0.7566296	0.6726190
	A-JIVE	36	0.6940476	0.4829471	0.4580430	0.7211722	0.7907898	0.7208333
	iCluster	3	0.7678571	0.5441494	0.5298306	0.7850480	0.8182207	0.7678571
	LRAcluster	3	0.7678571	0.5421434	0.5201970	0.7894569	0.8152267	0.7678571
	PCA-Con	4	0.7916667	0.5729951	0.5441936	0.8072570	0.8244226	0.7916667
	NormS	13	0.6964286	0.4610496	0.4593267	0.7190554	0.7931993	0.7023810
	SURE	4	0.7976190	0.5815764	0.5588514	0.8120413	0.8300542	0.7976190

Comparative Performance Analysis

	Different Algorithms	Rank of Subspace	External Evaluation Index					
			Accuracy	NMI	ARI	F-Measure	RAND	Purity
KIDNEY	COCA	-	0.9408280	0.7493140	0.8393954	0.9477422	0.9199568	0.9470828
	BCC	-	0.9122117	0.6783448	0.7299573	0.9139998	0.8657292	0.9122117
	JIVE-Perm	12	0.9308005	0.6955325	0.7786981	0.9300085	0.8893944	0.9308005
	JIVE-BIC	12	0.9253731	0.6777835	0.7724587	0.9250073	0.8863305	0.9253731
	A-JIVE	48	0.9582090	0.7902576	0.8695284	0.9585611	0.9349404	0.9582090
	iCluster	2	0.6065129	0.2547010	0.1717458	0.6514716	0.5842023	0.6811398
	LRAcluster	2	0.9538670	0.7862018	0.8579391	0.9545717	0.9292298	0.9538670
	PCA-Con	3	0.9511533	0.7670505	0.8489024	0.9516854	0.9246800	0.9511533
	NormS	35	0.9525102	0.7726162	0.8534490	0.9530685	0.9269512	0.9525102
	SURE	3	0.9525102	0.7726162	0.8534490	0.9530685	0.9269512	0.9525102
OV	COCA	-	0.5943114	0.3131466	0.2810761	0.6068513	0.7039183	0.5943114
	BCC	-	0.4610778	0.1567582	0.1254690	0.4755846	0.6268706	0.4622754
	JIVE-Perm	32	0.5718563	0.2629523	0.2027605	0.5653910	0.6885005	0.5718563
	A-JIVE	64	0.5191617	0.2124862	0.1981556	0.5111353	0.6942997	0.5221557
	iCluster	3	0.5089820	0.2249889	0.2005886	0.4808256	0.6916078	0.5119760
	LRAcluster	2	0.6287425	0.3745173	0.2999204	0.6384046	0.7322472	0.6287425
	PCA-con	4	0.6946108	0.4424701	0.4068449	0.6868295	0.7734621	0.6946108
	NormS	10	0.6976048	0.4504552	0.4142200	0.6910392	0.7766269	0.6976048
	SURE	4	0.7215569	0.4680312	0.4372574	0.7148805	0.7857258	0.7215569

Comparative Performance Analysis

	Different Algorithms	Rank of Subspace	External Evaluation Index					
			Accuracy	NMI	ARI	F-Measure	RAND	Purity
LGG	COCA	-	0.6591760	0.2772248	0.2533847	0.6608123	0.6454901	0.6591760
	BCC	-	0.6340824	0.2737596	0.248606	0.63111660	0.6382755	0.6355805
	JIVE-Perm	8	0.5617978	0.2299551	0.1606599	0.5757978	0.6056715	0.5730337
	JIVE-BIC	8	0.6741573	0.3441747	0.3050874	0.6679019	0.6642730	0.6741573
	A-JIVE	48	0.7168539	0.4267241	0.3376560	0.7172792	0.6869055	0.7168539
	iCluster	2	0.4382022	0.1379678	0.0996867	0.5187438	0.5821858	0.5355805
	LRAcluster	2	0.4719101	0.1240057	0.1030798	0.5137382	0.5831714	0.5280899
	PCA-con	3	0.6666667	0.3438738	0.3031312	0.6574834	0.6616823	0.6666667
	NormS	14	0.7940075	0.5325030	0.4649223	0.7916535	0.7465292	0.7940075
	SURE	3	0.7940075	0.5335888	0.4668931	0.790475	0.7465292	0.7940075
LUNG	COCA	-	0.9284650	0.6287671	0.7339231	0.9283705	0.8669662	0.9284650
	BCC	-	0.9372578	0.6648076	0.7645295	0.9371445	0.8822697	0.9372578
	JIVE-Perm	8	0.9269747	0.6333526	0.7288041	0.9266709	0.8644127	0.9269747
	JIVE-BIC	8	0.9388972	0.6883994	0.7701592	0.9385860	0.8850902	0.9388972
	A-JIVE	32	0.9478390	0.7192028	0.8019299	0.9476450	0.9009720	0.9478390
	iCluster	1	0.6333830	0.0627751	0.0696293	0.6299231	0.5348889	0.6333830
	LRAcluster	1	0.9344262	0.6535038	0.7545277	0.9342966	0.8772694	0.9344262
	PCA-Con	2	0.9388972	0.6773549	0.7701654	0.9386955	0.8850902	0.9388972
	NormS	27	0.9359165	0.6650183	0.7597192	0.9357050	0.8798674	0.9359165
	SURE	2	0.9418778	0.6878184	0.7806842	0.9417093	0.8903486	0.9418778

- **SURE approach** performs better than all the existing approaches for most of the external indices, on four data sets: GBM, LGG, LUNG, and OV
- NormS has best in CESC and BRCA data sets, LRAcluster has that for KIDNEY data set, SURE is second best in these three data sets

Highlights

- Feature-space integration
- Construct the rank k SVD eigenspace of integrated data from rank k eigenspaces of individual views
- Theoretically quantify how far the proposed approximate eigenspace deviates from the full eigenspace using matrix perturbation theory
- Extract principal subspace of integrated data in less time compared to PCA
- View evaluation measures: Relevance & Concordance
Select only relevant views with maximum shared cluster structure
- Clustering on joint eigenspace
- Experiments: TCGA multi-omics cancer data sets

FEATURE-SPACE INTEGRATION: DRAWBACK

$$\mathbf{X}_1 \in \mathbb{R}^{n \times d_1}$$

$$\mathbf{X}_2 \in \mathbb{R}^{n \times d_2}$$

...

$$\mathbf{X}_M \in \mathbb{R}^{n \times d_M}$$

β -values $\in [0, 1]$

segment mean in \log_2

RPKM $\sim 10^6$

DATA HETEROGENITY IN TERMS OF UNIT, VARIANCE, AND SCALE
AS FEATURE-SPACE METHODS WORK WITH RAW DATA MATRICES

CHAPTER 5

Approximate Graph Laplacians for Multi-View Data Clustering

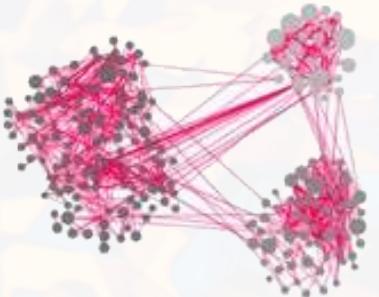
A. Khan and P. Maji, “Approximate Graph Laplacians for Multimodal Data Clustering,” IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI), vol. 43, no. 3, pp. 798-813, 2021. DOI: 10.1109/TPAMI.2019.2945574.

GRAPH BASED INTEGRATION

$$\mathbf{X}_1 \in \mathbb{R}^{n \times d_1}$$

$$\mathbf{X}_2 \in \mathbb{R}^{n \times d_2}$$

$$\mathbf{X}_M \in \mathbb{R}^{n \times d_M}$$



$$\mathbf{X}_1 \rightarrow \mathbf{G}_1$$

n node graph

Vertices: samples

Edge weights: pairwise similarities

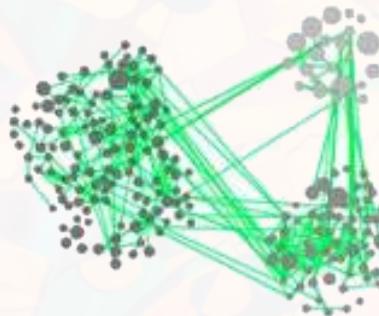
Gaussian kernel / k-nearest neighbors

GRAPH BASED INTEGRATION

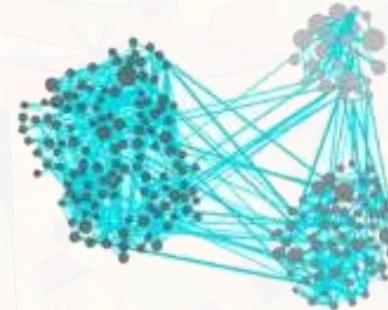
$X_1 \rightarrow G_1$



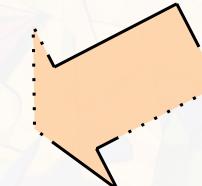
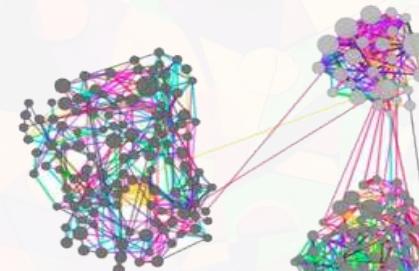
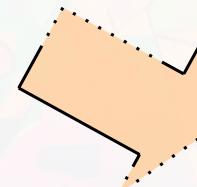
$X_2 \rightarrow G_2$



$X_M \rightarrow G_M$

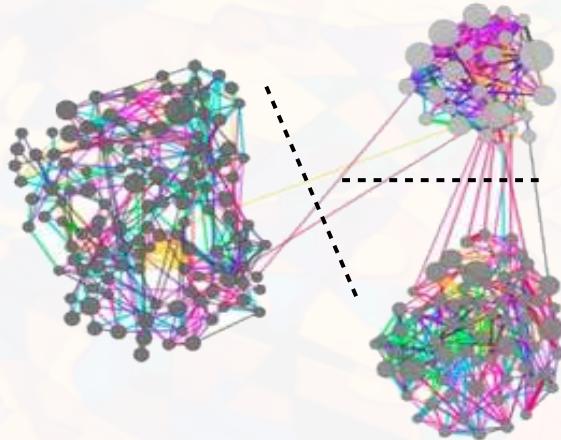


...



Clusters: Spectral Clustering

SPECTRAL CLUSTERING



Graph **G**

Weighted adjacency $W = [w(i, j)]_{n \times n}$

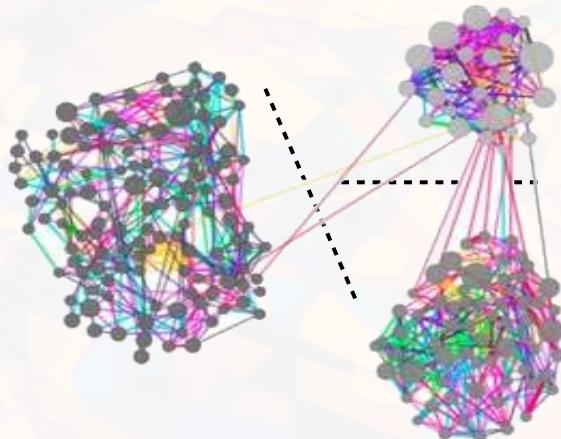
Degree $D = \text{diag}(d_1, \dots, d_n), d_i = \sum_{j=1}^n w(i, j)$

Laplacian $L = \mathbf{I}_n + D^{-1/2}WD^{-1/2}$



Shifted Laplacian

SPECTRAL CLUSTERING



$$\{0, 1\}^{n \times k}$$

Discrete cluster
indicator

Graph \mathbf{G}

$$\text{Laplacian } L = \mathbf{I}_n + D^{-1/2}WD^{-1/2}$$

$$\text{Eigen decomposition } L = U\Sigma U^T$$

$$\begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 0 & \dots & 1 \\ 0 & 1 \\ 0 & 1 \end{bmatrix}$$

$$\begin{bmatrix} u_1 & u_2 & \dots & u_k & u_{k+1} & \dots & u_n \end{bmatrix}$$

$$U^k$$

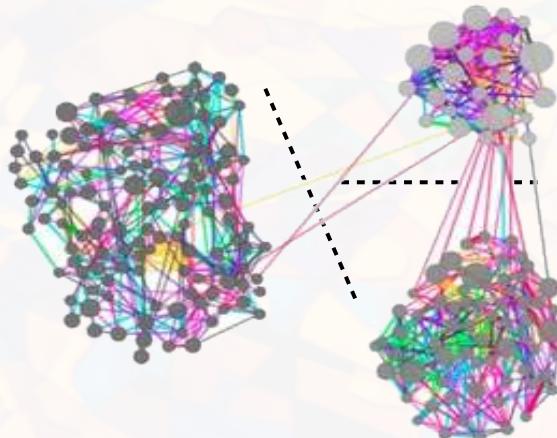
$$\mathbb{R}^{n \times (n-k)}$$

real-valued
relaxation

partition

noise

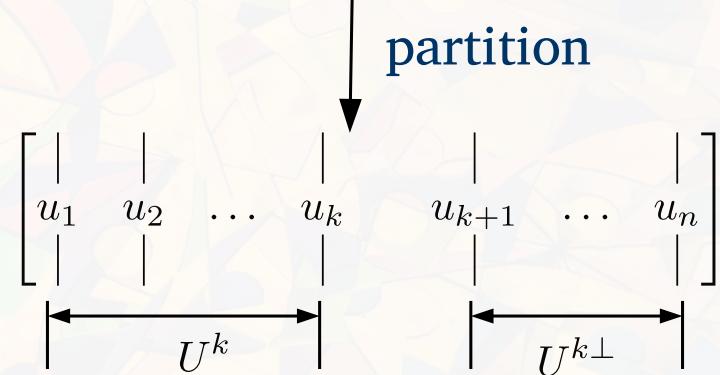
SPECTRAL CLUSTERING



Graph \mathbf{G}

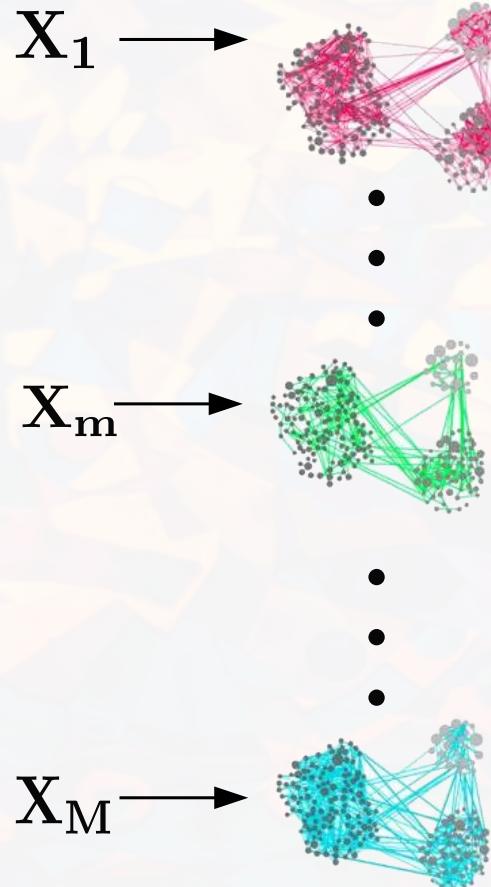
$$\text{Laplacian } L = \mathbf{I}_n + D^{-1/2}WD^{-1/2}$$

$$\text{Eigen decomposition } L = U\Sigma U^T$$

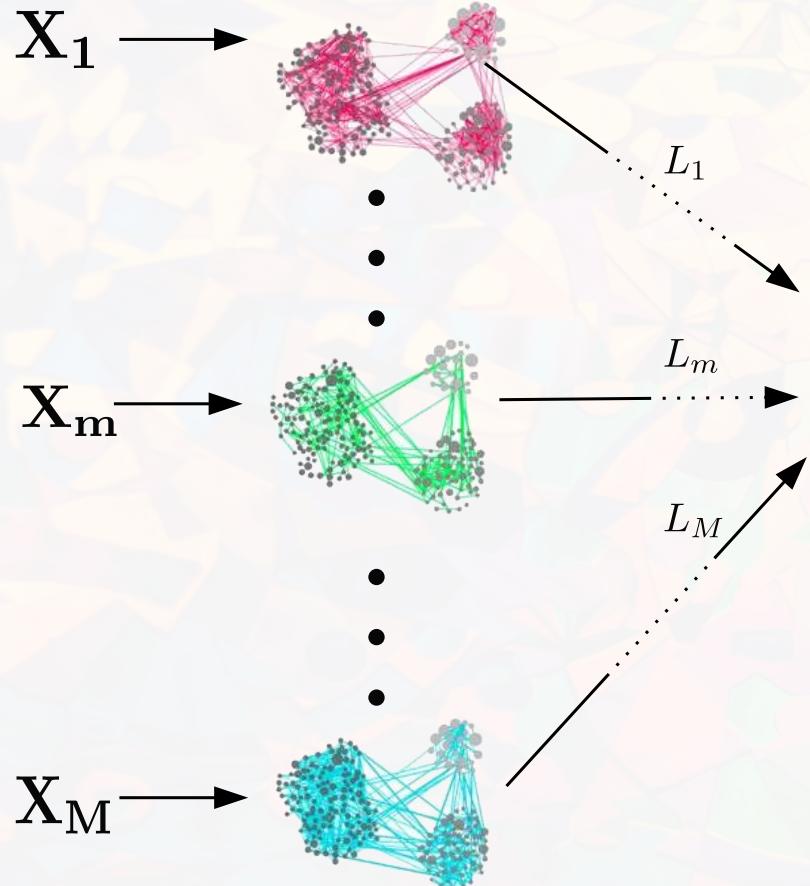


k-means

LAPLACIAN APPROXIMATION

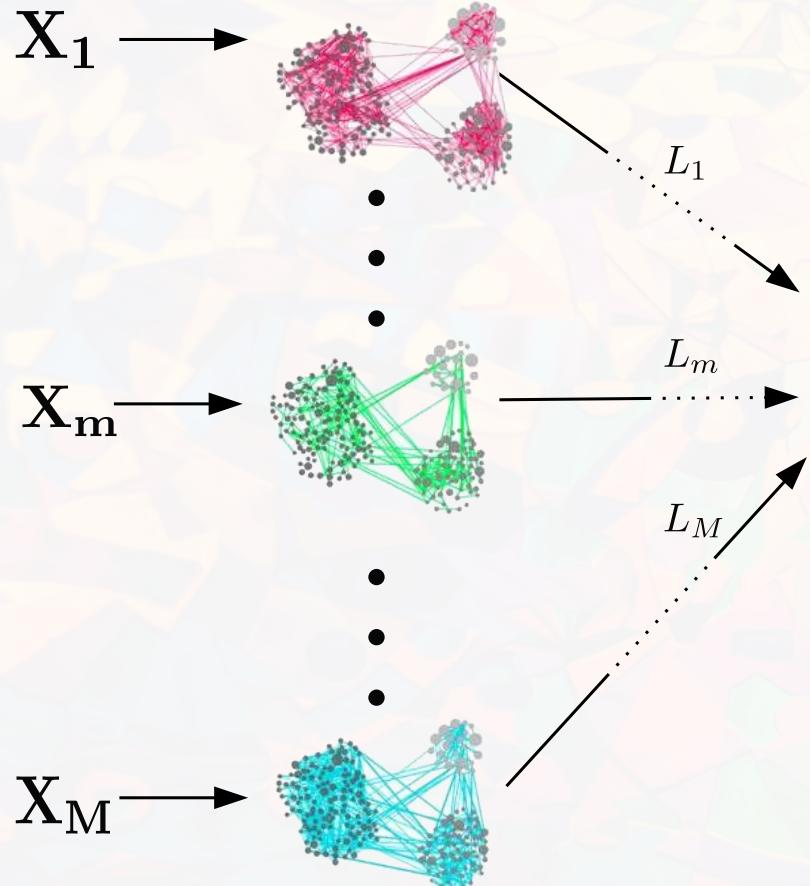


LAPLACIAN APPROXIMATION



Prevent noise of individual graphs from being propagated into unified graph

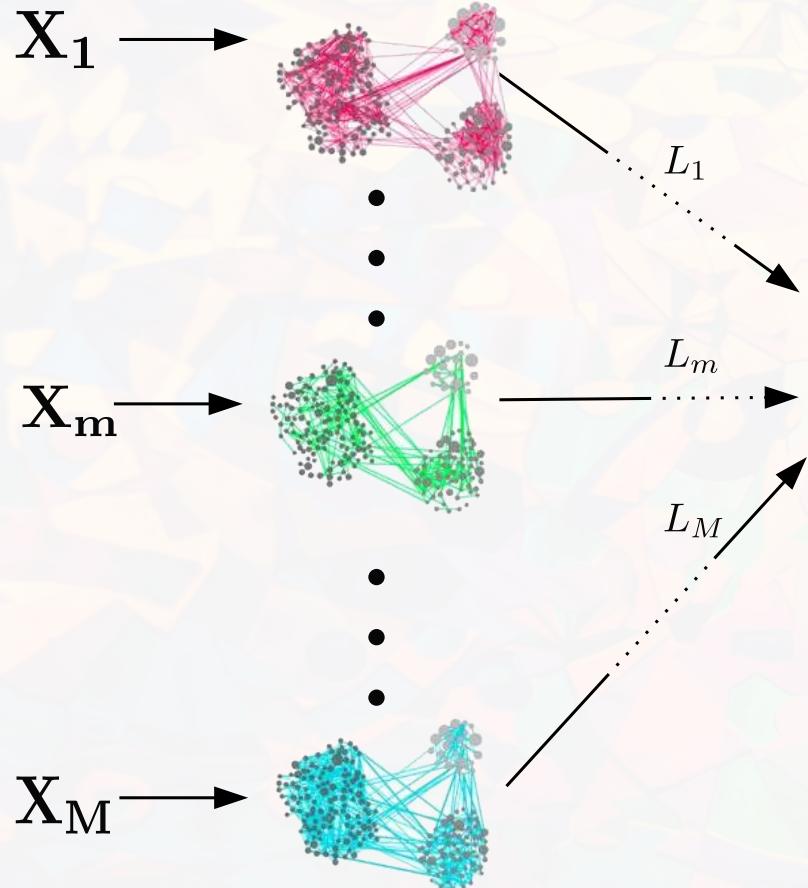
LAPLACIAN APPROXIMATION



Eigen decomposition $L_m = U_m \Sigma_m U_m^T$

$$\begin{array}{c} \text{partition} \\ [U_m^r \quad U_m^{r\perp}] \\ r \geq k \end{array} \quad \begin{bmatrix} \Sigma_m^r & \mathbf{0} \\ \mathbf{0} & \Sigma_m^{r\perp} \end{bmatrix}$$

