Oak Ridge National Laboratory

Matrix and Tensor Factorization with Scientific Constraints

Ramakrishnan(Ramki) Kannan

ORNL is managed by UT-Battelle, LLC for the US Department of Energy

U.S. DEPARTMENT OF ENERGY

Acknowledgements

This manuscript has been co-authored by UT-Battelle, LLC under Contract No. DE-AC05-00OR22725 with the U.S. Department of Energy. This project was partially funded by the Laboratory Director's Research and Development fund. This research used resources of the Oak Ridge Leadership Computing Facility at the Oak Ridge National Laboratory, which is supported by the Office of Science of the U.S. Department of Energy.
This research used resources of the National Energy Research Scientific Computing Center, a DOE Office of Science User Facility supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231

Scientific Engagement Model

Investigation

- Small Scientific Data w/ Ground Truths
- Bunch of Baselines with comparison ML metrics (Jupyter Notebook) (~ 6 months)
- Half-a-day Hackathon
- Feedback from Domain Scientists
- Add Domain constraints to Baselines (~ 2months)

Productivity

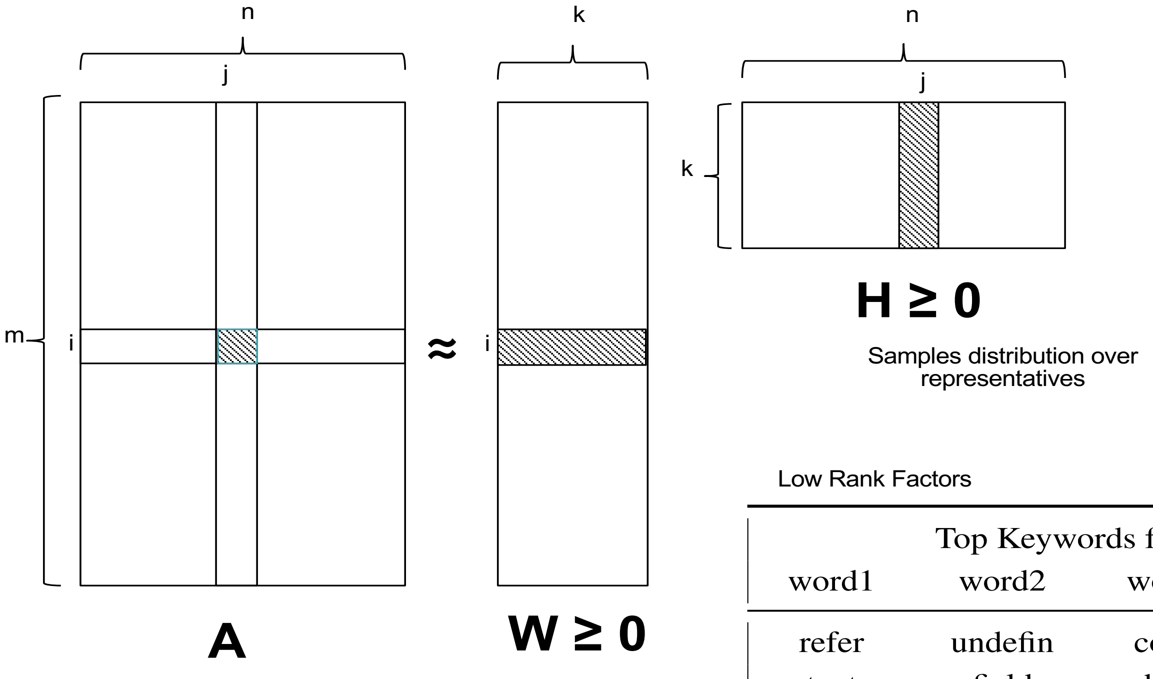
- Publications (~ 1 year)
  - Scientific
  - DM/ML
  - HPC
- Funding Opportunities
  - ASCR
  - LDRD
  - SEED
  - Sponsors
- Software (~ 3 mnths)
  - Pycroscopy
  - BEAM
  - DML
- Scalable Implementation
  - Algorithmic Framework Development with Theoretical Analysis
  - Implementation of the framework and the baselines w/ constraints
  - Benchmarking and OLCF Deployment (~ 3 months)