LAPLACIAN APPROXIMATION



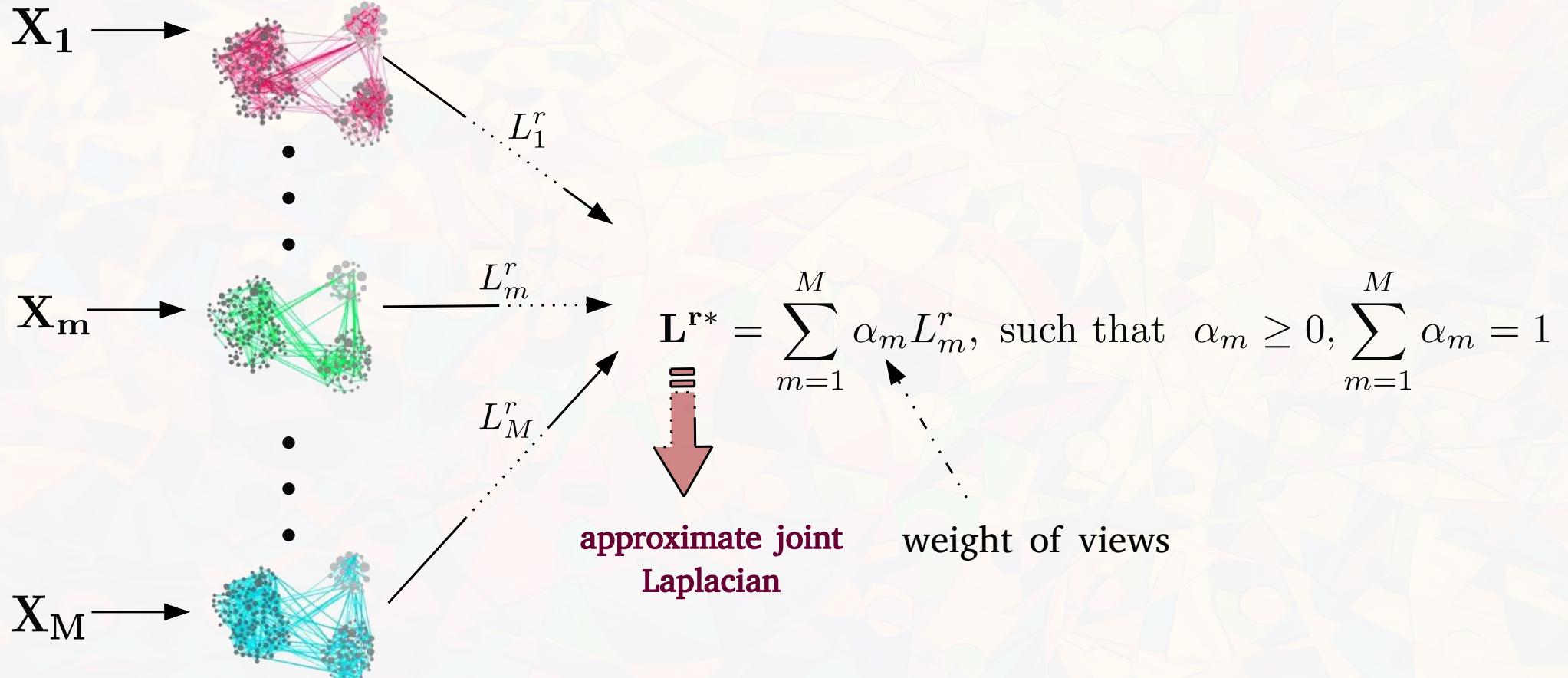
Eigen decomposition $L_m = U_m \Sigma_m U_m^T$

$$\begin{matrix} & \text{partition} \\ [U_m^r \quad U_m^{r\perp}] & \xrightarrow{r \geq k} \left[\begin{matrix} \Sigma_m^r & 0 \\ 0 & \Sigma_m^{r\perp} \end{matrix} \right] \end{matrix}$$

Only the subspace containing cluster information

$$L_m^r = U_m^r \Sigma_m^r (U_m^r)^T$$

LAPLACIAN APPROXIMATION



EIGENSPACE OF JOINT LAPLACIAN

$$\mathbf{L}^{r*} = \sum_{m=1}^M \alpha_m L_m^r = \sum_{m=1}^M \alpha_m U_m^r \Sigma_m^r (U_m^r)^T.$$

Let eigen decomposition be given by $\mathbf{L}^{r*} = \mathbf{V}\Pi\mathbf{V}^T$

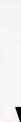
→ Construct components \mathbf{V} and Π from individual eigendecompositions U_m^r and Σ_m^r for $m = 1, 2, \dots, M$

CONSTRUCTION OF JOINT BASIS

$$\mathbf{L}^{r*} = \sum_{m=1}^M \alpha_m L_m^r = \sum_{m=1}^M \alpha_m U_m^r \Sigma_m^r (U_m^r)^T.$$

Let eigen decomposition be given by $\mathbf{L}^{r*} = \mathbf{V} \boldsymbol{\Pi} \mathbf{V}^T$

$$\text{span}(\mathbf{V}) = \bigcup_{m=1}^M \text{span}(U_m^r)$$



column space of \mathbf{V}

CONSTRUCTION OF JOINT BASIS

$$\mathbf{L}^{r*} = \sum_{m=1}^M \alpha_m L_m^r = \sum_{m=1}^M \alpha_m U_m^r \Sigma_m^r (U_m^r)^T.$$

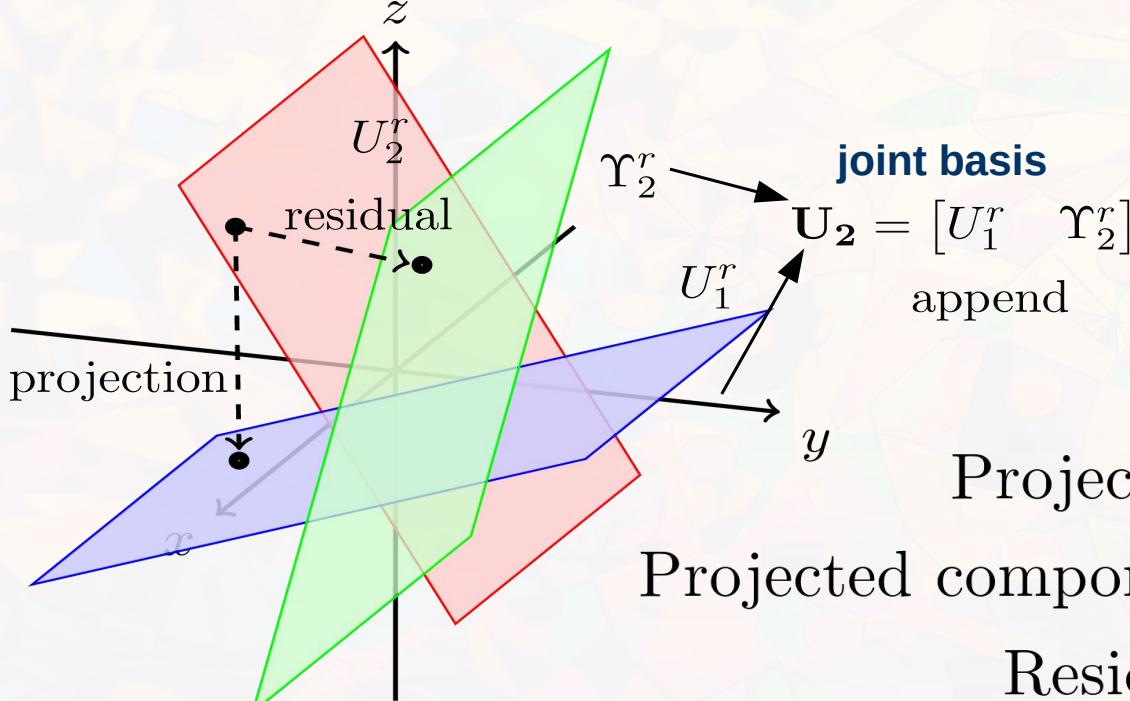
Let eigen decomposition be given by $\mathbf{L}^{r*} = \mathbf{V} \boldsymbol{\Pi} \mathbf{V}^T$

$$\text{span}(\mathbf{V}) = \bigcup_{m=1}^M \text{span}(U_m^r)$$

Construct the basis \mathbf{V} from M basis $U_1^r, U_2^r, \dots, U_M^r$

CONSTRUCTION OF JOINT BASIS

subspaces $U_1^r, U_2^r, \dots, U_M^r$



joint basis
 $\mathbf{U}_2 = [U_1^r \quad \Upsilon_2^r]$
append

Projection:

$$S_2 = \mathbf{U}_1^r{}^T U_2^r$$

Projected component:

$$P_2 = \mathbf{U}_1^r S_2$$

Residual:

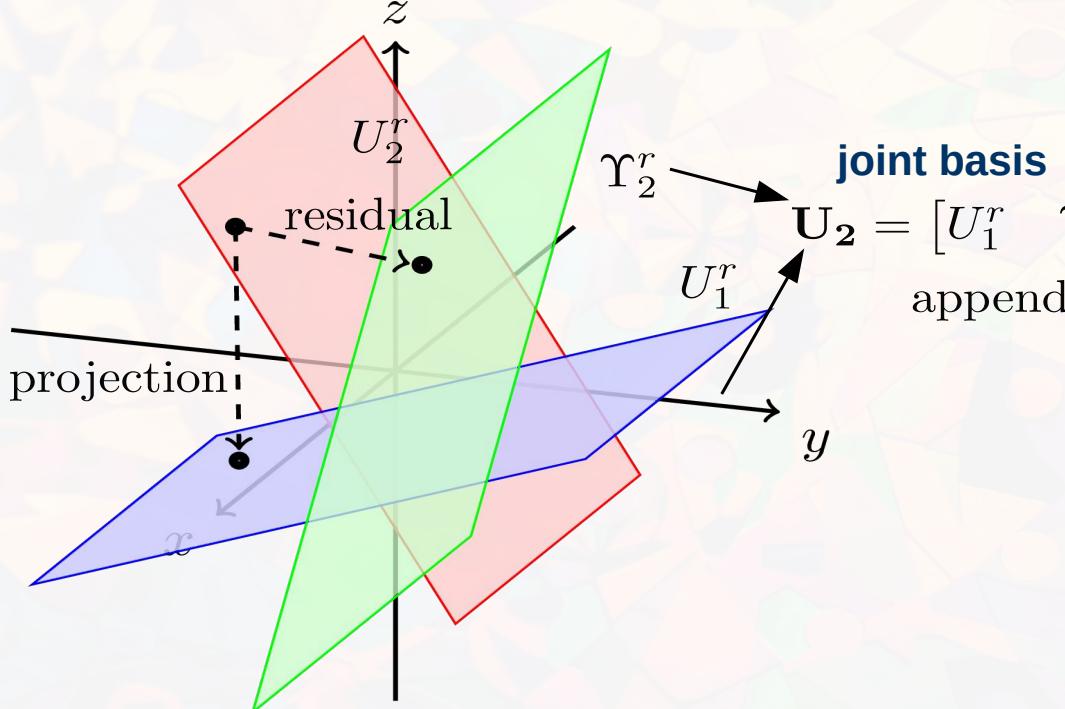
$$Q_2 = U_2^r - P_2$$

Orthogonal residual:

$$\Upsilon_2^r = \text{Gram-Schmidt}(Q_2)$$

CONSTRUCTION OF JOINT BASIS

subspaces $U_1^r, U_2^r, \dots, U_M^r$



joint basis

append

$$\mathbf{U}_M = [\Upsilon_1^r \quad \Upsilon_2^r \quad \dots \quad \Upsilon_M^r]$$

$$\mathbf{L}^{r*} = \mathbf{V} \Pi \mathbf{V}^T = \sum_{m=1}^M \alpha_m U_m^r \Sigma_m^r (U_m^r)^T$$

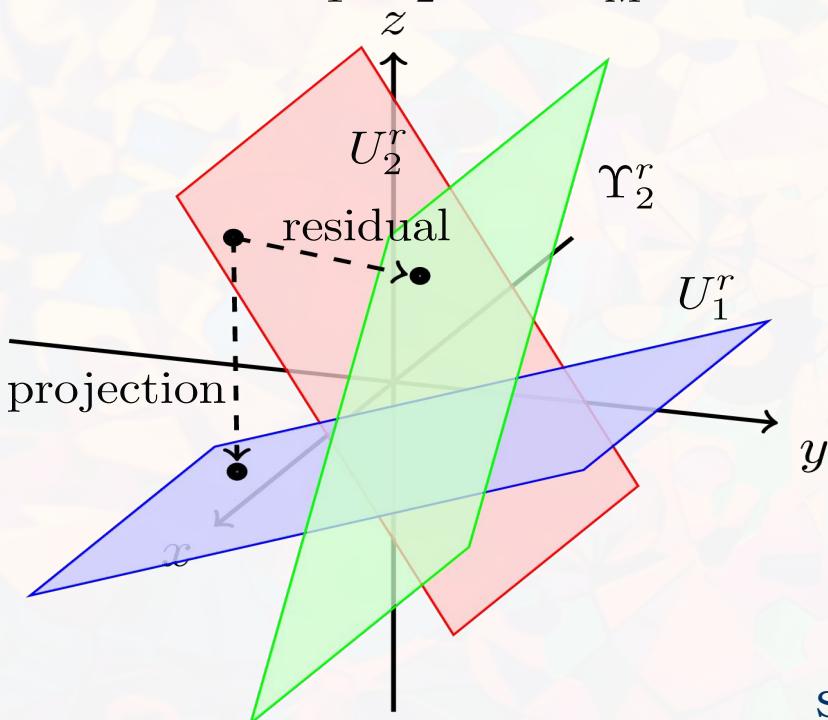
$$\text{span}(\mathbf{V}) = \text{span}(\mathbf{U}_M)$$

$$\mathbf{V} = \mathbf{U}_M \mathbf{R}$$

\mathbf{R} is a $(Mr \times Mr)$ rotation matrix

CONSTRUCTION OF JOINT EIGENSPACE

subspaces $U_1^r, U_2^r, \dots, U_M^r$



$$\mathbf{L}^{r*} = \mathbf{V} \Pi \mathbf{V}^T = \sum_{m=1}^M \alpha_m U_m^r \Sigma_m^r (U_m^r)^T$$

↑
⋮
.....
substitute

$$\mathbf{V} \equiv \mathbf{U}_M \mathbf{R}$$

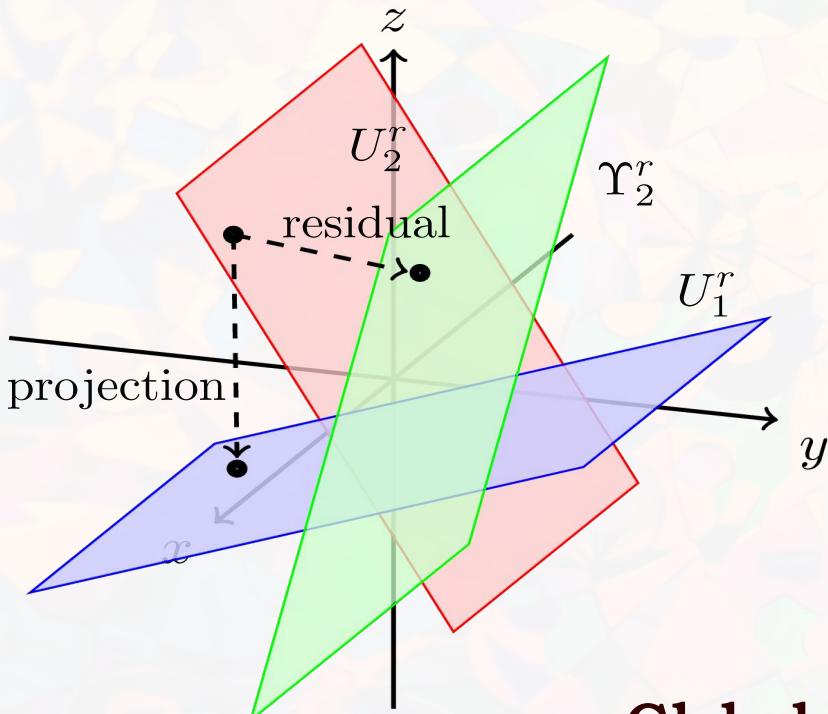
$$\mathbf{R}\Pi\mathbf{R}^T = \sum_{m=1}^M \alpha_m H_m$$

$$H_m(i, j) = \begin{cases} \Upsilon_i^T U_m^r \Sigma_m^r (U_m^r)^T \Upsilon_j & \text{if } i \leq m \text{ and } j \leq m, \\ 0 & \text{if } i > m \text{ or } j > m. \end{cases}$$

Smaller eigenvalue problem of size $(Mr \times Mr)$ where $M, r \ll n$

MULTI-VIEW CLUSTERING

subspaces $U_1^r, U_2^r, \dots, U_M^r$



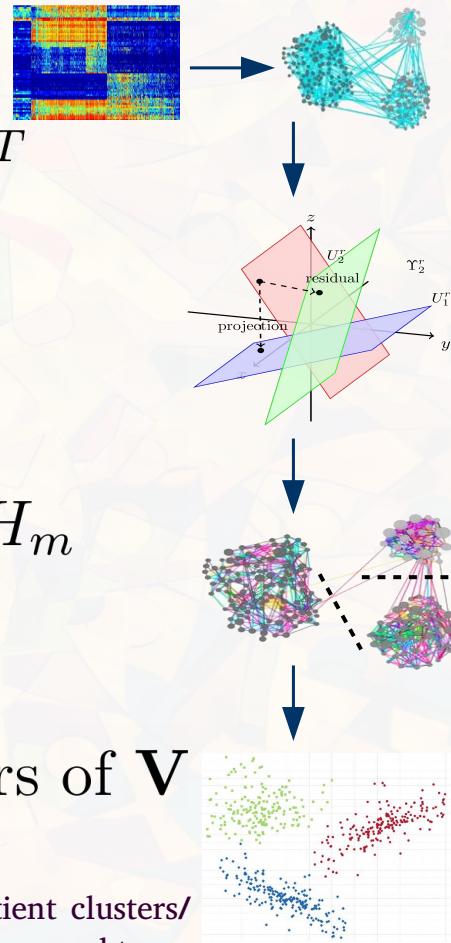
Global cluster
assignment:

$$\begin{aligned}\mathbf{L}^{r*} &= \mathbf{V}\Pi\mathbf{V}^T = \sum_{m=1}^M \alpha_m U_m^r \Sigma_m^r (U_m^r)^T \\ \mathbf{V} &= \mathbf{U}_M \mathbf{R} \\ \mathbf{R}\Pi\mathbf{R}^T &= \sum_{m=1}^M \alpha_m H_m\end{aligned}$$

Perform k -means clustering on
 k largest eigenvectors of \mathbf{V}

PROPOSED ALGORITHM: CoALA

- Construct Laplacian L_1, L_2, \dots, L_M
- Eigendecomposition of Laplacian $L_m = U_m \Sigma_m (U_m)^T$
- $U_m^r \leftarrow r$ largest eigenvectors of in U_m
- Construct joint basis $\mathbf{U}_M = [\Upsilon_1^r \quad \Upsilon_2^r \quad \dots \quad \Upsilon_M^r]$
- Solve smaller eigenvalue problem $\mathbf{R} \Pi \mathbf{R}^T = \sum_{m=1}^M \alpha_m H_m$
- Obtain eigenvectors of \mathbf{L}^{r*} : $\mathbf{V} = \mathbf{U}_M \mathbf{R}$
- Perform k -means clustering on k largest eigenvectors of \mathbf{V}



THEORETICAL BOUNDS ON EIGENSPACE APPROXIMATION

Approximate
Joint Laplacian

$$\mathbf{L}^{r*} = \sum_{m=1}^M \alpha_m L_m^r = \mathbf{V} \boldsymbol{\Pi} \mathbf{V}^T$$

Full-rank
Joint Laplacian

$$\mathbf{L} = \sum_{m=1}^M \alpha_m L_m = \mathbf{Z} \boldsymbol{\Gamma} \mathbf{Z}^T$$

How far does the rank r eigenspace of approximate Laplacian \mathbf{L}^{r*} deviate from that of the full-rank Laplacian \mathbf{L} ?

THEORETICAL BOUNDS ON EIGENSPACE APPROXIMATION

Approximate
Joint Laplacian

$$\mathbf{L}^{r*} = \sum_{m=1}^M \alpha_m L_m^r = \mathbf{V} \Pi \mathbf{V}^T$$

partition

$$\mathbf{V} = [\mathbf{V}^r \quad \mathbf{V}^{r\perp}] \text{ and } \Pi = \begin{bmatrix} \Pi^r & \mathbf{0} \\ \mathbf{0} & \Pi^{r\perp} \end{bmatrix}$$

Full-rank
Joint Laplacian

$$\mathbf{L} = \sum_{m=1}^M \alpha_m L_m = \mathbf{Z} \Gamma \mathbf{Z}^T$$

partition

$$\mathbf{Z} = [\mathbf{Z}^r \quad \mathbf{Z}^{r\perp}] \text{ and } \Gamma = \begin{bmatrix} \Gamma^r & \mathbf{0} \\ \mathbf{0} & \Gamma^{r\perp} \end{bmatrix}$$

How far does the rank r eigenspace of approximate Laplacian \mathbf{L}^{r*} deviate from that of the full-rank Laplacian \mathbf{L} ?

Eigenvectors: difference between $\text{span}(\mathbf{Z}^r)$ and $\text{span}(\mathbf{V}^r)$

in terms of principal angles between these subspaces.