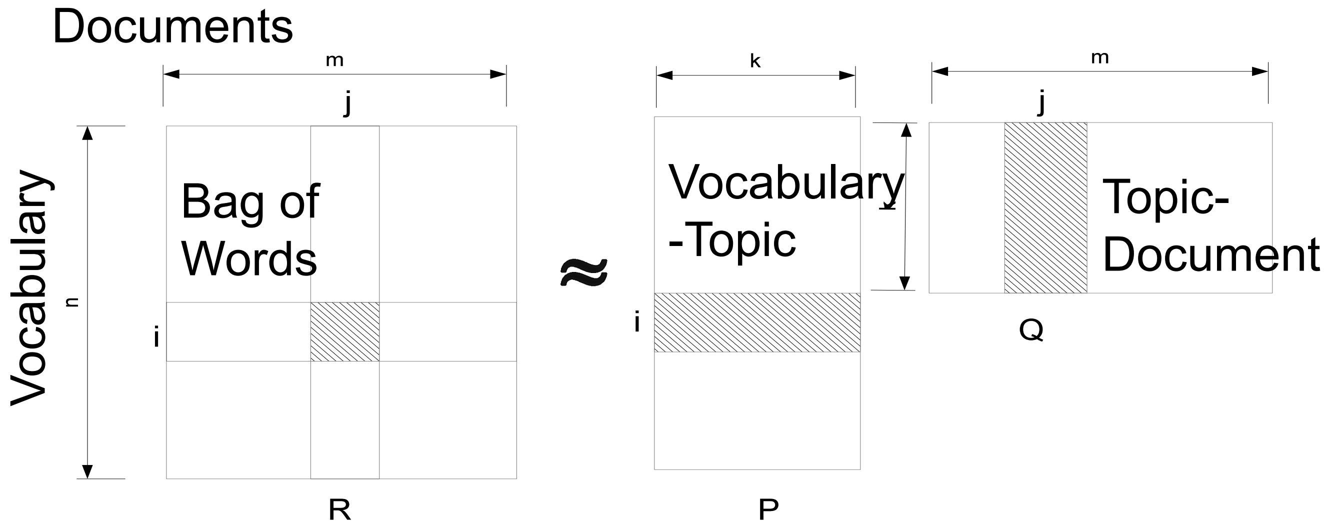
$\underbrace{\text{Algorithmic Framework Development with Theoretical Analysis; Implementation of the framework and the baselines w/ constraints}}_{\sim\ 10\ \text{months}}$