Eigenvalues: mean squared difference between eigenvalues

THEORETICAL BOUNDS ON EIGENSPACE APPROXIMATION

Approximate Joint Laplacian

$$\mathbf{L}^{r*} = \sum_{m=1}^M \alpha_m L_m^r = \mathbf{V} \Pi \mathbf{V}^T$$

partition

$$\mathbf{V} = [\mathbf{V}^r \quad \mathbf{V}^{r\perp}] \text{ and } \Pi = \begin{bmatrix} \Pi^r & \mathbf{0} \\ \mathbf{0} & \Pi^{r\perp} \end{bmatrix}$$

Eigenvectors

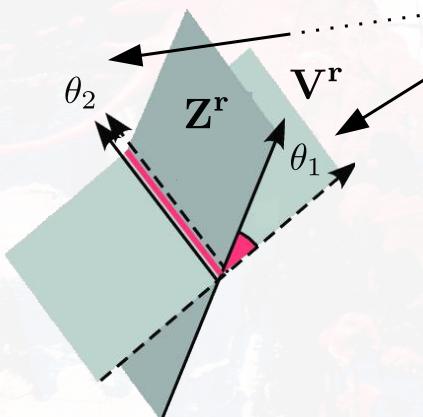
$$\| \sin \Theta (\text{span}(\mathbf{Z}^r), \text{span}(\mathbf{V}^r)) \|_F^2 = \sum_{i=1}^r \sin^2 \theta_i$$

Full-rank Joint Laplacian

$$\mathbf{L} = \sum_{m=1}^M \alpha_m L_m = \mathbf{Z} \Gamma \mathbf{Z}^T$$

$$\mathbf{Z} = [\mathbf{Z}^r \quad \mathbf{Z}^{r\perp}] \text{ and } \Gamma = \begin{bmatrix} \Gamma^r & \mathbf{0} \\ \mathbf{0} & \Gamma^{r\perp} \end{bmatrix}$$

Principal angles
between
subspaces



THEORETICAL BOUNDS ON EIGENSPACE APPROXIMATION

Approximate
Joint Laplacian

$$\mathbf{L}^{r*} = \sum_{m=1}^M \alpha_m L_m^r = \mathbf{V} \boldsymbol{\Pi} \mathbf{V}^T$$

Full-rank
Joint Laplacian

$$\mathbf{L} = \sum_{m=1}^M \alpha_m L_m = \mathbf{Z} \boldsymbol{\Gamma} \mathbf{Z}^T$$

$$\| \sin \Theta(\text{span}(\mathbf{Z}^r), \text{span}(\mathbf{V}^r)) \|_F^2 \leq \frac{\text{trace} \left((\mathbf{V}^r)^T \left(\sum_{m=1}^M \alpha_m L_m^{r\perp} \right)^2 \mathbf{V}^r \right)}{\left(\pi_r - \pi_{r+1} - \sum_{m=1}^M \alpha_m \lambda_{r+1}^m \right)}$$

THEORETICAL BOUNDS ON EIGENSPACE APPROXIMATION

Approximate Joint Laplacian

$$\mathbf{L}^{r*} = \sum_{m=1}^M \alpha_m L_m^r = \mathbf{V} \boldsymbol{\Pi} \mathbf{V}^T$$

$$\| \sin \Theta (\text{span}(\mathbf{Z}^r), \text{span}(\mathbf{V}^r)) \|_F^2 \leq \frac{\text{trace} \left((\mathbf{V}^r)^T \left(\sum_{m=1}^M \alpha_m L_m^{r\perp} \right)^2 \mathbf{V}^r \right)}{\left(\pi_r - \pi_{r+1} - \sum_{m=1}^M \alpha_m \lambda_{r+1}^m \right)}$$

← → **Gap between eigenvalues**

L = $\sum_{m=1}^M \alpha_m L_m = \sum_{m=1}^M \alpha_m (L_m^r + L_m^{r\perp})$

$= \mathbf{L}^{r*} + \sum_{m=1}^M \alpha_m L_m^{r\perp}$

perturbation of \mathbf{L}^{r*} by the factor $\sum_{m=1}^M \alpha_m L_m^{r\perp}$

Full-rank Joint Laplacian

**Matrix perturbation theory
Davis-Kahan theorem**

THEORETICAL BOUNDS ON EIGENSPACE APPROXIMATION

**Approximate
Joint Laplacian**

$$\mathbf{L}^{r*} = \sum_{m=1}^M \alpha_m L_m^r = \mathbf{V} \boldsymbol{\Pi} \mathbf{V}^T$$

**Full-rank
Joint Laplacian**

$$\mathbf{L} = \sum_{m=1}^M \alpha_m L_m = \mathbf{Z} \boldsymbol{\Gamma} \mathbf{Z}^T$$

$$\| \sin \Theta (\text{span}(\mathbf{Z}^r), \text{span}(\mathbf{V}^r)) \|_F^2 \leq \frac{\text{trace} \left((\mathbf{V}^r)^T \left(\sum_{m=1}^M \alpha_m L_m^{r\perp} \right)^2 \mathbf{V}^r \right)}{\left(\pi_r - \pi_{r+1} - \sum_{m=1}^M \alpha_m \lambda_{r+1}^m \right)}$$

Limiting conditions

$$\lim_{r \rightarrow n} \sum_{m=1}^M \alpha_m L_m^{r\perp} = 0$$

$$\lim_{r \rightarrow n} \Phi^r = \frac{1}{r} \| \sin \Theta (\text{span}(\mathbf{Z}^r), \text{span}(\mathbf{V}^r)) \|_F^2 = 0$$

THEORETICAL BOUNDS ON EIGENSPACE APPROXIMATION

Approximate
Joint Laplacian

$$\mathbf{L}^{r*} = \sum_{m=1}^M \alpha_m L_m^r = \mathbf{V} \boldsymbol{\Pi} \mathbf{V}^T$$

Full-rank
Joint Laplacian

$$\mathbf{L} = \sum_{m=1}^M \alpha_m L_m = \mathbf{Z} \boldsymbol{\Gamma} \mathbf{Z}^T$$

$$\| \sin \Theta (\text{span}(\mathbf{Z}^r), \text{span}(\mathbf{V}^r)) \|_F^2 \leq \frac{\text{trace} \left((\mathbf{V}^r)^T \left(\sum_{m=1}^M \alpha_m L_m^{r\perp} \right)^2 \mathbf{V}^r \right)}{\left(\pi_r - \pi_{r+1} - \sum_{m=1}^M \alpha_m \lambda_{r+1}^m \right)}$$

Bound on eigenvalues of \mathbf{L}^r and \mathbf{L}^{r*}

$$\Delta^r = \frac{1}{n} \sum_{j=1}^n (\gamma_j - \pi_j)^2 \leq \sum_{i=r+1}^n \sum_{m=1}^M \alpha_m (\lambda_i^m)^2$$

Convergence of eigenvalues

$$\lim_{r \rightarrow n} \lambda_r^m = 0$$

$$\lim_{r \rightarrow n} \Delta^r = \lim_{r \rightarrow n} \frac{1}{n} \sum_{j=r+1}^n \sum_{m=1}^M \alpha_m (\lambda_j^m)^2 = 0$$

EXPERIMENTAL RESULTS

Cancer Data Sets:

Colorectal Carcinoma (CRC)

[432 samples, 2 clusters]

Breast Carcinoma (BRCA)

[398 samples, 4 clusters]

Lower Grade Dlioma (LGG)

[267 samples, 3 clusters]

Stomach Adenocarcinoma (STAD)

[242 samples, 4 clusters]

CESC, KIDNEY, LUNG, OV

Modalities/ Views:

DNA methylation

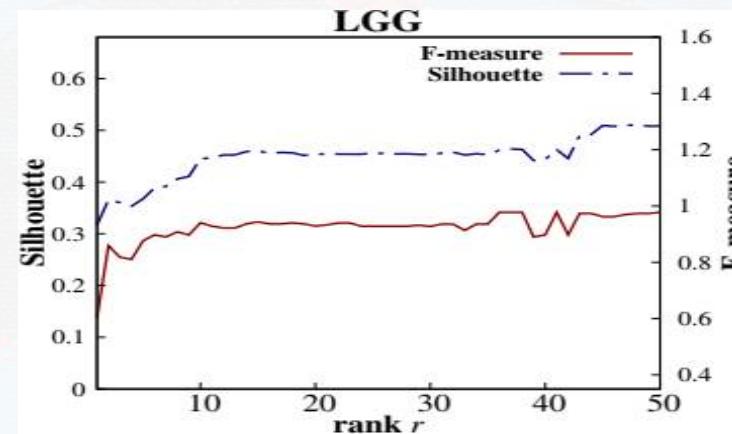
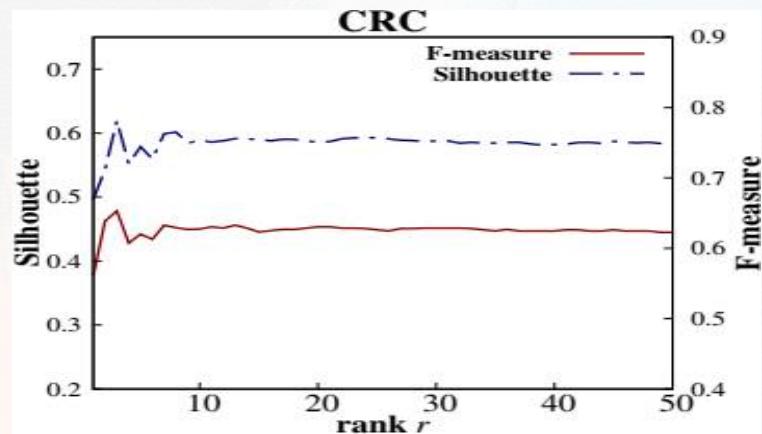
Gene expression

Micro-RNA expression

Protein expression

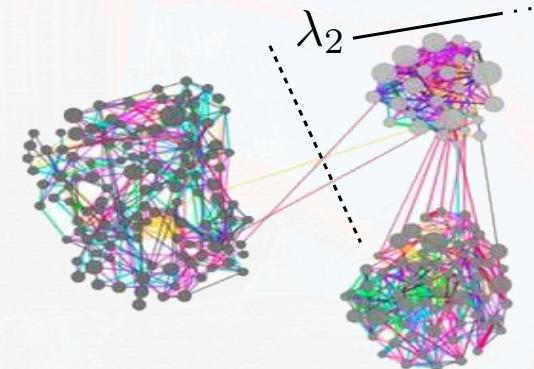
CHOICE OF PARAMETERS

Choice of rank: $r^* = \arg \max_r \{\text{Silhouette}(\mathbf{V}^r)\}.$



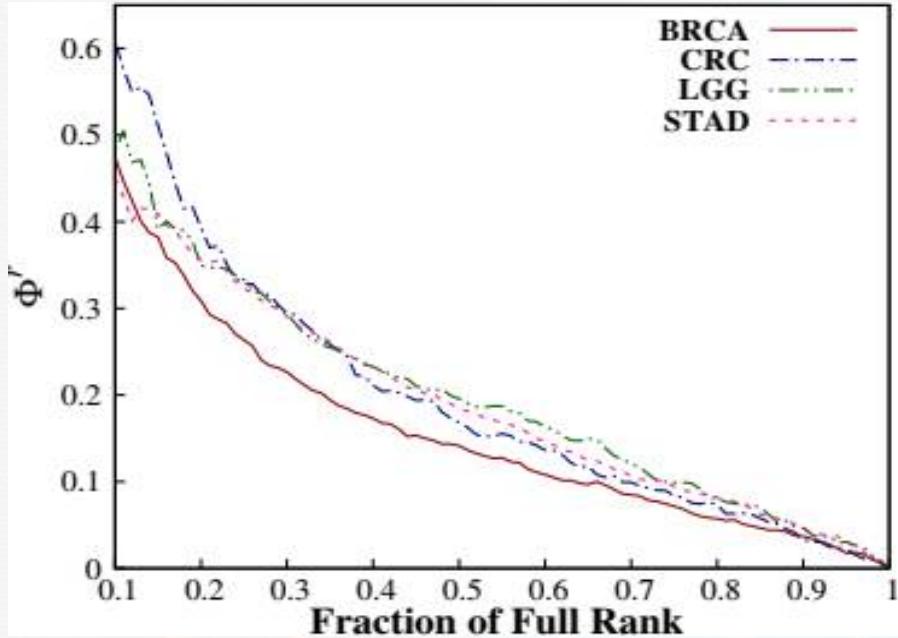
Choice of convex combination:

$$\alpha_m \propto \lambda_2^m \text{ (Fiedler value of } L_m\text{)}$$

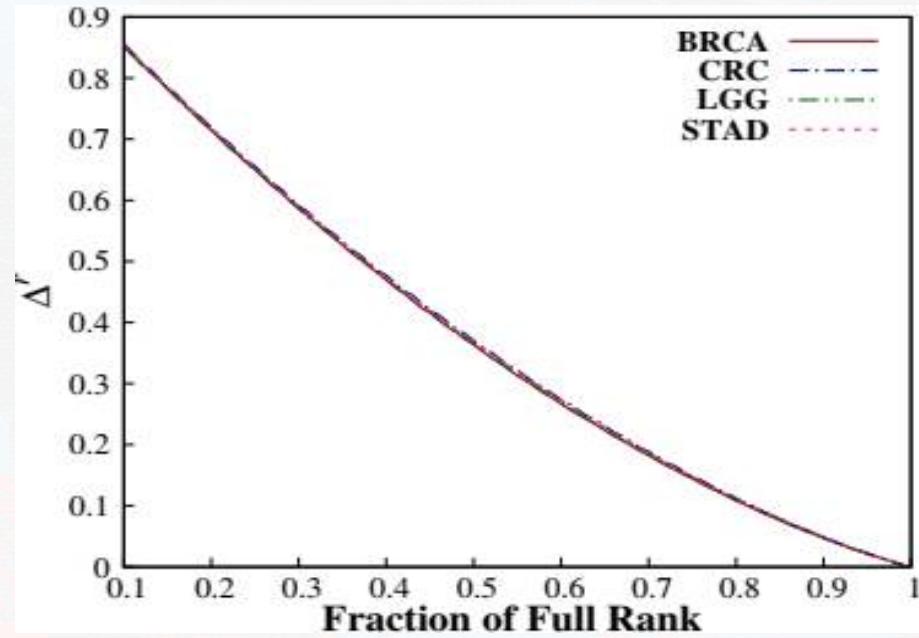


Separability
of graph into
component
subgraphs

EIGENSPACE APPROXIMATION RESULTS



(a) Difference between eigenvectors as approximation rank r increases to full-rank n



(b) Difference between eigenvalues as approximation rank r increases to full-rank n

As the approximation rank approaches to the full rank, the approximate eigenspace of L^{r^*} converges to the full-rank eigenspace of L , in terms of both eigenvectors and eigenvalues.

IMPORTANCE OF MULTI-VIEW INTEGRATION

COMPARATIVE PERFORMANCE OF INDIVIDUAL VIEWS AND PROPOSED COALA ALGORITHM

Views →		mDNA	RNA	miRNA	RPPA	CoALA
F-measure	CRC	0.5849894	0.5397796	0.5673758	0.5741394	0.6529565
Purity		0.7370690	0.7370690	0.7370690	0.7370690	0.7370690
Rand		0.4989573	0.4991528	0.5022809	0.5007448	0.5382531
Jaccard		0.3925508	0.3789509	0.3818306	0.3853947	0.4315561
Dice		0.5637867	0.5496220	0.5526446	0.5563681	0.6029189
F-measure	LGG	0.8269248	0.5875701	0.4717221	0.4326018	0.9737835
Purity		0.8352060	0.5917603	0.5318352	0.5280899	0.9737828
Rand		0.7861508	0.6149925	0.5593760	0.5476050	0.9622089
Jaccard		0.5814133	0.3235367	0.2476680	0.2328447	0.9056723
Dice		0.7353085	0.4888972	0.3970095	0.3777356	0.9505016

Integration of multiple views preserves better cluster structure compared to single view analysis.

IMPORTANCE OF MULTI-VIEW INTEGRATION

COMPARATIVE PERFORMANCE OF INDIVIDUAL VIEWS AND PROPOSED COALA ALGORITHM

Views →		mDNA	RNA	miRNA	RPPA	CoALA
F-measure	STAD	0.5469686	0.4781377	0.3998266	0.4469459	0.7778227
Purity		0.5867769	0.5495868	0.4917355	0.4917355	0.7685950
Rand		0.6509722	0.6239155	0.5989164	0.5883543	0.7661946
Jaccard		0.2869053	0.2234653	0.1994524	0.2076045	0.4535983
Dice		0.4458841	0.3652989	0.3325725	0.3438286	0.6241041
F-measure	BRCA	0.5982526	0.7690661	0.5105008	0.5630781	0.7660191
Purity		0.6532663	0.7688442	0.5703518	0.5879397	0.7613065
Rand		0.7193018	0.7995519	0.6455071	0.6689493	0.7922357
Jaccard		0.3318872	0.4857607	0.2672039	0.3132549	0.4612885
Dice		0.4983713	0.6538882	0.4217221	0.4770664	0.6313449

Integration of multiple views preserves better cluster structure compared to single view analysis.

IMPORTANCE OF LAPLACIAN APPROXIMATION

COMPARATIVE PERFORMANCE OF FULL-RANK AND APPROXIMATE SUBSPACES

Index	Full-Rank (L)	Approximate (L^{r*})	Full-Rank (L)	Approximate (L^{r*})
F-measure	0.6052757	0.6529565	CRC	0.6577440
Purity	0.7370690	0.7370690		0.6441948
Rand	0.5007448	0.5382531		0.6524739
Jaccard	0.4018471	0.4315561		0.4053390
Dice	0.5733108	0.6029189		0.5768558
F-measure	0.6158419	0.7778227	STAD	0.6197007
Purity	0.6157025	0.768595		0.7185930
Rand	0.6706560	0.7661946		0.7403390
Jaccard	0.2966164	0.4535983		0.3586770
Dice	0.4575237	0.6241041		0.5279798

Integrating the de-noised cluster information of only r largest eigenpairs of each Laplacian preserves significantly better cluster structure.

IMPORTANCE OF CONVEX COMBINATION

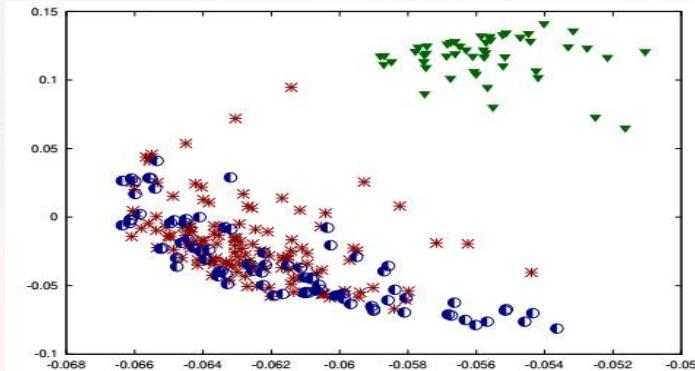
COMPARATIVE PERFORMANCE OF EQUALLY-WEIGHTED AND EIGEN-WEIGHTED COMBINATION OF VIEWS

Index		L^{r^*} -Equal	L^{r^*} -Eigen		L^{r^*} -Equal	L^{r^*} -Eigen
F-measure	CRC	0.6309431	0.6529565	LGG	0.9625844	0.9737835
Purity		0.7370690	0.7370690		0.9625468	0.9737828
Rand		0.5260669	0.5382531		0.9437921	0.9622089
Jaccard		0.4194417	0.4315561		0.8619640	0.9056723
Dice		0.5909953	0.6029189		0.9258654	0.9505016
F-measure	STAD	0.7788198	0.7778227	BRCA	0.6834253	0.7660191
Purity		0.7727273	0.7685950		0.6783920	0.7613065
Rand		0.7703782	0.7661946		0.7523132	0.7922357
Jaccard		0.4579454	0.4535983		0.3986848	0.4612885
Dice		0.6282066	0.6241041		0.5700852	0.6313449

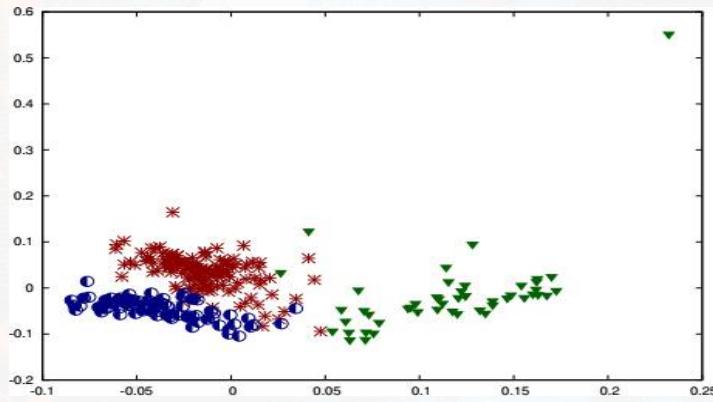
Weighting the views proportional to the Fidler value or second-largest eigenvalue of each Lapacian gives better cluster information in majority of the cases..

SCATTER PLOT ANALYSIS: LGG DATA SET

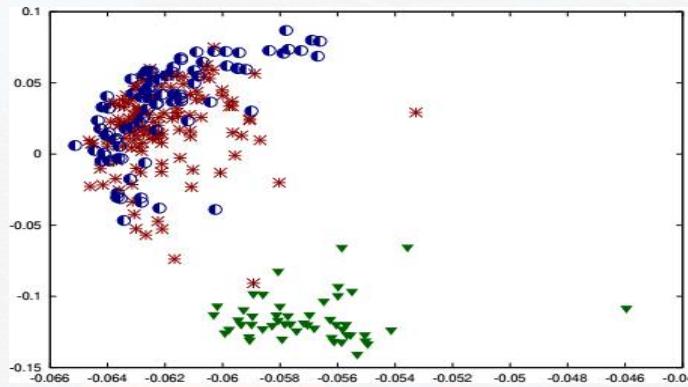
(a) Best view:
mDNA



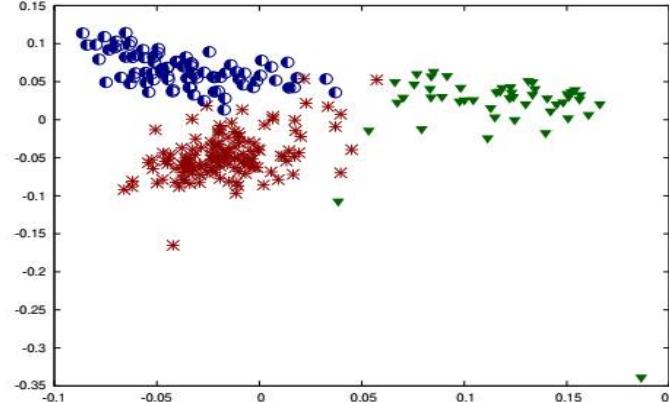
(b) Equally
weighted
combination



(c) Full-Rank



(d) Proposed:
CoALa



Proposed approach CoALa has the most compact and well-separated clusters.

COMPARATIVE PERFORMANCE ANALYSIS

	Different Algorithms	Rank of Subspace	External Evaluation Index					
			Accuracy	NMI	ARI	F-Measure	RAND	Purity
CRC	COCA	-	0.5323276	0.0120929	0.0007663	0.5586055	0.5010706	0.7370690
	BCC	-	0.5745690	0.0070894	0.0074889	0.5973067	0.5158300	0.7370690
	JIVE(Perm)	16	0.6034483	0.0071359	0.0256478	0.6210774	0.5203694	0.7370690
	A-JIVE	32	0.6034483	0.0064720	0.0246106	0.6206032	0.5203694	0.7370690
	iCluster	1	0.6163793	0.0069992	0.0293081	0.6298050	0.5260669	0.7370690
	LRACLuster	1	0.5129310	0.0030437	-0.001822	0.5410661	0.4992552	0.7370690
	PCA-Con	2	0.5366379	0.0057828	0.0036971	0.5641984	0.5016106	0.7370690
	SNF	2	0.5991379	0.0069730	0.0240692	0.6178576	0.5186192	0.7370690
	NormS	16	0.6206897	0.0093881	0.0347351	0.6345375	0.5281150	0.7370690
	SURE	2	0.5107759	0.0027977	-0.002148	0.5416716	0.4991528	0.7370690
LGG	CoALA	2	0.6400862	0.0185660	0.0548748	0.6529565	0.5382531	0.7370690
	COCA	-	0.6591760	0.2772248	0.2533847	0.6608123	0.6454901	0.6591760
	BCC	-	0.6340824	0.2737596	0.248606	0.63111660	0.6382755	0.6355805
	JIVE	8	0.5617978	0.2299551	0.1606599	0.5757978	0.6056715	0.5730337
	A-JIVE	48	0.7168539	0.4267241	0.3376560	0.7172792	0.6869055	0.7168539
	iCluster	2	0.4382022	0.1379678	0.0996867	0.5187438	0.5821858	0.5355805
	LRACLuster	2	0.4719101	0.1240057	0.1030798	0.5137382	0.5831714	0.5280899
	PCA-con	3	0.6666667	0.3438738	0.3031312	0.6574834	0.6616823	0.6666667
	SNF	3	0.8689139	0.6253254	0.6331662	0.8720595	0.8268142	0.8689139
	NormS	14	0.7940075	0.5325030	0.4649223	0.7916535	0.7465292	0.7940075
	SURE	3	0.7940075	0.5335888	0.4668931	0.7904750	0.7465292	0.7940075
	CoALA	4	0.9737828	0.8689965	0.9199392	0.9737835	0.9622089	0.9737828

Clusters identified by the proposed CoALA algorithm have the closest resemblance with the clinically established subtypes of colorectal carcinoma (CRC) and lower-grade glioma (LGG).

COMPARATIVE PERFORMANCE ANALYSIS

	Different Algorithms	Rank of Subspace	External Evaluation Index					
			Accuracy	NMI	ARI	F-Measure	RAND	Purity
STAD	COCA	-	0.4450413	0.1309746	0.0740987	0.4558087	0.5981242	0.5173554
	BCC	-	0.5392562	0.1500351	0.1421471	0.5520075	0.6081204	0.5673554
	JIVE(Perm)	8	0.4049587	0.1288122	0.0657955	0.4487487	0.5981619	0.5165289
	A-JIVE	64	0.4148760	0.1234864	0.0763413	0.4458621	0.6086142	0.5227273
	iCluster	3	0.3512397	0.0650589	0.0288255	0.3832114	0.5855423	0.4917355
	LRAcluster	1	0.4256198	0.1259879	0.0912460	0.4746753	0.6122218	0.5619835
	PCA-Con	2	0.6900826	0.3654109	0.3204142	0.6959782	0.7110524	0.6900826
	SNF	2	0.5661157	0.3216270	0.2694201	0.6333622	0.6945235	0.6363636
	NormS	27	0.5702479	0.1805281	0.1625013	0.5770884	0.6435993	0.5950413
	SURE	2	0.6983471	0.3511439	0.3445607	0.7056674	0.7216145	0.6983471
BRCA	CoALa	4	0.7685950	0.5107260	0.4559866	0.7778227	0.7661946	0.768595
	COCA	-	0.7434673	0.5002408	0.4864778	0.7457304	0.7905295	0.7434673
	BCC	-	0.6251256	0.3169187	0.3049874	0.6242493	0.7055783	0.6334171
	JIVE	12	0.6859296	0.4287142	0.3772649	0.6889363	0.7464906	0.6859296
	A-JIVE	64	0.6140704	0.4482479	0.3710317	0.6707575	0.7363682	0.6841709
	iCluster	3	0.7638191	0.5176193	0.4745746	0.7658865	0.7842867	0.7638191
	LRAcluster	2	0.7110553	0.4368520	0.4035040	0.7101385	0.7521740	0.7110553
	PCA-con	4	0.7587940	0.5506612	0.5038795	0.7601317	0.7984380	0.7587940
	SNF	4	0.6783920	0.4558955	0.4111794	0.6865447	0.7602370	0.6959799
	NormS	11	0.7688442	0.5437267	0.5090183	0.7699789	0.7999063	0.7688442
	SURE	4	0.7663317	0.5528011	0.5104814	0.7683344	0.8010455	0.7663317
	CoALa	4	0.7613065	0.5281849	0.4874579	0.7660191	0.7922357	0.7613065

CoALa algorithm has best performance for STAD data set, for BRCA the performance is comparable.

RESULTS ON BENCHMARK DATASETS

COMPARATIVE PERFORMANCE OF ANALYSIS OF PROPOSED AND EXISTING APPROACHES

Algorithms →		Best View	Full-Rank	SNF	CoALA		Best View	Full-Rank	SNF	CoALA
Subspace Rank	FOOTBALL	20	20	20	20	POLITICS-UK	5	5	5	5
F-measure		0.7747023	0.6616297	0.8431825	0.8683491		0.9175316	0.8192186	0.9701235	0.9736129
Purity		0.7282258	0.6572580	0.8266129	0.8584677		0.9713604	0.8591885	0.9761337	0.9785203
Rand		0.9472965	0.8843737	0.9735862	0.9739682		0.9196880	0.8603076	0.9814665	0.9826084
Jaccard		0.3963918	0.2612106	0.6125478	0.6005824		0.8019766	0.7257236	0.9529074	0.9559279
Dice		0.5667814	0.4136485	0.7597267	0.7504383		0.8901077	0.8410659	0.9758859	0.9774674
Subspace Rank	RUGBY	15	15	15	15	DIGITS	17	17	17	17
F-measure		0.7426962	0.6845209	0.7778990	0.8349647		0.7209662	0.8481826	0.8932872	0.8839913
Purity		0.7796253	0.6803279	0.8454333	0.8606557		0.7100000	0.8500000	0.8835000	0.8835000
Rand		0.8672685	0.8578210	0.8818113	0.9067597		0.9173923	0.9503602	0.9715983	0.9576618
Jaccard		0.4447761	0.4883211	0.4446208	0.5982183		0.4163257	0.6055477	0.7534116	0.6502019
Dice		0.6155136	0.6562039	0.6155536	0.7486065		0.5878948	0.7543192	0.8593665	0.7880271

Football : Twitter data set, 248 samples, 20 clusters, 9 views

Politics-UK : Twitter data set, 419 samples, 5 clusters, 9 views

Rugby : Twitter data set, 854 samples, 15 clusters, 9 views

Digits : Image data set, 2000 samples, 10 clusters, 6 views

RESULTS ON BENCHMARK DATASETS

COMPARATIVE PERFORMANCE OF ANALYSIS OF PROPOSED AND EXISTING APPROACHES

Algorithms →		Best View	Full-Rank	SNF	CoALa		Best View	Full-Rank	SNF	CoALa
Subspace Rank	FOOTBALL	20	20	20	20	POLITICS-UK	5	5	5	5
F-measure		0.7747023	0.6616297	0.8431825	0.8683491		0.9175316	0.8192186	0.9701235	0.9736129
Purity		0.7282258	0.6572580	0.8266129	0.8584677		0.9713604	0.8591885	0.9761337	0.9785203
Rand		0.9472965	0.8843737	0.9735862	0.9739682		0.9196880	0.8603076	0.9814665	0.9826084
Jaccard		0.3963918	0.2612106	0.6125478	0.6005824		0.8019766	0.7257236	0.9529074	0.9559279
Dice		0.5667814	0.4136485	0.7597267	0.7504383		0.8901077	0.8410659	0.9758859	0.9774674
Subspace Rank	RUGBY	15	15	15	15	DIGITS	17	17	17	17
F-measure		0.7426962	0.6845209	0.7778990	0.8349647		0.7209662	0.8481826	0.8932872	0.8839913
Purity		0.7796253	0.6803279	0.8454333	0.8606557		0.7100000	0.8500000	0.8835000	0.8835000
Rand		0.8672685	0.8578210	0.8818113	0.9067597		0.9173923	0.9503602	0.9715983	0.9576618
Jaccard		0.4447761	0.4883211	0.4446208	0.5982183		0.4163257	0.6055477	0.7534116	0.6502019
Dice		0.6155136	0.6562039	0.6155536	0.7486065		0.5878948	0.7543192	0.8593665	0.7880271

Proposed algorithm has best performance compared to best individual view, full-rank integration, and graph based SNF approach on benchmark data sets from varying application domains like social networking, image processing.

Highlights

- Integrates multiple similarity graphs
- Prevents noise of individual graphs from being propagated
- Theoretical bound on the difference between the full-rank and approximate eigenspaces
- Clustering in approximate subspace preserves better cluster structure compared to individual subspaces and also the full-rank subspace
- Proposed algorithm significantly and consistently outperforms state-of-the-art integrative clustering approaches on multi-view data sets from various application domains

CHAPTER 6

Multi-Manifold Optimization for Multi-View Subspace Clustering

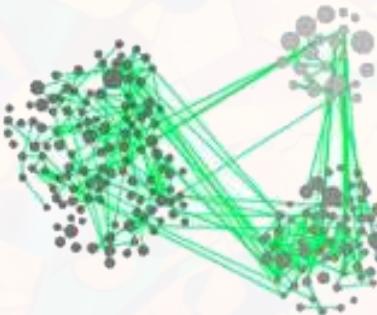
A. Khan and P. Maji, “Multi-Manifold Optimization for Multi-View Subspace Clustering,” **IEEE Transactions on Neural Networks and Learning Systems (TNNLS)**, pp. 1-13, 2021. DOI:10.1109/TNNLS.2021.3054789.

Manifold Based Clustering

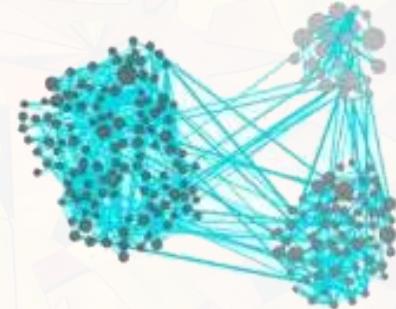
$\mathbf{X}_1 \rightarrow \mathbf{G}_1$



$\mathbf{X}_2 \rightarrow \mathbf{G}_2$



$\mathbf{X}_M \rightarrow \mathbf{G}_M$



Separate Similarity Graphs



Data heterogeneity

How to capture the low-rank non-linear geometry of views?

Manifold Based Clustering

Equivalant formulation of the k-means clustering objective [2]

$$\min_{U \in \Re^{n \times k}} -\text{tr}(U^T L U) + \xi \|U_{-}\|_F^2$$

such that $U^T U = \mathbf{I}_k; \quad U U^T \mathbf{1} = \mathbf{1}$

Manifold Based Clustering

Equivalent formulation of the k-means clustering objective [2]

$$\min_{U \in \mathbb{R}^{n \times k}} -\text{tr}(U^T L U) + \xi \|U_{-}\|_F^2$$

such that

$$U^T U = \mathbf{I}_k; \quad U U^T \mathbf{1} = \mathbf{1}$$

$$\text{St}(n, r) := \{U \in \mathbb{R}^{n \times r} : U^T U = \mathbf{I}_r\}$$

Stiefel manifold

$$\text{Km}(n, r) := \{U \in \mathbb{R}^{n \times r} : U^T U = \mathbf{I}_r, U U^T \mathbf{1} = \mathbf{1}\}$$

k-means manifold

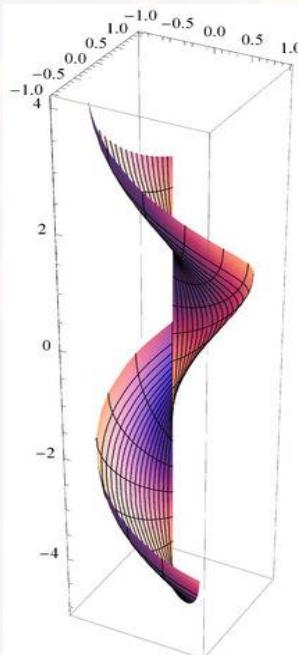
Constraint space forms different Riemannian
submanifolds of the Euclidean space!

Manifold Based Clustering

Equivalent formulation of the k-means clustering objective [2]

$$\min_{U \in \Re^{n \times k}} -\text{tr}(U^T L U) + \xi \|U_{-}\|_F^2$$

such that $U^T U = \mathbf{I}_k; \quad U U^T \mathbf{1} = \mathbf{1}$



Manifold: A topological space that locally resembles
the Euclidean space

Manifold Based Clustering

Equivalent formulation of the k-means clustering objective [2]

$$\min_{U \in \Re^{n \times k}} -\text{tr}(U^T L U) + \xi \|U_{-}\|_F^2$$

such that $U^T U = \mathbf{I}_k; U U^T \mathbf{1} = \mathbf{1}$

Constrained
optimization in
Euclidean space

$$\min_{U \in \mathbb{Km}(n,k)} -\text{tr}(U^T L U) + \xi \|U_{-}\|_F^2$$

Un-constrained
optimization over
Riemannian manifolds

Multi-View Integration

Approximate Joint Laplacian:

$$\mathbf{L}_{\text{Joint}}^r = \sum_{m=1}^M \alpha_m \mathbf{L}_m^r, \text{ such that } \alpha_m \geq 0 \text{ and } \sum_{m=1}^M \alpha_m = 1.$$

Joint clustering objective $\longrightarrow \dots \rightarrow$ Global/Consensus clustering

$$\min_{U_{\text{Joint}} \in \Re^{n \times r}} -\frac{1}{2} \text{tr}(\mathbf{U}_{\text{Joint}}^T \mathbf{L}_{\text{Joint}}^r \mathbf{U}_{\text{Joint}}) + \frac{\xi}{2} \| \mathbf{U}_{\text{Joint}} - \mathbf{U}_{\text{Joint-}} \|_F^2$$

such that $\mathbf{U}_{\text{Joint}}^T \mathbf{U}_{\text{Joint}} = \mathbf{I}_r; \mathbf{U}_{\text{Joint}} \mathbf{U}_{\text{Joint}}^T \mathbf{1} = \mathbf{1}$

Multi-View Integration

Joint clustering objective

$$\min_{U_{\text{Joint}} \in \Re^{n \times r}} -\frac{1}{2} \text{tr}(U_{\text{Joint}}^T \mathbf{L}_{\text{Joint}}^r U_{\text{Joint}}) + \frac{\xi}{2} \|U_{\text{Joint}}\|_F^2$$

such that $U_{\text{Joint}}^T U_{\text{Joint}} = \mathbf{I}_r; U_{\text{Joint}} U_{\text{Joint}}^T \mathbf{1} = \mathbf{1}$

Disagreement between views

$$\begin{aligned} \mathfrak{D}(U_{\text{Joint}}, U_j) &= \|U_{\text{Joint}} U_{\text{Joint}}^T - U_j U_j^T\|_F^2 \\ &= 2r - 2 \text{tr}(U_{\text{Joint}} U_{\text{Joint}}^T U_j U_j^T) \end{aligned}$$

Multi-View Integration

Joint objective: Clustering + Disagreement minimization

$$\min_{U_{\text{Joint}}, U_j \in \Re^{n \times r}} -\frac{1}{2r} \text{tr} \left(U_{\text{Joint}}^T \mathbf{L}_{\text{Joint}}^r U_{\text{Joint}} \right) + \frac{\xi}{2r} \| U_{\text{Joint}} - \|_F^2$$

$$-\frac{1}{2rM} \sum_{j=1}^M \left[\text{tr} \left(U_{\text{Joint}} U_{\text{Joint}}^T U_j U_j^T \right) + \text{tr} \left(U_j^T L_j^r U_j \right) \right]$$

such that $U_{\text{Joint}}^T U_{\text{Joint}} = \mathbf{I}_r, U_{\text{Joint}} U_{\text{Joint}}^T \mathbf{1} = \mathbf{1}, U_j^T U_j = \mathbf{I}_r$

$$\begin{array}{c} \xleftarrow{\hspace{1cm}} \text{k-means manifold} \xrightarrow{\hspace{1cm}} \\ U_{\text{Joint}} \in \text{Km}(n, r) \end{array} \quad \begin{array}{c} \xleftarrow{\hspace{1cm}} \text{Stiefel manifold} \xrightarrow{\hspace{1cm}} \\ U_j \in \text{St}(n, r) \end{array}$$

Manifold Optimization

Stiefel and k-means manifolds:
Non-linear manifolds, not necessarily vector spaces

Gradient descent: $y^{(t+1)} = y^{(t)} \boxed{-} \eta \nabla f(y^{(t)})$

Line-search optimization

Manifold Optimization

Initial iterates: $U_{\text{Joint}}^{(0)}, U_j^{(0)}$ for $j = 1, \dots, M$ Iteration t : $U_{\text{Joint}}^{(t)}, U_j^{(t)}$

Optimization of U_{Joint}

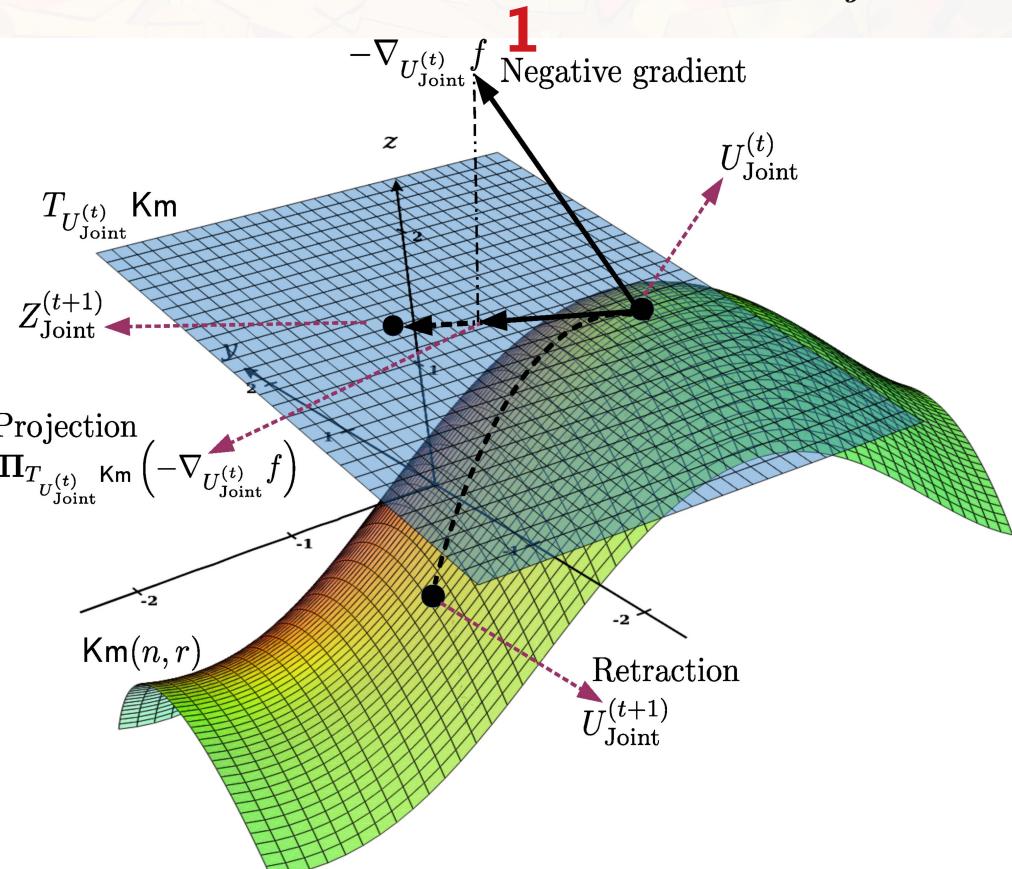
Manifold Optimization

Initial iterates: $U_{\text{Joint}}^{(0)}, U_j^{(0)}$ for $j = 1, \dots, M$

Optimization of U_{Joint}

$$1. -\nabla_{U_{\text{Joint}}^{(t)}} f = \left(\mathbf{L}_{\text{Joint}}^r + \sum_{j=1}^M U_j U_j^T \right) U_{\text{Joint}}^{(t)}$$

Iteration t : $U_{\text{Joint}}^{(t)}, U_j^{(t)}$



Manifold Optimization

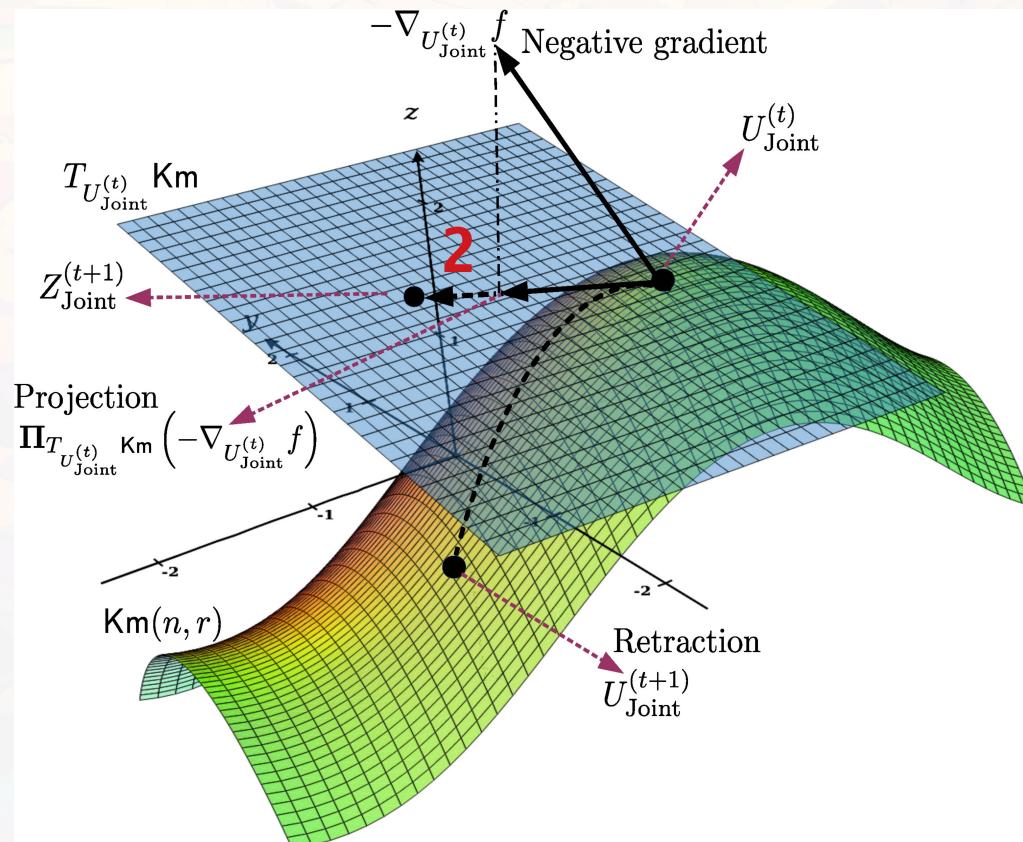
Initial iterates: $U_{\text{Joint}}^{(0)}, U_j^{(0)}$ for $j = 1, \dots, M$