Performance

OAK RIDGE National Laboratory | OAK RIDGE LEADERSHIP COMPUTING FACILITY

NMF and Applications



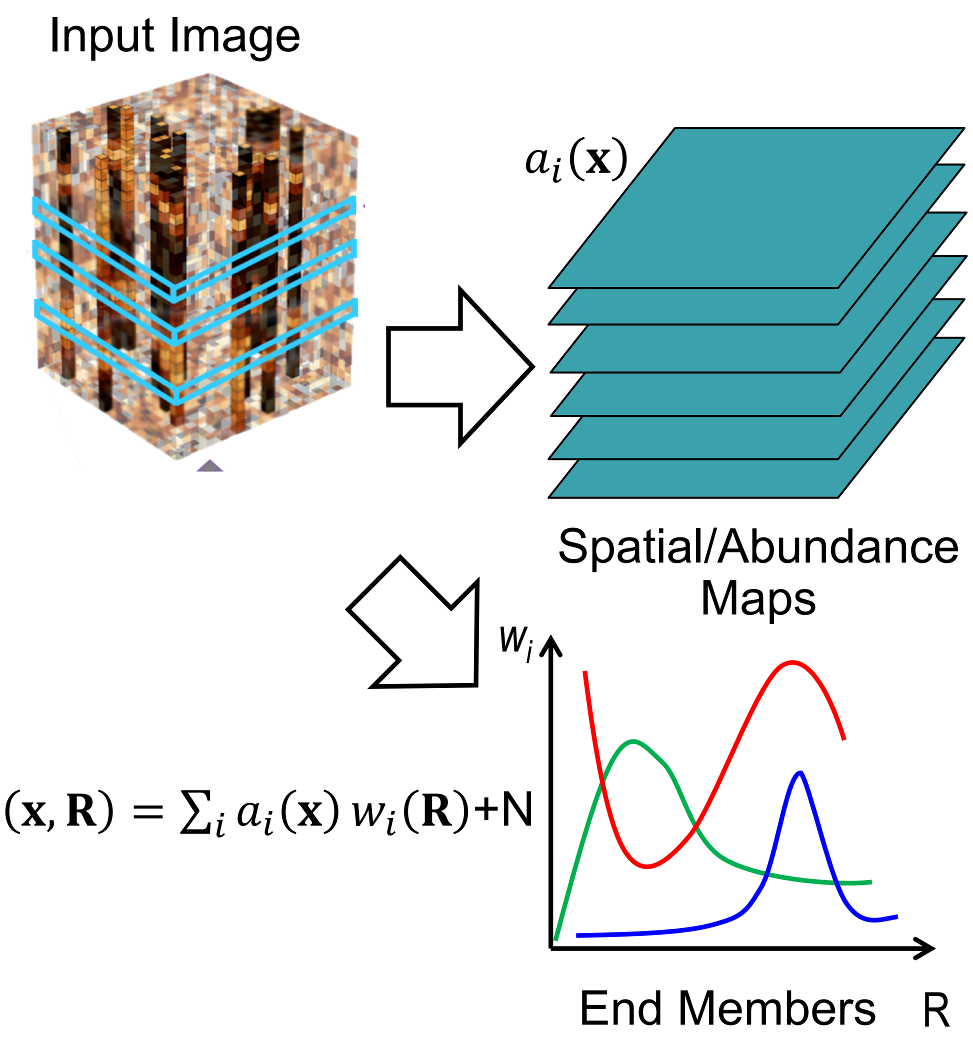


Top Keywords from Topics 1-25					Top Keywords from Topics 26-50				
word1	word2	word3	word4	word5	word1	word2	word3	word4	word5
refer	undefin	const	key	compil	echo	type=text	php	form	result
text	field	box	word	static	test	perform	fail	unit	result
imag	src	descript	alt=ent	size	tabl	key	queri	databas	insert
button	click	event	form	add	user	email	usernam	login	log
creat	bean	add	databas	except	data	json	store	read	databas
string	static	final	catch	url	page	load	content	url	link
width	height	color	left	display	privat	static	final	import	float
app	applic	servic	thread	work	row	column	date	cell	valu
ipsum	lorem	dolor	sit	amet	line	import	command	print	recent
node	list	root	err	element	var	map	marker	match	url
0x00	0xff	byte	0x01	0xc0	server	connect	client	messag	request
file	directori	read	open	upload	number	byte	size	print	input

Ramakrishnan Kannan, Grey Ballard, Haesun Park: MPI-FAUN: An MPI-Based Framework for Alternating-Updating Nonnegative Matrix Factorization. IEEE Trans. Knowl. Data Eng.30(3): 544-558 (2018)

Motivation

- Understanding terrestrial information in an unknown place from satellite images
- Identifying presence of hidden unknown/foreign bodies in a scanned image - Eg., contamination in food articles, camouflaged explosives etc.
- Biological application - spectral karyotyping, immunofluorescence, live-cell imaging, drug discovery, and tissue pathology – Eg., Unmixing on Spectral imaging of the stained tissues using multiple dyes.
- Physics and Material Sciences – Mapping properties to end-members. Comparing different materials



$$S(\mathbf{x},\mathbf{R}) = \sum_i a_i(\mathbf{x})\, w_i(\mathbf{R}) + \mathrm{N}$$

MPI-FAUN

- Distributed Communication avoiding NMF Algorithms
- https://github.com/ramkikannan/nmflibrary
- http://dx.doi.org/10.1109/TKDE.2017.2767592

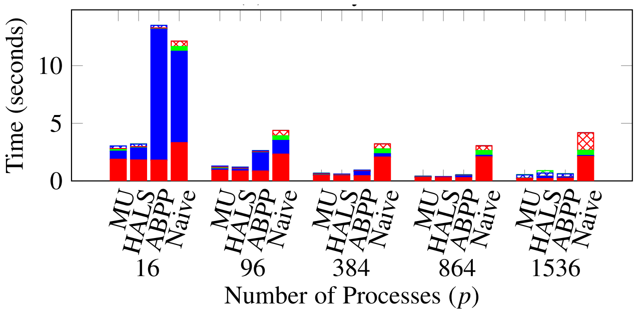
Titan – Dense Matrix, Low Rank 50, 100 Iterations, 12650 Nodes, 202500 Cores,

Matrix Size	Algos	NMF Time (in Secs)
3.03 million x 3.03 million	MU	554
	HALS	197.75
	ANLS/BPP	219.8

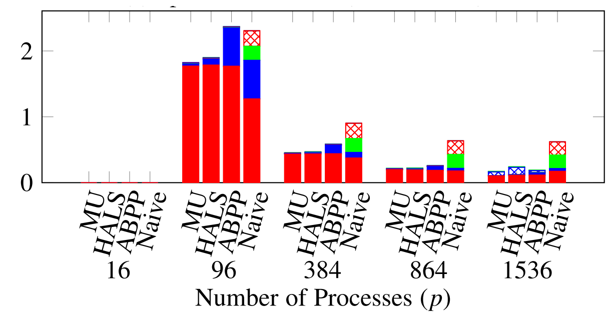
Rhea, 100 nodes, 1600 cores, Low Rank 50,

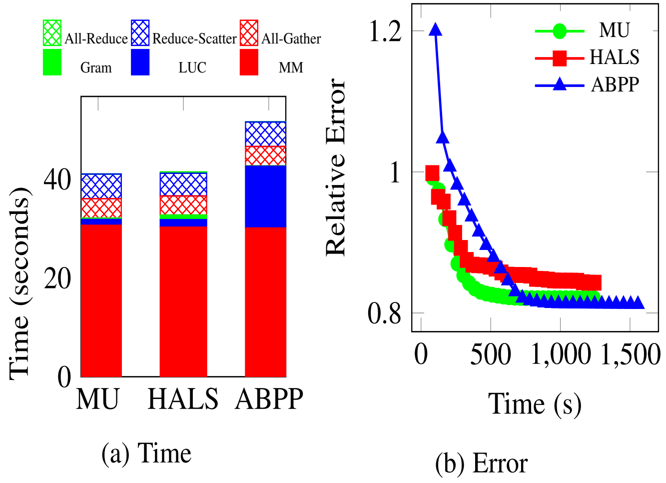
Dataset	Type	Matrix size	NMF Time
Video	Dense	1 Million x 13,824	5.73 seconds
Stack Exchange	Sparse	627,047 x 12 Million	67 seconds
Webbase-2001	Sparse	118 Million x 118 Million	25 minutes

Sparse Webbase – 1 Million Vertex Graph



Dense Real world – Video

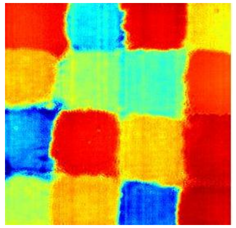




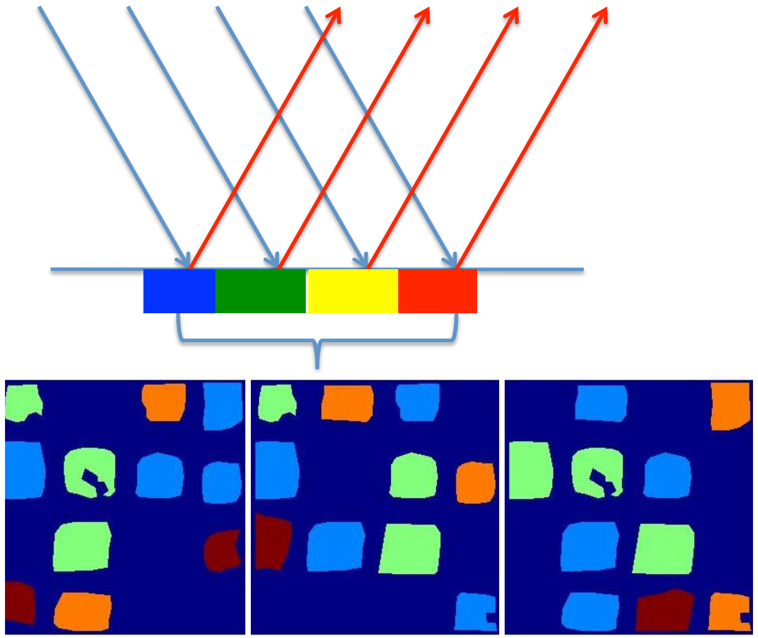
NMF on118 million Web-graph

Existing Approach : Linear Unmixing

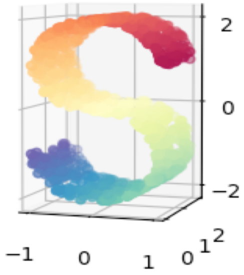
1. Good at Capturing Macroscopic Information
2. Spatially segregated patterns