Optimization of U_{Joint}

$$1. -\nabla_{U_{\text{Joint}}^{(t)}} f = \left(\mathbf{L}_{\text{Joint}}^r + \sum_{j=1}^M U_j U_j^T \right) U_{\text{Joint}}^{(t)}$$

$$\begin{aligned} 2. \Pi_{T_{U_{\text{Joint}}^{(t)}} \mathcal{K}_m} & \left(-\nabla_{U_{\text{Joint}}^{(t)}} f \right) \\ &= [-\nabla_{U_{\text{Joint}}^{(t)}} f] - 2U_{\text{Joint}}^{(t)} \Omega - (\mathbf{z}\mathbf{1}^T + \mathbf{1}\mathbf{z}^T) U_{\text{Joint}}^{(t)} = \mathbf{Z}_{\text{Joint}}^{(t)} \end{aligned}$$

Iteration t : $U_{\text{Joint}}^{(t)}, U_j^{(t)}$



Manifold Optimization

Initial iterates: $U_{\text{Joint}}^{(0)}, U_j^{(0)}$ for $j = 1, \dots, M$

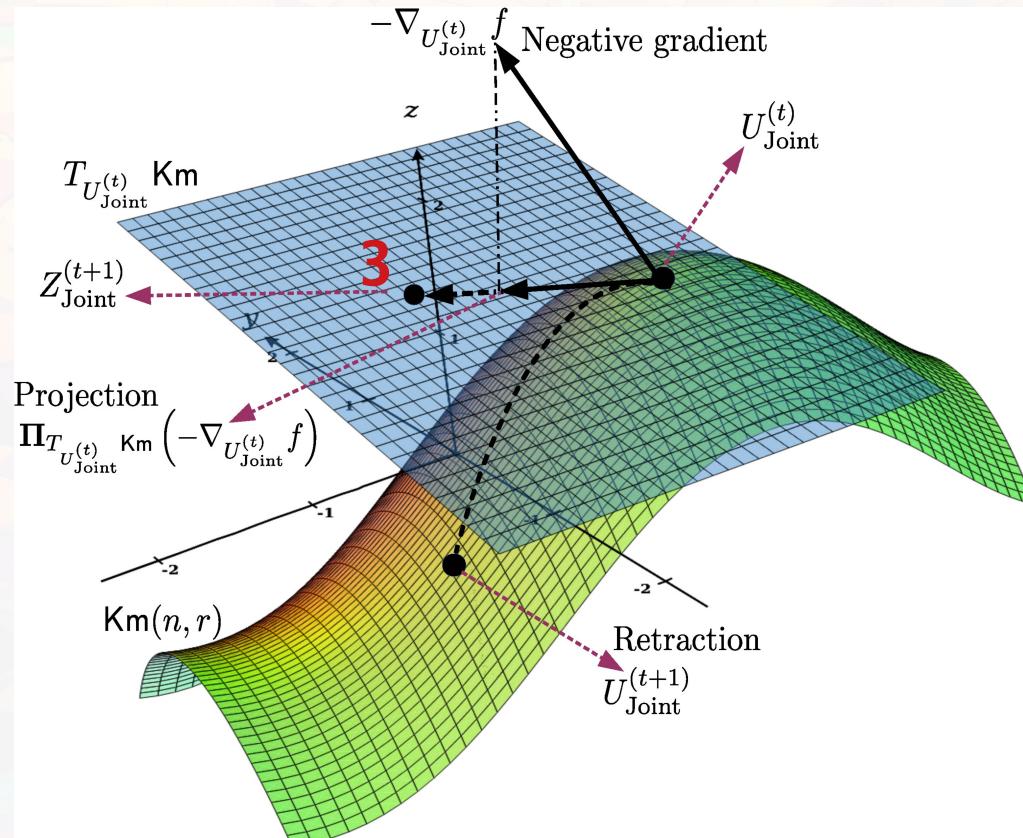
Optimization of U_{Joint}

$$1. -\nabla_{U_{\text{Joint}}^{(t)}} f = \left(\mathbf{L}_{\text{Joint}}^r + \sum_{j=1}^M U_j U_j^T \right) U_{\text{Joint}}^{(t)}$$

$$\begin{aligned} 2. \Pi_{T_{U_{\text{Joint}}^{(t)}} \mathcal{K}_m} (-\nabla_{U_{\text{Joint}}^{(t)}} f) \\ = [-\nabla_{U_{\text{Joint}}^{(t)}} f] - 2U_{\text{Joint}}^{(t)} \Omega - (\mathbf{z}\mathbf{1}^T + \mathbf{1}\mathbf{z}^T)U_{\text{Joint}}^{(t)} = \mathbf{Z}_{\text{Joint}}^{(t)} \end{aligned}$$

$$3. \mathbf{Z}_{\text{Joint}}^{(t+1)} = U_{\text{Joint}}^{(t)} + \eta_{\mathcal{K}} \mathbf{Z}_{\text{Joint}}^{(t)}$$

Iteration t : $U_{\text{Joint}}^{(t)}, U_j^{(t)}$



Manifold Optimization

Initial iterates: $U_{\text{Joint}}^{(0)}, U_j^{(0)}$ for $j = 1, \dots, M$

Optimization of U_{Joint}

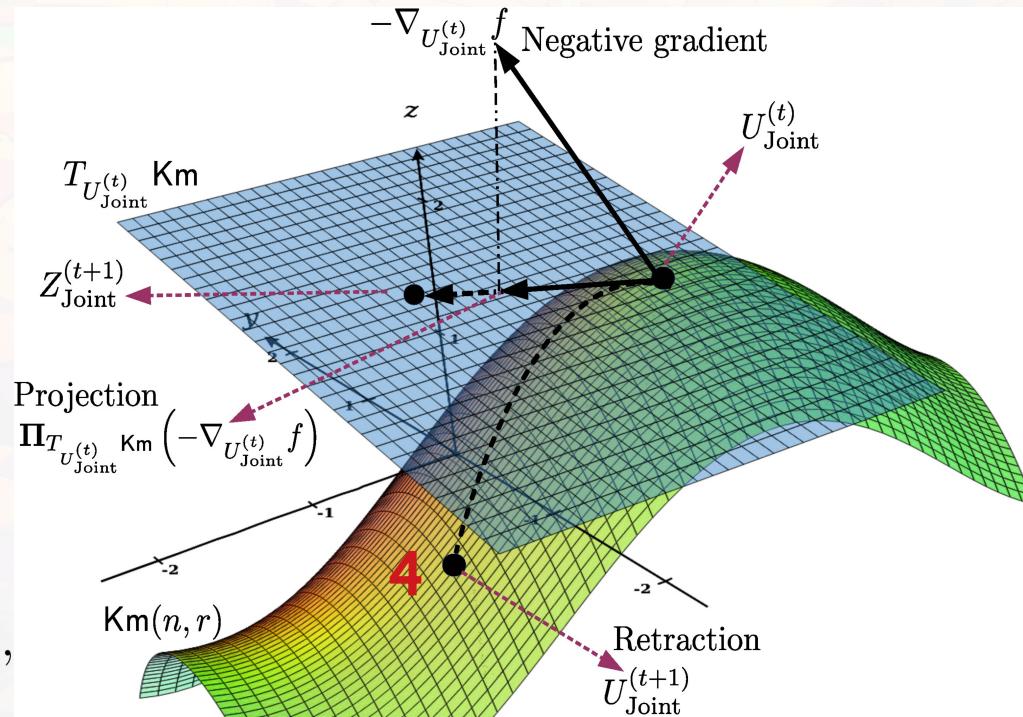
$$1. -\nabla_{U_{\text{Joint}}^{(t)}} f = \left(\mathbf{L}_{\text{Joint}}^r + \sum_{j=1}^M U_j U_j^T \right) U_{\text{Joint}}^{(t)}$$

$$\begin{aligned} 2. \Pi_{T_{U_{\text{Joint}}^{(t)}} \mathcal{K}_m} (-\nabla_{U_{\text{Joint}}^{(t)}} f) \\ = [-\nabla_{U_{\text{Joint}}^{(t)}} f] - 2U_{\text{Joint}}^{(t)} \Omega - (\mathbf{z}\mathbf{1}^T + \mathbf{1}\mathbf{z}^T)U_{\text{Joint}}^{(t)} = \mathbf{Z}_{\text{Joint}}^{(t)} \end{aligned}$$

$$3. \mathbf{Z}_{\text{Joint}}^{(t+1)} = U_{\text{Joint}}^{(t)} + \eta_{\mathcal{K}} \mathbf{Z}_{\text{Joint}}^{(t)}$$

$$4. \mathbf{P} \mathbf{K}_m_{U_{\text{Joint}}^{(t)}} (\mathbf{Z}_{\text{Joint}}^{(t+1)}) = \exp(B) \exp(Q') U_{\text{Joint}}^{(t)},$$

Iteration t : $U_{\text{Joint}}^{(t)}, U_j^{(t)}$



Manifold Optimization

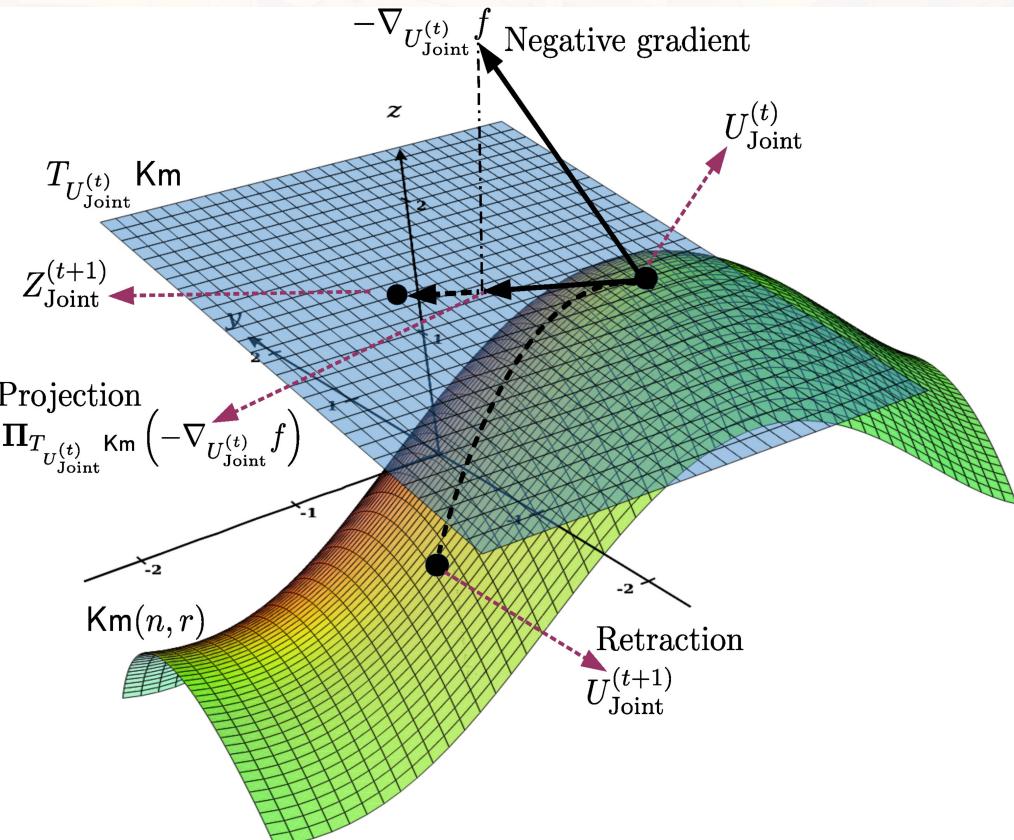
Initial iterates: $U_{\text{Joint}}^{(0)}, U_j^{(0)}$ for $j = 1, \dots, M$

Optimization of U_{Joint}

$$U_{\text{Joint}}^{(t+1)} = \text{PKm}_{U_{\text{Joint}}^{(t)}} (\mathbf{Z}_{\text{Joint}}^{(t+1)})$$

$$4.\text{PKm}_{U_{\text{Joint}}^{(t)}} (\mathbf{Z}_{\text{Joint}}^{(t+1)}) = \exp(B) \exp(Q') U_{\text{Joint}}^{(t)},$$

Iteration t : $U_{\text{Joint}}^{(t)}, U_j^{(t)}$



Manifold Optimization

Initial iterates: $U_{\text{Joint}}^{(0)}, U_j^{(0)}$ for $j = 1, \dots, M$ Iteration t : $U_{\text{Joint}}^{(t)}, U_j^{(t)}$

Optimization of U_j over Stiefel manifold

- 1: Compute negative gradient $\mathbf{Q}_j^{(t)} \leftarrow -\nabla_{U_j^{(t)}} \mathbf{f}$
- 2: Project negative gradient onto tangent space:
$$\mathbf{Z}_j^{(t)} \leftarrow \Pi_{T_{U_j^{(t)}} \text{St}} (\mathbf{Q}_j^{(t)})$$
 using (6.20).
- 3: $\mathbf{Z}_j^{(t+1)} \leftarrow U_j^{(t)} + \eta_S \mathbf{Z}_j^{(t)}.$
- 4: Find retractive projection $\mathbf{PSt}_{U_j^{(t)}} (\mathbf{Z}_j^{(t+1)})$ using (6.21)
- 5: $U_j^{(t+1)} \leftarrow \mathbf{PSt}_{U_j^{(t)}} (\mathbf{Z}_j^{(t+1)}).$

MIMIC: Proposed Algorithm

For each view:

Compute Laplacian L_j

$U_j^{(0)} \leftarrow r$ largest eig-vec of L_j
 $\alpha_j \leftarrow$ eig-vals of L_j

Compute joint Laplacian $\mathbf{L}_{\text{Joint}}^r$

$U_{\text{Joint}}^{(0)} \leftarrow r$ largest eig-vec of $\mathbf{L}_{\text{Joint}}^r$
 $t = 0, f^{(0)} \leftarrow f(U_{\text{Joint}}^{(0)}, U_1^{(0)}, \dots, U_M^{(0)})$

Repeat until convergence: $(f^{(t)} - f^{(t+1)}) > \epsilon$

$U_j^{(t+1)} \leftarrow \text{Optimize_Stiefel}(L_j^r, U_{\text{Joint}}^{(t)}, U_j^{(t)}, \eta_S), \forall j = 1, \dots, M$

$U_{\text{Joint}}^{(t+1)} \leftarrow \text{Optimize_}k\text{-means}(\mathbf{L}_{\text{Joint}}^r, U_{\text{Joint}}^{(t)}, U_1^{(t)}, \dots, U_M^{(t)}, \eta_K, \xi)$

Compute $f^{(t+1)} \leftarrow f(U_{\text{Joint}}^{(t+1)}, U_1^{(t+1)}, \dots, U_M^{(t+1)}), t = t + 1$

Optimal solution: $U_{\text{Joint}}^* \leftarrow U_{\text{Joint}}^{(t+1)}$
 k -means on first k columns of U_{Joint}^*

Convergence and Asymptotic Bounds

Theorem.(Convergence) Every limit point of the sequence $\{U_{\text{Joint}}^{(t)}\}_{t=0,1,2,\dots}$, generated by the proposed algorithm for a set of given U_j 's for $j \in \{1, \dots, M\}$, is a critical point of the cost function f .

Theorem.(Asymptotic Bound) There exists an integer $t' \geq 0$ such that

$$f(U_{\text{Joint}}^{(t+1)}) - f(U_{\text{Joint}}^*) \leq c \left(f(U_{\text{Joint}}^{(t)}) - f(U_{\text{Joint}}^*) \right),$$

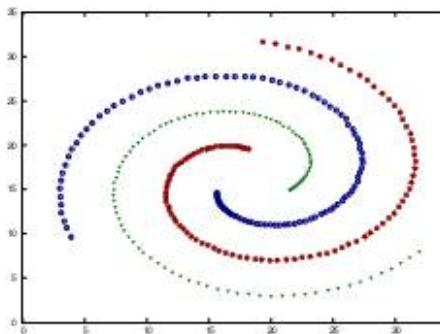
for all $t \geq t'$, where

$$c = 1 - 2\sigma(\lambda_{k+1} - \lambda_k) \min \left\{ \eta, \frac{2\beta(1-\sigma)}{(\lambda_n - \lambda_1)} \right\},$$

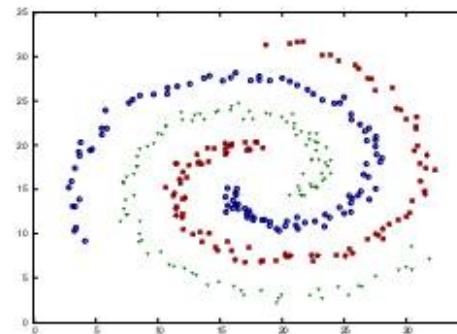
where η is the step length, σ and β are Armijo criterion parameters, $\lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_n$ are the eigenvalues of the Hessian of f given by

$$\boldsymbol{\Xi} = \left(\xi \mathbf{I}_n - \mathbf{L}_{\text{Joint}}^k - \frac{1}{M} \sum_{j=1}^M U_j U_j^T \right)$$

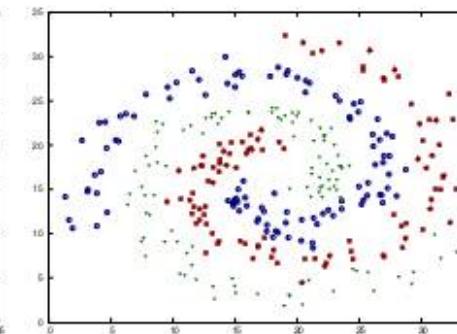
Significance of Asymptotic Convergence Bound



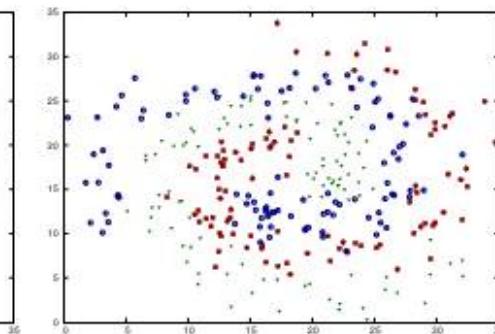
(a) Original Data Set



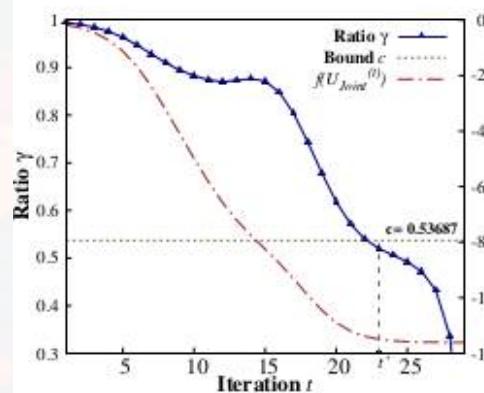
(b) Noise with STD= 0.5



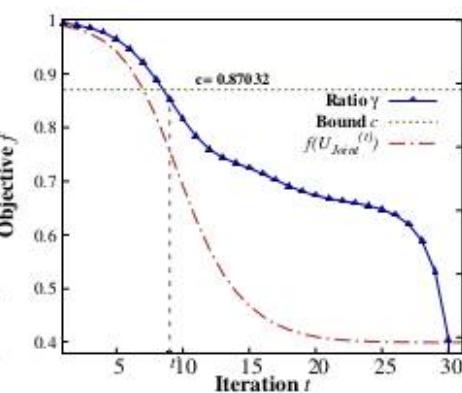
(c) Noise with STD= 1



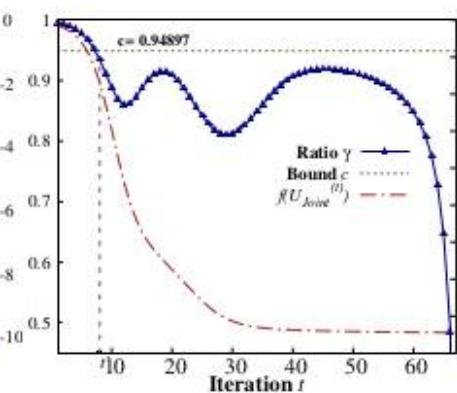
(d) Noise with STD= 1.5



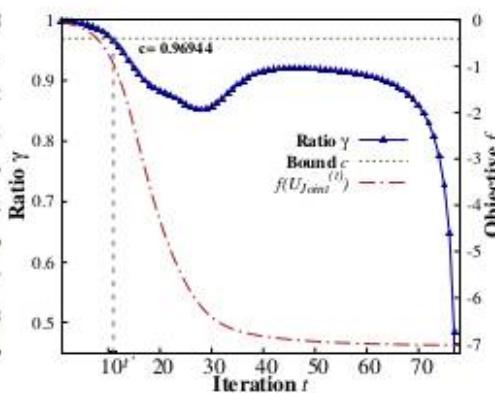
(a) $c = 0.53687$



(b) $c = 0.87032$



(c) $c = 0.94897$



(d) $c = 0.96944$

- Asymptotic bound c can be used to infer the separability of clusters in data set

Experimental Results & Data Sets

Multi-Omics Cancer Data Sets:

CESC, LGG, BRCA, OV, STAD, CRC, LUNG, and KIDNEY

Views: mDNA, RNA, miRNA, RPPA Clusters: 2, 3, 4

Benchmark Data Sets:

Multi-source News Article Clustering: 3Sources and BBC

Views: Reuters, BBC News, The Guardian Clusters: 5, 6

Image Data Sets: Digits, 100Leaves, and ALOI

Views: RGB histogram, HSV histogram, Fourier coefficients

Clusters: 10, 100, 100

Importance of Data Integration

Views→		Shape	Texture	Margin	MiMIC
Accuracy	100Leaves	0.3095(9.1e-3)	0.4777(1.4e-2)	0.5786(1.1e-2)	0.8185 (1.5e-2)
NMI		0.6479(6.7e-3)	0.7327(5.6e-3)	0.7940(4.4e-3)	0.9302 (4.1e-3)
ARI		0.1820(5.8e-3)	0.3265(1.4e-2)	0.4478(9.8e-3)	0.7431 (2.5e-2)
F-measure		0.3525(7.4e-3)	0.5139(1.1e-2)	0.6113(9.7e-3)	0.8492 (1.1e-2)
Rand		0.9699(1.5e-3)	0.9839(7.2e-4)	0.9880(2.9e-4)	0.9913 (1.17e-3)
Purity		0.3696(7.5e-3)	0.5216(1.2e-2)	0.6203(7.6e-3)	0.7772 (1.53e-2)
Views→		BBC	Guardian	Reuters	MiMIC
Accuracy	3Sources	0.7159(0.0)	0.6508(0.0)	0.5562(0.0)	0.7360 (5.9e-2)
NMI		0.6390(0.0)	0.5270(0.0)	0.5347(0.0)	0.6433 (3.5e-2)
ARI		0.6082 (0.0)	0.4119(0.0)	0.41434(0.0)	0.5957(6.6e-2)
F-measure		0.7656 (0.0)	0.7036(0.0)	0.6482(0.0)	0.7581(5.0e-2)
Rand		0.8624 (0.0)	0.7983(0.0)	0.7982(0.0)	0.8514(2.6e-2)
Purity		0.7869(0.0)	0.6982(0.0)	0.6982(0.0)	0.7946 (2.2e-2)

- MiMIC has better performance compared to spectral clustering on individual views

Importance of Data Integration

Views→	Segment1	Segment2	Segment3	Segment4	MiMIC
Accuracy	0.6202(2.1e-3)	0.6202(3.6e-2)	0.6102(3.6e-2)	0.5550(3.0e-3)	0.8715(0.0)
NMI	0.4312(1.7e-3)	0.4459(5.7e-2)	0.4097(1.3e-3)	0.4033(8.2e-3)	0.7182(→0)
ARI	BBC	0.3405(6.6e-2)	0.3895(8.9e-2)	0.3429(7.0e-3)	0.7273(0.0)
F-measure		0.6514(1.2e-2)	0.6363(3.9e-2)	0.6435(1.2e-2)	0.8613(0.0)
Rand		0.7256(1.7e-2)	0.7174(6.3e-2)	0.7425(7.4e-3)	0.8959(0.0)
Purity		0.6212(3.5e-3)	0.6218(3.7e-2)	0.6120(3.6e-2)	0.8715(0.0)
Views→	RGB	HSV	Haralick	ColorSimilarity	MiMIC
Accuracy	0.4215(1.1e-2)	0.4433(7.0e-3)	0.1001(2.3e-3)	0.5191(1.1e-2)	0.5742(7.4e-3)
NMI	0.7179(3.9e-3)	0.7093(5.1e-3)	0.3659(4.1e-3)	0.7683(4.9e-3)	0.7805(2.3e-3)
ARI	ALOI	0.2915(1.4e-2)	0.2979(1.9e-2)	0.0550(6.8e-4)	0.4233(6.6e-3)
F-measure		0.4789(1.0e-2)	0.5136(7.9e-3)	0.1209(1.5e-3)	0.6221(4.9e-3)
Rand		0.9745(1.8e-3)	0.9759(2.2e-3)	0.8938(7.0e-3)	0.9840(3.7e-4)
Purity		0.4717(9.9e-3)	0.4876(7.3e-3)	0.1094(2.4e-3)	0.6119(5.7e-3)

- Data integration preserves better cluster structure

Importance of k -Means & Stiefel Manifolds

Manifold→	k -Means	Stiefel	MiMIC	k -Means	Stiefel	MiMIC	
Digits	0.8205(0.0)	0.6480(2.2e-2)	0.9207 (4.2e-4)	BBC	0.7345(8.5e-2)	0.7732(1.3e-3)	0.8715 (0.0)
	0.8350(→0)	0.6535(8.8e-3)	0.8597 (4.8e-4)		0.6167(6.1e-2)	0.5983(2.4e-3)	0.7182 (→0)
	0.7687(0.0)	0.5155(1.5e-2)	0.8352 (8.1e-4)		0.5910(1.2e-1)	0.6382(3.2e-3)	0.7273 (0.0)
	0.8756(0.0)	0.6922(1.8e-2)	0.9209 (4.1e-4)		0.7614(6.1e-1)	0.7806(1.3e-2)	0.8613 (0.0)
	0.9564(0.0)	0.9005(5.7e-03)	0.9703 (1.4e-4)		0.8315(6.1e-2)	0.8659(1.1e-3)	0.8959 (0.0)
	0.8315(0.0)	0.6665(1.8e-02)	0.9207 (4.2e-4)		0.7356(8.5e-2)	0.7747(1.3e-3)	0.8715 (0.0)
Sources	0.6497(2.4e-3)	0.6798(7.2e-2)	0.7360 (5.9e-2)	LGG	0.9288(0.0)	0.6292(0.0)	0.9625 (0.0)
	0.6221(4.6e-3)	0.6020(7.2e-2)	0.6433 (3.5e-2)		0.7949(→0)	0.4305(→0)	0.8543 (→0)
	0.5173(2.1e-3)	0.5226(1.2e-1)	0.5957 (6.6e-2)		0.7790(0.0)	0.2842(0.0)	0.8790 (0.0)
	0.6927(2.7e-4)	0.7330(6.3e-2)	0.7581 (5.0e-2)		0.9269(0.0)	0.6313(0.0)	0.9623 (0.0)
	0.8170(1.7e-4)	0.8310(4.1e-2)	0.8514 (2.61e-2)		0.8940(0.0)	0.6632(0.0)	0.9424 (0.0)
	0.7739(2.4e-3)	0.7455(5.9e-2)	0.7946 (2.28e-2)		0.9288(0.0)	0.6779(0.0)	0.9625 (0.0)

- Performance of k -means manifold is better than Stiefel manifold:
- Reason: k -means manifold optimizes joint view U_{joint} , Stiefel manifold here reports performance of best individual view

Importance of k -Means & Stiefel Manifolds

Manifold→	k -Means	Stiefel	MiMIC	k -Means	Stiefel	MiMIC	
Accuracy	0.7139(2.2e-2)	0.6457(1.8e-2)	0.8185 (1.5e-2)	STAD	0.6867(3.3e-2)	0.5165(0.0)	0.7727 (0.0)
	0.8887(5.4e-3)	0.8278(5.9e-3)	0.9302 (4.1e-3)		0.4412(4.2e-2)	0.2985(→0)	0.5220 (→0)
	0.6374(2.1e-2)	0.5323(1.5e-2)	0.7431 (2.5e-2)		0.3615(5.1e-2)	0.1616(0.0)	0.4650 (0.0)
	0.7543(1.7e-2)	0.6741(1.5e-2)	0.8492 (1.1e-2)		0.6930(3.1e-2)	0.5241(0.0)	0.7830 (0.0)
	0.9920(6.4e-4)	0.9903(3.8e-4)	0.9913 (1.1e-3)		0.6938(1.1e-2)	0.6433(0.0)	0.7698 (0.0)
	0.7502(1.8e-2)	0.6764(1.5e-2)	0.7772 (1.5e-2)		0.6884(2.9e-2)	0.5909(0.0)	0.7727 (0.0)
NMI	0.5044(1.4e-2)	0.5068(1.8e-2)	0.5742 (7.4e-3)	BRCA	0.7085(0.0)	0.7889(0.0)	0.7964 (0.0)
	0.7461(4.5e-3)	0.7462(5.2e-3)	0.7805 (2.3e-3)		0.4964(→0)	0.5373(→0)	0.5553 (→0)
	0.3874(1.6e-2)	0.3850(1.7e-2)	0.4233 (6.6e-3)		0.4291(0.0)	0.5331(0.0)	0.5474 (0.0)
	0.5739(1.0e-2)	0.5748(1.4e-2)	0.6221 (4.9e-3)		0.7072(0.0)	0.7905(0.0)	0.7997 (0.0)
	0.9828(1.1e-3)	0.9826(1.2e-3)	0.9840 (3.7e-4)		0.7670(0.0)	0.8075(0.0)	0.8152 (0.0)
	0.5461(1.1e-2)	0.5483(1.5e-2)	0.6119 (5.7e-3)		0.7085(0.0)	0.7889(0.0)	0.7964 (0.0)
ARI	0.7139(2.2e-2)	0.6457(1.8e-2)	0.8185 (1.5e-2)	ALOI	0.6867(3.3e-2)	0.5165(0.0)	0.7727 (0.0)
	0.8887(5.4e-3)	0.8278(5.9e-3)	0.9302 (4.1e-3)		0.4412(4.2e-2)	0.2985(→0)	0.5220 (→0)
	0.6374(2.1e-2)	0.5323(1.5e-2)	0.7431 (2.5e-2)		0.3615(5.1e-2)	0.1616(0.0)	0.4650 (0.0)
	0.7543(1.7e-2)	0.6741(1.5e-2)	0.8492 (1.1e-2)		0.6930(3.1e-2)	0.5241(0.0)	0.7830 (0.0)
	0.9920(6.4e-4)	0.9903(3.8e-4)	0.9913 (1.1e-3)		0.6938(1.1e-2)	0.6433(0.0)	0.7698 (0.0)
	0.7502(1.8e-2)	0.6764(1.5e-2)	0.7772 (1.5e-2)		0.6884(2.9e-2)	0.5909(0.0)	0.7727 (0.0)

- Best performance when considering both k -means and Steifel manifold in MiMIC algorithm

Comparative Performance on Benchmark Data Sets

	Algorithm → MKC	CoregSC	MSC	ASMV	MGL	MCGL	GMC	CoAla	MiMIC
Digits	Accuracy	0.4924(2.77e-1)	0.7556(5.96e-2)	0.7918(8.21e-2)	0.5745(→0)	0.7440(8.19e-2)	0.8530(0.0)	0.8820(→0)	0.9207 (4.21e-4)
	NMI	0.5325(3.68e-1)	0.7421(3.27e-2)	0.7560(3.24e-2)	0.6709(→0)	0.8264(4.73e-2)	0.9055 (0.0)	0.9050(→0)	0.8597(4.88e-4)
	ARI	0.4280(2.99e-1)	0.6885(5.73e-2)	0.6803(6.28e-2)	0.4047(→0)	0.6888(1.07e-1)	0.8313(→0)	0.8502 (→0)	0.7645(0.0)
	F-measure	0.5130(2.33e-2)	0.6934(5.11e-2)	0.7129(5.58e-2)	0.4852(→0)	0.7238(9.37e-2)	0.8493(→0)	0.8658(→0)	0.9209 (4.15e-4)
3Sources	Accuracy	0.4663(1.06e-1)	0.5479(2.99e-2)	0.4751(2.97e-2)	0.3373(→0)	0.6751(6.67e-2)	0.3077(→0)	0.6923(→0)	0.6508(0.0)
	NMI	0.3665(1.00e-1)	0.5238(1.98e-2)	0.3850(2.27e-2)	0.0896(→0)	0.5768(8.61e-2)	0.1034(→0)	0.6216 (0.0)	0.6198(→0)
	ARI	0.2461(1.40e-1)	0.3339(2.85e-2)	0.2618(3.81e-2)	-0.021(→0)	0.4431(1.17e-1)	-0.033(→0)	0.4431(0.0)	0.5957 (6.69e-2)
	F-measure	0.4114(1.08e-1)	0.4775(1.91e-2)	0.4087(3.05e-2)	0.3528(→0)	0.5966(7.12e-2)	0.3417(0.0)	0.6047(0.0)	0.7581 (5.92e-2)
BBC	Accuracy	0.6034(1.10e-1)	0.4701(0.0)	0.6732(4.94e-2)	0.3372(0.0)	0.5396(1.10e-1)	0.3533(→0)	0.6934(→0)	0.8108 (4.36e-3)
	NMI	0.4786(8.51e-2)	0.2863(0.0)	0.5531(1.44e-2)	0.0348(0.0)	0.3697(1.89e-1)	0.0741(→0)	0.5628(0.0)	0.6536 (1.96e-2)
	ARI	0.3450(1.21e-1)	0.2727(0.0)	0.4658(2.20e-2)	0.0018(→0)	0.3153(1.66e-1)	0.0053(→0)	0.4789(→0)	0.7102 (2.78e-2)
	F-measure	0.5018(9.03e-2)	0.4879(0.0)	0.5877(1.83e-2)	0.3781(0.0)	0.5402(8.53e-2)	0.3762(0.0)	0.6333(→0)	0.8138 (9.93e-4)
100Leaves	Accuracy	0.0100(0.0)	0.7706(2.58e-2)	0.7379(2.21e-2)	0.7906(→0)	0.6904(2.42e-2)	0.8106(→0)	0.8238 (→0)	0.7384(1.34e-2)
	NMI	0.0000(0.0)	0.9165(5.90e-3)	0.9014(7.60e-3)	0.9009(→0)	0.8753(7.60e-3)	0.9130(0.0)	0.9292(0.0)	0.8893(4.06e-3)
	ARI	0.0000(0.0)	0.7229(1.92e-2)	0.6788(2.26e-2)	0.6104(→0)	0.3858(5.65e-2)	0.5155(→0)	0.4974(→0)	0.6550(1.41e-2)
	F-measure	0.0186(0.0)	0.7257(1.90e-2)	0.6821(2.23e-2)	0.6148(→0)	0.3944(5.53e-2)	0.5217(0.0)	0.5042(→0)	0.7672(1.19e-2)
ALOI	Accuracy	0.0101(→0)	0.5217(2.13e-2)	0.4738(7.65e-2)	0.4555(→0)	0.4807(1.51e-2)	0.4625(→0)	0.5705 (→0)	0.5594(1.44e-2)
	NMI	0.0000(0.0)	0.6993(1.32e-2)	0.6358(5.44e-2)	0.6767(→0)	0.7052(7.00e-3)	0.6657(→0)	0.7350(0.0)	0.7654(3.72e-3)
	ARI	0.0000(→0)	0.4097(4.52e-2)	0.3305(4.81e-2)	0.0533(0.0)	0.1987(4.37e-2)	0.0441(0.0)	0.4305(→0)	0.4352 (1.18e-2)
	F-measure	0.0196(→0)	0.4051(2.38e-2)	0.3366(3.68e-2)	0.0712(→0)	0.2112(4.22e-2)	0.0621(→0)	0.4366(→0)	0.6213(1.15e-2)

- MiMIC algorithm gives best performance on all five benchmark data sets across all measures, except for three cases: ARI and NMI on Digits, Accuracy on 100Leaves, and ARI on ALOI data set.

Comparative Performance on Omics Data Sets

Algorithm→	Consensus	Statistical Model Based			Subspace Based			Graph Based		Manifold	
	COCA	NormS	LR Acluster	iCluster	PCA-con	SURE	JIVE	SNF	CoALa	MiMIC	
BRCA	Accuracy	0.7434(7.94e-4)	0.7688(0.0)	0.7110(0.0)	0.7638(0.0)	0.7587(0.0)	0.7663(0.0)	0.6859(0.0)	0.6783(0.0)	0.7613(0.0)	0.7964(0.0)
	NMI	0.5002(3.48e-4)	0.4287(→0)	0.5437(→0)	0.5176(→0)	0.5506(→0)	0.4558(0.0)	0.4368(0.0)	0.5528(→0)	0.5281(→0)	0.5553(→0)
	ARI	0.4864(4.50e-4)	0.5090(0.0)	0.4035(0.0)	0.4745(0.0)	0.5038(0.0)	0.5104(0.0)	0.3772(0.0)	0.4111(0.0)	0.4874(0.0)	0.5474(0.0)
	F-measure	0.7457(8.13e-4)	0.7699(0.0)	0.7101(0.0)	0.7658(0.0)	0.7601(0.0)	0.7683(0.0)	0.6889(0.0)	0.6865(0.0)	0.7660(0.0)	0.7997(0.0)
	Rand	0.7905(1.92e-4)	0.7999(0.0)	0.7521(0.0)	0.7842(0.0)	0.7984(0.0)	0.8010(0.0)	0.7464(0.0)	0.7602(0.0)	0.7922(0.0)	0.8152(0.0)
	Purity	0.7434(7.95e-4)	0.7688(0.0)	0.7110(0.0)	0.7638(0.0)	0.7587(0.0)	0.7663(0.0)	0.6859(0.0)	0.6959(0.0)	0.7613(0.0)	0.7964(0.0)
LGG	Accuracy	0.6591(0.0)	0.7940(0.0)	0.4719(0.0)	0.4382(0.0)	0.6666(0.0)	0.7940(0.0)	0.5617(0.0)	0.8689(0.0)	0.9737(0.0)	0.9625(0.0)
	NMI	0.2772(0.0)	0.5325(0.0)	0.1240(→0)	0.1379(→0)	0.3438(0.0)	0.5335(0.0)	0.2299(→0)	0.6253(0.0)	0.8689(→0)	0.8543(→0)
	ARI	0.2533(0.0)	0.4649(0.0)	0.1030(0.0)	0.0996(0.0)	0.3031(0.0)	0.4668(0.0)	0.1606(0.0)	0.6331(0.0)	0.9199(0.0)	0.8790(0.0)
	F-measure	0.6608(0.0)	0.7916(0.0)	0.5137(0.0)	0.5187(0.0)	0.6574(0.0)	0.7904(0.0)	0.5757(0.0)	0.8720(0.0)	0.9737(0.0)	0.9623(0.0)
	Rand	0.6454(0.0)	0.7465(0.0)	0.5831(0.0)	0.5821(0.0)	0.6616(0.0)	0.7465(0.0)	0.6056(0.0)	0.8268(0.0)	0.9622(0.0)	0.9424(0.0)
	Purity	0.6591(0.0)	0.7940(0.0)	0.5280(0.0)	0.5355(0.0)	0.6666(0.0)	0.7940(0.0)	0.5730(0.0)	0.8689(0.0)	0.9737(0.0)	0.9625(0.0)
STAD	Accuracy	0.4450(3.34e-2)	0.5702(0.0)	0.4256(0.0)	0.3512(0.0)	0.6900(0.0)	0.6983(0.0)	0.4049(0.0)	0.5661(0.0)	0.7685(0.0)	0.7727(0.0)
	NMI	0.1309(4.77e-3)	0.1805(→0)	0.1259(→0)	0.0650(→0)	0.3654(0.0)	0.3511(→0)	0.1288(→0)	0.3216(0.0)	0.5107(0.0)	0.5220(→0)
	ARI	0.0740(1.02e-2)	0.1625(0.0)	0.0912(0.0)	0.0288(0.0)	0.3204(0.0)	0.3445(0.0)	0.0657(0.0)	0.2694(0.0)	0.4559(0.0)	0.4650(0.0)
	F-measure	0.4558(2.50e-2)	0.5770(0.0)	0.4746(0.0)	0.3832(0.0)	0.6959(0.0)	0.7056(0.0)	0.4487(0.0)	0.6333(0.0)	0.7778(0.0)	0.7830(0.0)
	Rand	0.5981(1.32e-2)	0.6435(0.0)	0.6122(0.0)	0.5855(0.0)	0.7110(0.0)	0.7216(0.0)	0.5981(0.0)	0.6945(0.0)	0.7661(0.0)	0.7698(0.0)
	Purity	0.5173(9.50e-3)	0.5950(0.0)	0.5619(0.0)	0.4917(0.0)	0.6900(0.0)	0.6983(0.0)	0.5165(0.0)	0.6363(0.0)	0.7685(0.0)	0.7727(0.0)
LUNG	Accuracy	0.9284(0.0)	0.9359(0.0)	0.9344(0.0)	0.6333(0.0)	0.9388(0.0)	0.9418(0.0)	0.9269(0.0)	0.9493(0.0)	0.9403(0.0)	0.9463(0.0)
	NMI	0.6287(0.0)	0.6650(0.0)	0.6535(0.0)	0.0627(0.0)	0.6773(→0)	0.6878(0.0)	0.6333(0.0)	0.7152(0.0)	0.6970(0.0)	0.7173(0.0)
	ARI	0.7339(0.0)	0.7597(0.0)	0.7545(0.0)	0.0696(0.0)	0.7701(0.0)	0.7806(0.0)	0.7288(0.0)	0.8072(0.0)	0.7754(0.0)	0.7965(0.0)
	F-measure	0.9283(0.0)	0.9357(0.0)	0.9342(0.0)	0.6299(0.0)	0.9386(0.0)	0.9417(0.0)	0.9266(0.0)	0.9492(0.0)	0.9400(0.0)	0.9461(0.0)
	Rand	0.8669(0.0)	0.8798(0.0)	0.8772(0.0)	0.5348(0.0)	0.8850(0.0)	0.8903(0.0)	0.8644(0.0)	0.9036(0.0)	0.8877(0.0)	0.8983(0.0)
	Purity	0.9284(0.0)	0.9359(0.0)	0.9344(0.0)	0.6333(0.0)	0.9388(0.0)	0.9418(0.0)	0.9269(0.0)	0.9493(0.0)	0.9403(0.0)	0.9463(0.0)

- MiMIC algorithm gives best performance on BRCA and STAD data sets
- Graph based algorithms CoALa and SNF has best performance on LGG and LUNG data sets.

Highlights

- Manifold optimization to capture low-rank non-linear geometry of views
- K-means Manifold: Joint clustering subspaces, Stiefel manifold: Disagreement minimization
- Asymptotic convergence bound can be used to make inference regarding the separability of the clusters
- Proposed algorithm outperforms state-of-the-art integrative clustering approaches on all benchmark and majority of multi-omics cancer data sets

CHAPTER 7

Geometry Aware Multi-View Clustering over Riemannian Manifolds

A. Khan and P. Maji, “Geometry Aware Multi-View Clustering over Riemannian Manifolds,” **IEEE Transactions on Pattern Analysis and Machine Intelligence**, pages 1-13, 2021
(Manuscript ID: TPAMI-2021-08-1458).

Motivation

MiMIC

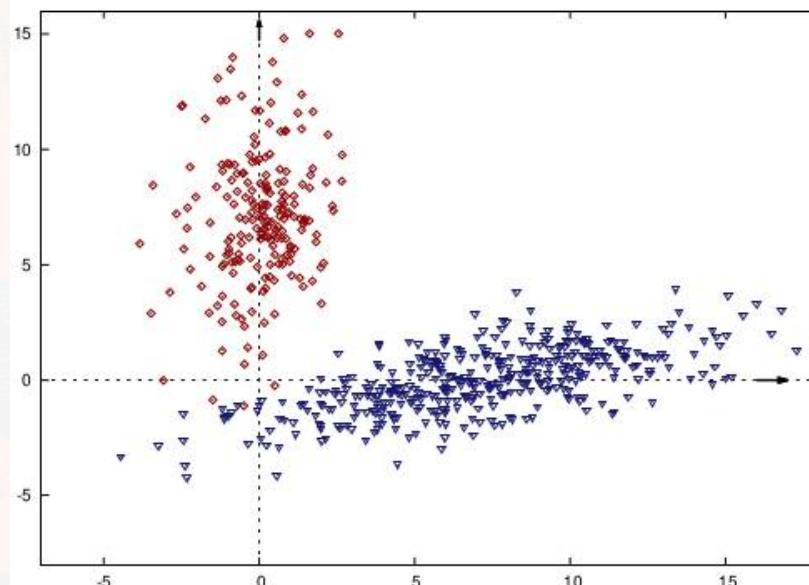
- Optimizes only the joint and the individual cluster indicator matrices.
- View/Graph weights are fixed [eigenvalue based heuristic].

Simultaneous optimization of graph connectivity and graph weights along with the cluster indicator subspaces is expected to preserve better cluster structure

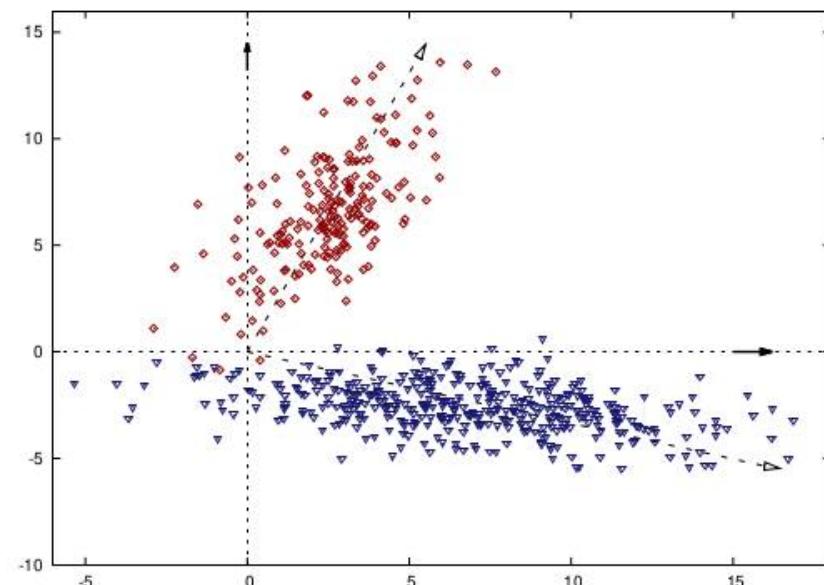
Effect of the Change of Basis on Cluster Structure

Spectral clustering on individual views

$$\underset{U_m \in \Re^{n \times k}}{\text{minimize}} -\text{tr}(U_m^T L_m U_m) \text{ such that } U_m^T U_m = \mathbf{I}_k$$



(a) Axes aligned basis



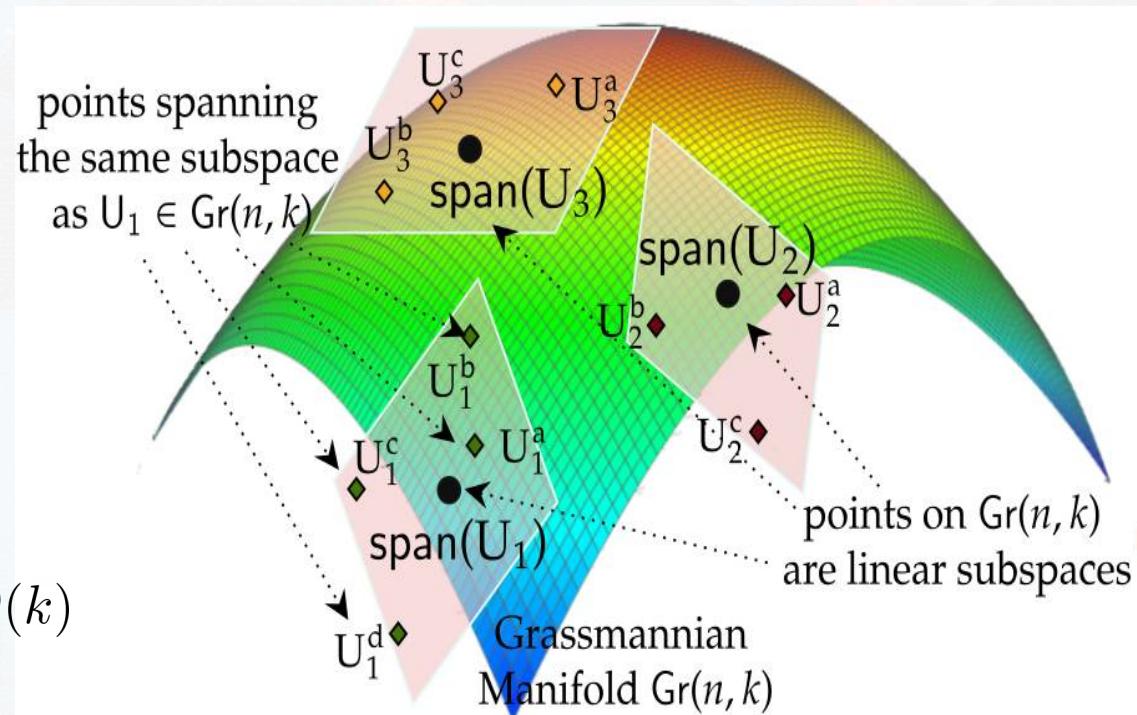
(b) Rotated basis

Incorporating Geometry Awareness

$$\text{Gr}(n, k) := \{\text{span}(U) \in \Re^{n \times k} \mid U^T U = \mathbf{I}_k\}$$

Space formed by k -dimensional linear subspaces of n -dimensional Euclidean space

$$\text{Gr}(n, k) := \{U \in \Re^{n \times k} \mid U^T U = \mathbf{I}_k\}/O(k)$$



Incorporating Geometry Awareness

$$\text{Gr}(n, k) := \{\text{span}(U) \in \Re^{n \times k} \mid U^T U = \mathbf{I}_k\}$$

Unconstrained optimization of spectral clustering objective over Grassmannian manifold:

$$\underset{\begin{array}{c} \text{minimize} \\ \text{span}(U_{\text{Joint}}) \\ \text{span}(U_m) \end{array} \Big\} \in \text{Gr}(n, k)} - \text{tr} \left(U_{\text{Joint}}^T \mathbf{L}_{\text{Joint}}^r U_{\text{Joint}} \right) - \frac{1}{M} \sum_{m=1}^M \text{tr}(U_m^T L_m^r U_m)$$

Disagreement between subspaces:

$$d_\theta(U_{\text{Joint}}, U_m) = \| U_{\text{Joint}} U_{\text{Joint}}^T - U_m U_m^T \|_F^2$$

Multi-View Integration

Joint clustering

$$\underset{\substack{\text{span}(U_{\text{Joint}}) \\ \text{span}(U_m)}}{\underset{\in \text{Gr}(n,k)}{\text{minimize}}} \quad f(U_{\text{Joint}}, U_1, \dots, U_M) = -\frac{1}{2} \text{tr} \left(U_{\text{Joint}}^T \mathbf{L}_{\text{Joint}}^r U_{\text{Joint}} \right)$$

$$+ \frac{1}{2M(M-1)} \sum_{\substack{i,j=1 \\ i \neq j}}^M d_\theta(U_i, U_j) + \frac{1}{2M} \sum_{m=1}^M \left[- \text{tr}(U_m^T L_m^r U_m) + d_\theta(U_{\text{Joint}}, U_m) \right]$$

Pairwise disagreement Individual clustering Joint disagreement

Updation of Graph Connectivity

L_m^r is symmetric and positive definite (SPD)

SPD manifold

$$\mathcal{S}_{++}^n = \{A \in \Re^{n \times n} \mid A = A^T \text{ and } v^T A v > 0 \text{ for } v \in \Re^n, v \neq 0\}$$

Proposed Objective

$$\underset{\substack{\text{span}(U_{\text{Joint}}), \text{span}(U_m) \in \text{Gr}(n,k) \\ L_m^r \in \mathcal{S}_{++}^n, \alpha_m \in \Re}}{\text{minimize}} -\frac{1}{2} \text{tr} \left(U_{\text{Joint}}^T \left(\sum_{m=1}^M \alpha_m^\kappa L_m^r \right) U_{\text{Joint}} \right)$$

$$-\frac{1}{2M} \sum_{m=1}^M \left[\text{tr}(U_m^T L_m^r U_m) + \text{tr}(U_{\text{Joint}} U_{\text{Joint}}^T U_m U_m^T) \right] - \frac{1}{2M(M-1)} \sum_{\substack{i,j=1 \\ i \neq j}}^M \text{tr}(U_i U_i^T U_j U_j^T),$$

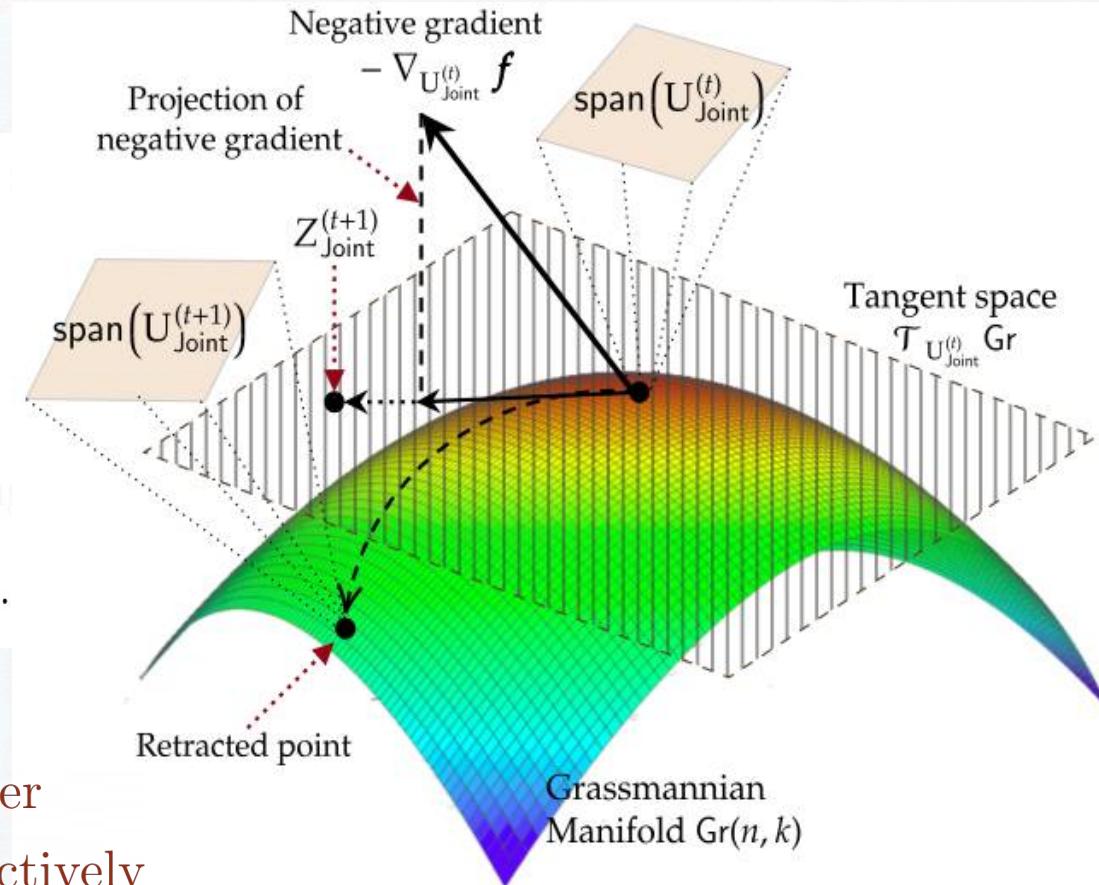
$$\text{such that } \alpha_m \geq 0, \quad \sum_{m=1}^M \alpha_m = 1.$$

Manifold Optimization

Optimization of U_{Joint}

- 1: Compute negative gradient $\mathbf{Q}_{\text{Joint}}^{(t)} \leftarrow \left[-\nabla_{U_{\text{Joint}}^{(t)}} f \right]$
- 2: Project negative gradient onto tangent space:
 $Z_{\text{Joint}}^{(t)} \leftarrow \Pi_{U_{\text{Joint}}^{(t)}} \left(\mathbf{Q}_{\text{Joint}}^{(t)} \right)$ using (7.16).
- 3: $Z_{\text{Joint}}^{(t+1)} \leftarrow U_{\text{Joint}}^{(t)} + \eta_G Z_{\text{Joint}}^{(t)}$.
- 4: Find retractive projection $\mathbf{PGr}_{U_{\text{Joint}}^{(t)}} \left(Z_{\text{Joint}}^{(t+1)} \right)$ using
- 5: Next iterate: $\text{span} \left(U_{\text{Joint}}^{(t+1)} \right) \leftarrow \mathbf{PGr}_{U_{\text{Joint}}^{(t)}} \left(Z_{\text{Joint}}^{(t+1)} \right)$.

Similarly optimize each U_j and L_j^r over Grassmannian and SPD manifolds, respectively



Weight Optimization

$$\underset{\alpha_m \in \Re}{\text{minimize}} \quad \sum_{m=1}^M -\alpha_m^\kappa \mathfrak{g}_m \text{ such that } \alpha_m \geq 0, \quad \sum_{m=1}^M \alpha_m = 1,$$

$$\text{where } \mathfrak{g}_m = \frac{1}{2} \text{tr} \left(U_{\text{Joint}}^T L_m^r U_{\text{Joint}} \right)$$



$$\alpha_m = \frac{(\mathfrak{g}_m)^{\frac{1}{1-\kappa}}}{\sum_{m=1}^M (\mathfrak{g}_m)^{\frac{1}{1-\kappa}}}$$

GEARS: Proposed Algorithm

For each view:

Compute Laplacian L_j

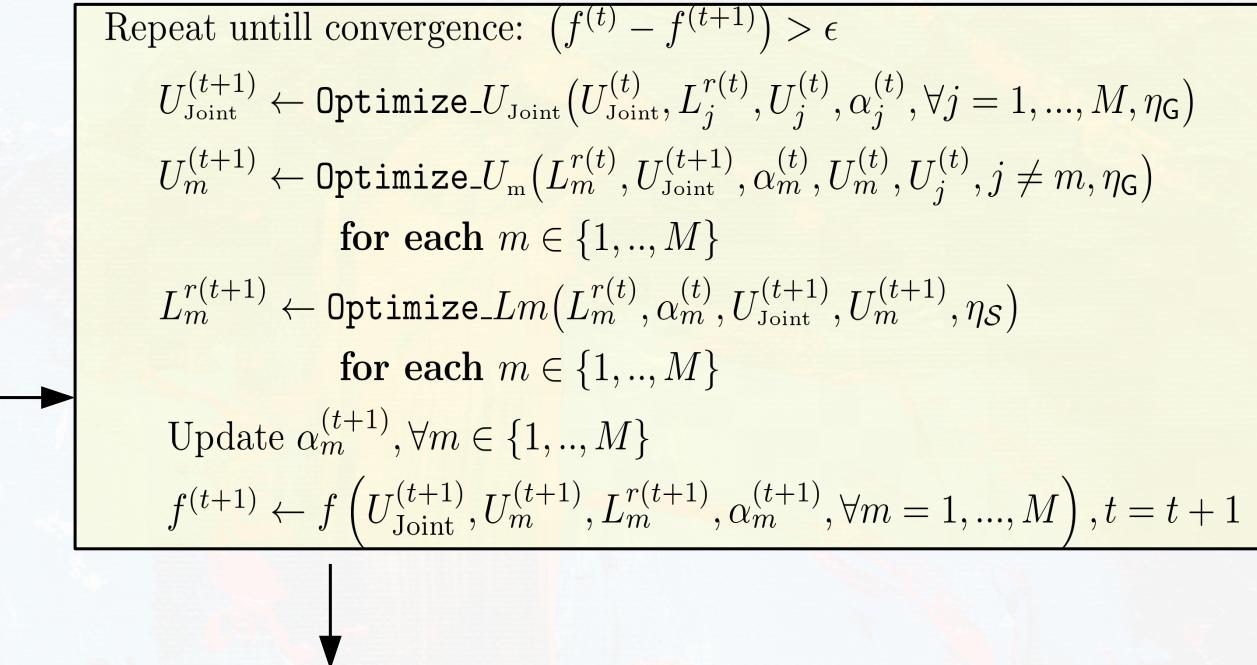
$U_j^{(0)} \leftarrow r$ largest eig-vec of L_j
 $\alpha_j \leftarrow$ eig-vals of L_j



Compute joint Laplacian $\mathbf{L}_{\text{Joint}}^r$

$U_{\text{Joint}}^{(0)} \leftarrow r$ largest eig-vec of $\mathbf{L}_{\text{Joint}}^r$

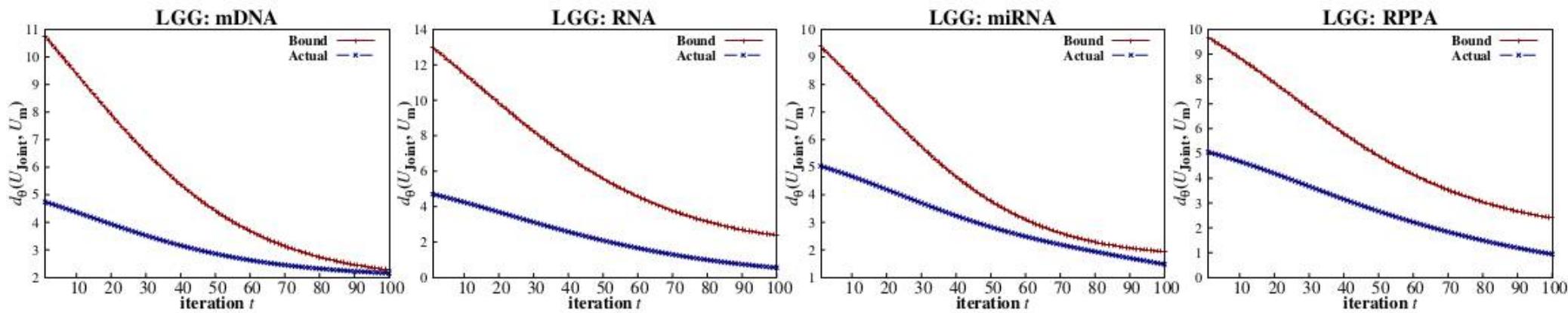
$t = 0, f^{(0)} \leftarrow f \left(U_{\text{Joint}}^{(0)}, U_1^{(0)}, \dots, U_M^{(0)} \right)$



Convergence result of MiMIC
hold analogically for GeARS

Grassmannian Disagreement Bounds

$$d_\theta(U_{\text{Joint}}^{(t)}, U_m^{(t)}) \leq \frac{\| \mathbf{E}_m^{(t)} U_m^{(t)} \|_F^2}{\Psi_{\mathbf{V}_{\text{Joint}}^{(t)}}(k) - \Phi_{\mathbf{W}_m^{(t)}}(k+1)}$$



Variation of the theoretical upper bound and observed Grassmannian distance between U_{Joint} and U_m with increase in iteration number t for LGG data set.

Experimental Results and Data Sets

Multi-Omics Cancer Data Sets:

CESC, LGG, BRCA, OV, STAD, CRC, LUNG, and KIDNEY

Views: mDNA, RNA, miRNA, RPPA Clusters: 2, 3, 4

Benchmark Data Sets:

Multi-source News Article Clustering: 3Sources and BBC

Views: Reuters, BBC News, The Guardian Clusters: 5, 6

Image Data Sets: Digits and 100Leaves

Views: Haralick features, Fourier coefficients

Clusters: 10, 100, 100

Ablation Study on Benchmark Data Sets

Module	Accuracy	NMI	ARI	F-measure	Rand	Purity
Best view	0.7096(7.7e-4)	0.6443(3.9e-4)	0.5416(9.2e-4)	0.7206(6.9e-4)	0.9173(2.1e-4)	0.7096(7.7e-4)
U_{Joint}	0.7735(0.0)	0.7256($\rightarrow 0$)	0.6401(0.0)	0.7843(0.0)	0.9347(0.0)	0.7775(0.0)
U_m	0.9055(0.0)	0.8392($\rightarrow 0$)	0.8064(0.0)	0.9067(0.0)	0.9650(0.0)	0.9055(0.0)
$d_\theta(U_i, U_j)$	0.9315(0.0)	0.8668($\rightarrow 0$)	0.8543(0.0)	0.9318(0.0)	0.9738(0.0)	0.9315(0.0)
L_m	0.9060(0.0)	0.8395($\rightarrow 0$)	0.8072(0.0)	0.9072(0.0)	0.9652(0.0)	0.9060 (0.0)
α_{Equal}	0.8965(0.0)	0.8363($\rightarrow 0$)	0.7973(0.0)	0.8971(0.0)	0.9634(0.0)	0.8965(0.0)
$\alpha^{(0)}_{\text{Eigen}}$	0.9055(0.0)	0.8385($\rightarrow 0$)	0.8060(0.0)	0.9067(0.0)	0.9650(0.0)	0.9055(0.0)
GeARS	0.9321 (3.1e-4)	0.8683 (5.5e-4)	0.8555 (6.2e-4)	0.9325 (3.2e-4)	0.9740 (1.1e-4)	0.9321 (3.1e-4)
<hr/>						
Best view	0.7159(0.0)	0.6390(0.0)	0.6082(0.0)	0.7656(0.0)	0.8624(0.0)	0.7869(0.0)
U_{Joint}	0.6343(1.3e-2)	0.5535(7.3e-3)	0.4432(1.1e-2)	0.6844(1.0e-2)	0.8136(2.8e-3)	0.7236(9.2e-3)
U_m	0.7165(4.3e-2)	0.6292(2.2e-2)	0.5635(4.9e-2)	0.7560(3.4e-2)	0.8455(1.9e-2)	0.7621(1.9e-2)
$d_\theta(U_i, U_j)$	0.7526(3.3e-2)	0.6615(3.7e-2)	0.6176(6.7e-2)	0.7855(3.8e-2)	0.8627(2.5e-2)	0.7881(3.3e-2)
L_m	0.7526(3.3e-2)	0.6590(3.4e-2)	0.6149(6.4e-2)	0.7838(3.7e-2)	0.5464(2.4e-2)	0.7881(3.3e-2)
α_{Equal}	0.7289(4.93e-2)	0.6434(3.44e-2)	0.5832(5.57e-2)	0.7629(3.78e-2)	0.8509(2.11e-2)	0.7745(2.53e-2)
$\alpha^{(0)}_{\text{Eigen}}$	0.7526(3.3e-2)	0.6590(3.4e-2)	0.6149(6.4e-2)	0.7838(3.7e-2)	0.8617(2.4e-2)	0.7881(3.3e-2)
GeARS	0.7786 (7.4e-3)	0.6823 (1.0e-2)	0.6591 (2.0e-2)	0.8063 (1.0e-2)	0.8770 (6.6e-3)	0.8142 (7.4e-3)

- Best performance when optimizing clustering subspaces, graph connectivity, and graph weights

Ablation Study on Omics Data Sets

Module	Accuracy	NMI	ARI	F-measure	Rand	Purity
Best view	0.6497(0.0)	0.3748 (→ 0)	0.3548(0.0)	0.6444(0.0)	0.7536(0.0)	0.6497(0.0)
U_{Joint}	0.6556(0.0)	0.3297(→ 0)	0.3109(0.0)	0.6566(0.0)	0.7409(0.0)	0.6556(0.0)
U_m	0.6766(0.0)	0.3360(0.0)	0.3369(0.0)	0.6775(0.0)	0.7484(0.0)	0.6766(0.0)
$d_\theta(U_i, U_j)$	0.5889(8.0e-2)	0.3075(4.5e-2)	0.2684(7.6e-2)	0.5963(7.6e-2)	0.7208(3.0e-2)	0.6035(7.0e-2)
L_m	0.6485(6.5e-2)	0.3257(3.2e-2)	0.3152(5.7e-2)	0.6510(6.1e-2)	0.7397(2.3e-2)	0.6526(5.6e-2)
α_{Equal}	0.5598(0.0)	0.2588(→ 0)	0.2155(0.0)	0.5691(0.0)	0.7030(0.0)	0.5628(0.0)
$\alpha^{(0)}_{\text{Eigen}}$	0.5658(0.0)	0.2673(→ 0)	0.2443(0.0)	0.5699(0.0)	0.7162(0.0)	0.5658(0.0)
GeARS	0.7023 (2.8e-3)	0.3687(1.7e-3)	0.3735 (4.8e-3)	0.7035 (3.0e-3)	0.7621 (2.2e-3)	0.7023 (2.8e-3)
Best view	0.8352(0.0)	0.5734(→ 0)	0.5567(0.0)	0.8269(0.0)	0.7861(0.0)	0.8352(0.0)
U_{Joint}	0.6292(0.0)	0.4106(→ 0)	0.2765(0.0)	0.6316(0.0)	0.6590(0.0)	0.6741(0.0)
U_m	0.8764(0.0)	0.6502(→ 0)	0.6328(0.0)	0.8742(0.0)	0.8194(0.0)	0.8764(0.0)
$d_\theta(U_i, U_j)$	0.9812(0.0)	0.9001(→ 0)	0.9449(0.0)	0.9812(0.0)	0.9740(0.0)	0.9812(0.0)
L_m	0.9565(3.9e-2)	0.8406(9.5e-2)	0.8689(1.2e-1)	0.9561(4.0e-2)	0.9367(6.0e-2)	0.9565(3.9e-2)
α_{Equal}	0.9101(0.0)	0.7167(→ 0)	0.7541(0.0)	0.9107(0.0)	0.8834(0.0)	0.9101(0.0)
$\alpha^{(0)}_{\text{Eigen}}$	0.9850(0.0)	0.9189(→ 0)	0.9527(0.0)	0.9850(0.0)	0.9776(0.0)	0.9850(0.0)
GeARS	0.9887 (0.0)	0.9397 (→ 0)	0.9655 (0.0)	0.9887 (0.0)	0.9837 (0.0)	0.9887 (0.0)

- Best performance when optimizing clustering subspaces, graph connectivity, and graph weights

Comparative Performance on Benchmark Data Sets

Algorithm	Accuracy	NMI	ARI	F-measure
MKC	0.6034(1.10e-1)	0.4786(8.51e-2)	0.3450(1.21e-1)	0.5018(9.03e-2)
CoregSC	0.4701(0.0)	0.2863(0.0)	0.2727(0.0)	0.4879(0.0)
MSC	0.6732(4.94e-2)	0.5531(1.44e-2)	0.4658(2.20e-2)	0.5877(1.83e-2)
ASMV	0.3372(0.0)	0.0348(0.0)	0.0018(\rightarrow 0)	0.3781(0.0)
MGL	BBC	0.5396(1.10e-1)	0.3697(1.89e-1)	0.3153(1.66e-1)
MCGL		0.3533(\rightarrow 0)	0.0741(\rightarrow 0)	0.0053(\rightarrow 0)
CoALA		0.8108(4.36e-3)	0.6536(1.96e-2)	0.7102(2.78e-2)
GMC		0.6934(\rightarrow 0)	0.5628(0.0)	0.4789(\rightarrow 0)
MiMIC		0.8715(0.0)	0.7182(\rightarrow 0)	0.7273(0.0)
GeARS		0.8804(1.74e-3)	0.7335(5.84e-3)	0.7566(8.23e-3)
MKC		0.0100(0.0)	0.0000(0.0)	0.0186(0.0)
CoregSC		0.7706(2.58e-2)	0.9165(5.90e-3)	0.7229(1.92e-2)
MSC		0.7379(2.21e-2)	0.9014(7.60e-3)	0.6788(2.26e-2)
ASMV		0.7906(\rightarrow 0)	0.9009(\rightarrow 0)	0.6104(\rightarrow 0)
MGL	100Leaves	0.6904(2.42e-2)	0.8753(7.60e-3)	0.3858(5.65e-2)
MCGL		0.8106(\rightarrow 0)	0.9130(0.0)	0.5155(\rightarrow 0)
CoALA		0.7384(1.34e-2)	0.8893(4.06e-3)	0.6550(1.41e-2)
GMC		0.8238(\rightarrow 0)	0.9292(0.0)	0.4974(\rightarrow 0)
MiMIC		0.8185(1.56e-2)	0.9302(4.12e-3)	0.7431(2.53e-2)
GeARS		0.8372(1.79e-2)	0.9346(3.71e-3)	0.7643(1.97e-2)

Comparative Performance on Benchmark Data Sets

Algorithm	Accuracy	NMI	ARI	F-measure
MKC	0.4924(2.77e-1)	0.5325(3.68e-1)	0.4280(2.99e-1)	0.5130(2.33e-2)
CoregSC	0.7556(5.96e-2)	0.7421(3.27e-2)	0.6885(5.73e-2)	0.6934(5.11e-2)
MSC	0.7918(8.21e-2)	0.7560(3.24e-2)	0.6803(6.28e-2)	0.7129(5.58e-2)
ASMV	0.5745($\rightarrow 0$)	0.6709($\rightarrow 0$)	0.4047($\rightarrow 0$)	0.4852($\rightarrow 0$)
MGL	Digits	0.7440(8.19e-2)	0.8264(4.73e-2)	0.6888(1.07e-1)
MCGL		0.8530(0.0)	0.9055 (0.0)	0.8313($\rightarrow 0$)
CoALa		0.8835(0.0)	0.7981($\rightarrow 0$)	0.7645(0.0)
GMC		0.8820($\rightarrow 0$)	0.9050($\rightarrow 0$)	0.8502($\rightarrow 0$)
MiMIC		0.9207(4.21e-4)	0.8597(4.88e-4)	0.8352(8.18e-4)
GeARS		0.9321 (3.16e-4)	0.8683(5.59e-4)	0.9325 (3.21e-4)
MKC	3Sources	0.4663(1.06e-1)	0.3665(1.00e-1)	0.2461(1.40e-1)
CoregSC		0.5479(2.99e-2)	0.5238(1.98e-2)	0.3339(2.85e-2)
MSC		0.4751(2.97e-2)	0.3850(2.27e-2)	0.2618(3.81e-2)
ASMV		0.3373($\rightarrow 0$)	0.0896($\rightarrow 0$)	-0.021($\rightarrow 0$)
MGL		0.6751(6.67e-2)	0.5768(8.61e-2)	0.4431(1.17e-1)
MCGL		0.3077($\rightarrow 0$)	0.1034($\rightarrow 0$)	-0.033($\rightarrow 0$)
CoALa		0.6508(0.0)	0.6198($\rightarrow 0$)	0.5183(0.0)
GMC		0.6923($\rightarrow 0$)	0.6216(0.0)	0.4431(0.0)
MiMIC		0.7360(5.92e-2)	0.6433(3.59e-2)	0.5957(6.69e-2)
GeARS		0.7786 (7.48e-3)	0.6823 (1.06e-2)	0.6591 (2.09e-2)

- **GeARS algorithm gives best performance on all benchmark data sets across all measures, except NMI measure on Digits data set.**
- **Outperforms the MiMIC algorithm that optimizes only the cluster indicator matrices**

Comparative Performance on Omics Data Sets

Algorithm	Accuracy	NMI	ARI	F-measure	Rand	Purity
COCA	0.5943(7.0e-3)	0.3131(1.2e-2)	0.2810(6.8e-3)	0.6068(4.2e-3)	0.7039(2.6e-3)	0.5943(7.0e-3)
NormS	0.6976(0.0)	0.4504(0.0)	0.4142(0.0)	0.6910(0.0)	0.7766(0.0)	0.6976(0.0)
LRACLuster	0.6287(0.0)	0.3745(\rightarrow 0)	0.2999(0.0)	0.6384(0.0)	0.7322(0.0)	0.6287(0.0)
iCluster	0.5089(0.0)	0.2249(\rightarrow 0)	0.2005(0.0)	0.4808(0.0)	0.6916(0.0)	0.5119(0.0)
PCA-con	0.6946(0.0)	0.4424(\rightarrow 0)	0.4068(0.0)	0.6868(0.0)	0.7734(0.0)	0.6946(0.0)
SURE	OV	0.7215 (0.0)	0.4680 (\rightarrow 0)	0.4372 (0.0)	0.7148 (0.0)	0.7857 (0.0)
JIVE	0.5718(7.7e-3)	0.2629(8.4e-3)	0.2027(4.2e-3)	0.5653(7.8e-3)	0.6885(2.8e-3)	0.5718(7.7e-3)
SNF	0.5269(0.0)	0.2753(0.0)	0.2058(0.0)	0.5642(0.0)	0.6557(0.0)	0.5389(0.0)
CoALA	0.6736(0.0)	0.3381(\rightarrow 0)	0.3199(0.0)	0.6700(0.0)	0.7379(0.0)	0.6736(0.0)
MiMIC	0.6595(2.8e-3)	0.3271(3.9e-4)	0.3112(4.2e-3)	0.6611(2.5e-3)	0.7383(1.9e-3)	0.6595(2.8e-3)
GeARS	0.7023(2.8e-3)	0.3687(1.7e-3)	0.3735(4.8e-3)	0.7035(3.0e-3)	0.7621(2.2e-3)	0.7023(2.8e-3)
COCA	0.6591(0.0)	0.2772(0.0)	0.2533(0.0)	0.6608(0.0)	0.6454(0.0)	0.6591(0.0)
NormS	0.7940(0.0)	0.5325(0.0)	0.4649(0.0)	0.7916(0.0)	0.7465(0.0)	0.7940(0.0)
LRACLuster	0.4719(0.0)	0.1240(\rightarrow 0)	0.1030(0.0)	0.5137(0.0)	0.5831(0.0)	0.5280(0.0)
iCluster	0.4382(0.0)	0.1379(\rightarrow 0)	0.0996(0.0)	0.5187(0.0)	0.5821(0.0)	0.5355(0.0)
PCA-con	LGG	0.6666(0.0)	0.3438(0.0)	0.3031(0.0)	0.6574(0.0)	0.6616(0.0)
SURE	LGG	0.7940(0.0)	0.5335(0.0)	0.4668(0.0)	0.7904(0.0)	0.7465(0.0)
JIVE	LGG	0.5617(0.0)	0.2299(\rightarrow 0)	0.1606(0.0)	0.5757(0.0)	0.6056(0.0)
SNF	0.8689(0.0)	0.6253(0.0)	0.6331(0.0)	0.8720(0.0)	0.8268(0.0)	0.8689(0.0)
CoALA	0.9737(0.0)	0.8689(\rightarrow 0)	0.9199(0.0)	0.9737(0.0)	0.9622(0.0)	0.9737(0.0)
MiMIC	0.9625(0.0)	0.8543(\rightarrow 0)	0.8790(0.0)	0.9623(0.0)	0.9424(0.0)	0.9625(0.0)
GeARS	0.9887 (0.0)	0.9397 (\rightarrow 0)	0.9655 (0.0)	0.9887 (0.0)	0.9837 (0.0)	0.9887 (0.0)

Comparative Performance on Omics Data Sets

Algorithm	Accuracy	NMI	ARI	F-measure	Rand	Purity
COCA	0.7434(7.9e-4)	0.5002(3.4e-4)	0.4864(4.5e-4)	0.7457(8.1e-4)	0.7905(1.9e-4)	0.7434(7.9e-4)
NormS	0.7688(0.0)	0.4287(\rightarrow 0)	0.5090(0.0)	0.7699(0.0)	0.7999(0.0)	0.7688(0.0)
LRACLuster	0.7110(0.0)	0.5437(\rightarrow 0)	0.4035(0.0)	0.7101(0.0)	0.7521(0.0)	0.7110(0.0)
iCluster	0.7638(0.0)	0.5176(\rightarrow 0)	0.4745(0.0)	0.7658(0.0)	0.7842(0.0)	0.7638(0.0)
PCA-con	BRCA	0.7587(0.0)	0.5506(\rightarrow 0)	0.5038(0.0)	0.7601(0.0)	0.7984(0.0)
SURE		0.7663(0.0)	0.4558(0.0)	0.5104(0.0)	0.7683(0.0)	0.8010(0.0)
JIVE		0.6859(0.0)	0.4368(0.0)	0.3772(0.0)	0.6889(0.0)	0.7464(0.0)
SNF		0.6783(0.0)	0.5528(\rightarrow 0)	0.4111(0.0)	0.6865(0.0)	0.7602(0.0)
CoALA		0.7613(0.0)	0.5281(\rightarrow 0)	0.4874(0.0)	0.7660(0.0)	0.7922(0.0)
MiMIC		0.7964(0.0)	0.5553(\rightarrow 0)	0.5474(0.0)	0.7997(0.0)	0.8152(0.0)
GeARS		0.7914(0.0)	0.5444(\rightarrow 0)	0.5376(0.0)	0.7936(0.0)	0.8104(0.0)
COCA	STAD	0.4450(3.3e-2)	0.1309(4.7e-3)	0.0740(1.0e-2)	0.4558(2.5e-2)	0.5981(1.3e-2)
NormS		0.5702(0.0)	0.1805(\rightarrow 0)	0.1625(0.0)	0.5770(0.0)	0.6435(0.0)
LRACLuster		0.4256(0.0)	0.1259(\rightarrow 0)	0.0912(0.0)	0.4746(0.0)	0.6122(0.0)
iCluster		0.3512(0.0)	0.0650(\rightarrow 0)	0.0288(0.0)	0.3832(0.0)	0.5855(0.0)
PCA-con		0.6900(0.0)	0.3654(0.0)	0.3204(0.0)	0.6959(0.0)	0.7110(0.0)
SURE		0.6983(0.0)	0.3511(\rightarrow 0)	0.3445(0.0)	0.7056(0.0)	0.7216(0.0)
JIVE		0.4049(0.0)	0.1288(\rightarrow 0)	0.0657(0.0)	0.4487(0.0)	0.5981(0.0)
SNF		0.5661(0.0)	0.4558(0.0)	0.1522(0.0)	0.5521(0.0)	0.6945(0.0)
CoALA		0.7685(0.0)	0.5107(0.0)	0.4559(0.0)	0.7778(0.0)	0.7661(0.0)
MiMIC		0.7727(0.0)	0.5220(\rightarrow 0)	0.4650(0.0)	0.7830(0.0)	0.7685(0.0)
GeARS		0.7933(0.0)	0.4970(\rightarrow 0)	0.5059(0.0)	0.7942(0.0)	0.7847(0.0)

- For LGG and STAD data sets, GeARS outperforms all the existing approaches except for the NMI measure on STAD data set
- SVD eigenspace based SURE algorithm has best performance for OV data set, MiMIC has that for BRCA data set.

Comparative Performance on Additional Benchmark Data Sets

Algorithms→	Graph Based		Manifold Based	
	SNF	CoALa	MiMIC	GeARS
Accuracy	0.6907(2.57e-2)	0.7715 (2.18e-2)	0.7307(2.36e-2)	0.8052 (3.06e-2)
NMI	0.8616(1.00e-2)	0.8980 (1.15e-2)	0.8814(1.35e-2)	0.9118 (1.73e-2)
ARI	0.6054(3.04e-2)	0.6932 (2.82e-2)	0.6208(3.83e-2)	0.7201 (4.22e-2)
F-measure	0.7257(2.44e-2)	0.7962 (1.78e-2)	0.7677(2.29e-2)	0.8297 (2.61e-2)
Rand	0.9804(2.04e-3)	0.9850 (1.63e-3)	0.9802(2.65e-3)	0.9864 (2.28e-3)
Purity	0.7450(2.26e-2)	0.8090 (1.75e-2)	0.7737(1.78e-2)	0.8315 (2.66e-2)
Accuracy	0.5440(3.42e-2)	0.5685(0.0)	0.5773(0.0)	0.5852 (2.31e-2)
NMI	0.5676(2.41e-2)	0.5650(→0)	0.5880 (→0)	0.5734(2.69e-2)
ARI	0.4126(2.86e-2)	0.4397(0.0)	0.4608 (0.0)	0.4582(3.65e-2)
F-measure	0.6363(4.12e-2)	0.6689 (0.0)	0.6600(0.0)	0.6761 (2.99e-2)
Rand	0.7482(1.13e-2)	0.7583(0.0)	0.7674 (0.0)	0.7666(1.49e-2)
Purity	0.8516(1.09e-2)	0.8548(0.0)	0.8751 (0.0)	0.8603(7.65e-3)
Accuracy	0.5450(2.79e-2)	0.5896 (3.41e-3)	0.6120 (2.46e-3)	0.5801(3.35e-2)
NMI	0.3829(1.14e-2)	0.4364(2.81e-3)	0.4686 (6.17e-3)	0.4416(1.64e-2)
ARI	0.2941(1.86e-2)	0.3256 (2.87e-3)	0.3479 (3.73e-3)	0.3079(4.17e-2)
F-measure	0.5957(1.96e-2)	0.5844(4.98e-3)	0.6373 (3.50e-3)	0.6140(1.96e-2)
Rand	0.7936 (9.31e-3)	0.7460(2.14e-3)	0.7709(9.35e-4)	0.7736 (9.66e-3)
Purity	0.6012(1.62e-2)	0.6206(3.41e-3)	0.6423 (2.46e-3)	0.6217(2.38e-2)

- ORL: face clustering; Caltech7: object recognition; CORA: author/paper citation network.
- Proposed GeARS has best performance for ORL face clustering data set and second best performance for Caltech7 and CORA after MiMIC algorithm.

Highlights

- Manifold optimization over **Grassmannian** and **SPD manifolds** for multi-view clustering.
- Simultaneous optimization of **cluster indicator subspaces**, **graph connectivity**, and **view weights** preserves better cluster structure.
- Matrix perturbation theory to **upper bound Grassmannian distance** between the joint and individual subspaces. Distance empirically minimizes as the algorithm converges.
- GeARS has superior performance in several data sets, compared to existing algorithms that optimize over Euclidean space or only a subset of variables.

CONCLUSION

NormS: Normality based rank, relevance, and redundancy estimation.

SURE: Incremental updation of SVD eigenspace from individual low-rank eigenspaces.

CoALa: Integration of approximate graph Laplacians.
Upper bound on gap between approximate and full-rank eigenspaces.

MiMIC: Manifold optimization over k-means and Steifel manifolds.
Asymptotic convergence bound to infer separability of clusters in data set.

GeARS: Simultaneous ioptimization of subspaces, graphs, and view weights over Grassmannian and SPD manifolds.
Grassmannian disagreement bound between joint and individual subspaces.

FUTURE DIRECTIONS

- Eigenspace model when data does not follow a mixture of Guassians
- Parallel computation of joint eigenspace
- Non-linear combination of graph Laplacians
- Tensor spectral clustering
- Incomplete views
- Views observed in heterogeneous measurement spaces
- Deep network based optimization

PUBLICATION

Published/Accepted:

- A. Khan and P. Maji, "Multi-Manifold Optimization for Multi-View Subspace Clustering," **IEEE Transactions on Neural Networks and Learning Systems (TNNLS)**, pp. 1-13, 2021. DOI:10.1109/TNNLS.2021.3054789.
- A. Khan and P. Maji, "Approximate Graph Laplacians for Multimodal Data Clustering," **IEEE Transactions on Pattern Analysis and Machine Intelligence(TPAMI)**, vol. 43, no. 3, pp. 798-813, 2021. DOI: 10.1109/TPAMI.2019.2945574.
- A. Khan and P. Maji, "Selective Update of Relevant Eigenspaces for Integrative Clustering of Multimodal Data," **IEEE Transactions on Cybernetics (TCYB)**, pp. 1-13, 2020. DOI: 10.1109/TCYB.2020.2990112.
- A. Khan and P. Maji, "Low-Rank Joint Subspace Construction for Cancer Subtype Discovery," **IEEE/ACM Transactions on Computational Biology and Bioinformatics (TCBB)**, vol. 17, no. 4, pp. 1290–1302, 2020. DOI: 10.1109/TCBB.2019.2894635.

Submitted:

- A. Khan and P. Maji, "Geometry Aware Multi-View Clustering over Riemannian Manifolds," **IEEE Transactions on Pattern Analysis and Machine Intelligence**, pages 1-13, 2021 (Manuscript ID: TPAMI-2021-08-1458).

REFERENCES

- [1] F. R. Chung, Spectral graph theory. American Mathematical Society, 1997, no. 92.
- [2] A. Y. Ng, M. I. Jordan, and Y. Weiss, “On spectral clustering: Analysis and an algorithm,” in Advances in neural information processing systems, 2002, pp. 849–856.
- [3] G. W. Stewart and J.-g. Sun, Matrix perturbation theory. Academic press New York, 1990.
- [4] C. Davis and W. Kahan, “The rotation of eigenvectors by a perturbation. III,” SIAM Journal on Numerical Analysis, vol. 7, no. 1, pp. 1–46, 1970.
- [5] S. Ji-guang, “Perturbation of angles between linear subspaces,” Journal of Computational Mathematics, vol. 5, no. 1, pp. 58–61, 1987.
- [6] D. Wu, D. Wang, M. Q. Zhang, and J. Gu, “Fast dimension reduction and integrative clustering of multi-omics data using low-rank approximation: application to cancer molecular classification,” BMC genomics, vol. 16, no. 1, p. 1022, 2015.
- [7] E. F. Lock, K. A. Hoadley, J. S. Marron, and A. B. Nobel, “Joint and individual variation explained (jive) for integrated analysis of multiple data types,” The annals of applied statistics, vol. 7, no. 1, pp. 523–542, 2013.
- [8] Q. Feng, M. Jiang, J. Hannig, and J. Marron, “Angle-based joint and individual variation explained,” Journal of Multivariate Analysis, vol. 166, pp. 241 – 265, 2018.
- [9] R. Shen, A. B. Olshen, and M. Ladanyi, “Integrative clustering of multiple genomic data types using a joint latent variable model with application to breast and lung cancer subtype analysis,” Bioinformatics, vol. 25, no. 22, pp. 2906–2912, 2009.
- [10] B. Wang et al., “Similarity network fusion for aggregating data types on a genomic scale,” Nature methods, vol. 11, no. 3, pp. 333–337, 2014.

THANK YOU