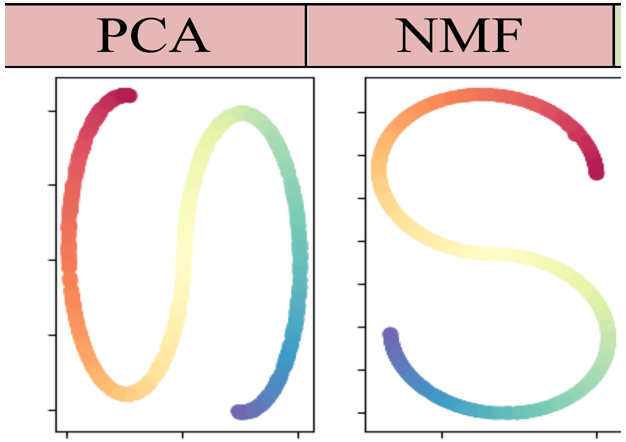


E_0 $\quad E_0 e^{-\alpha_1}$ .... $E_0 e^{-\alpha_N}$ $\Rightarrow$ $E_S = E_0 \sum_{k=1}^{N} e^{-\alpha_k}$



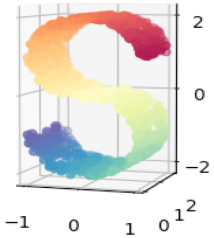
Spatial Maps

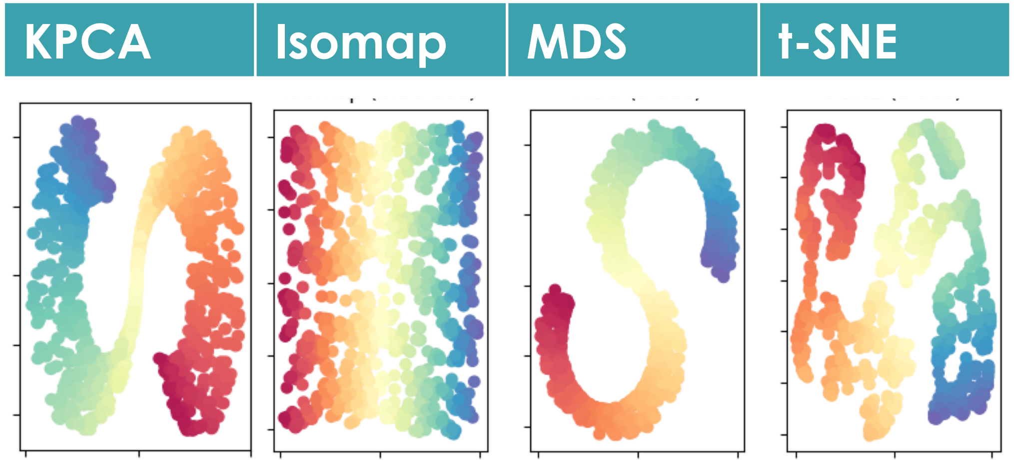


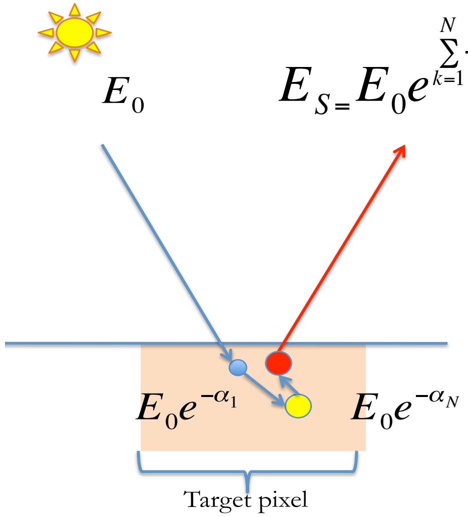


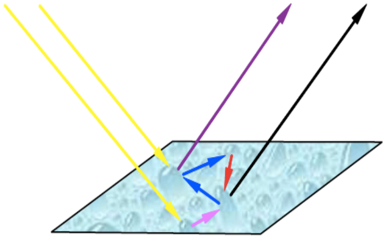
http://spectronet.de/story_docs/vortraege_2013/130306_ocm/130307_13_gro%C3%9F_fhg_iosb.pdf
http://www.sfpt.fr/hyperspectral/wp-content/uploads/2013/01/cours_Licciardi.pdf

Existing Non-linear Unmixing (NLUM)









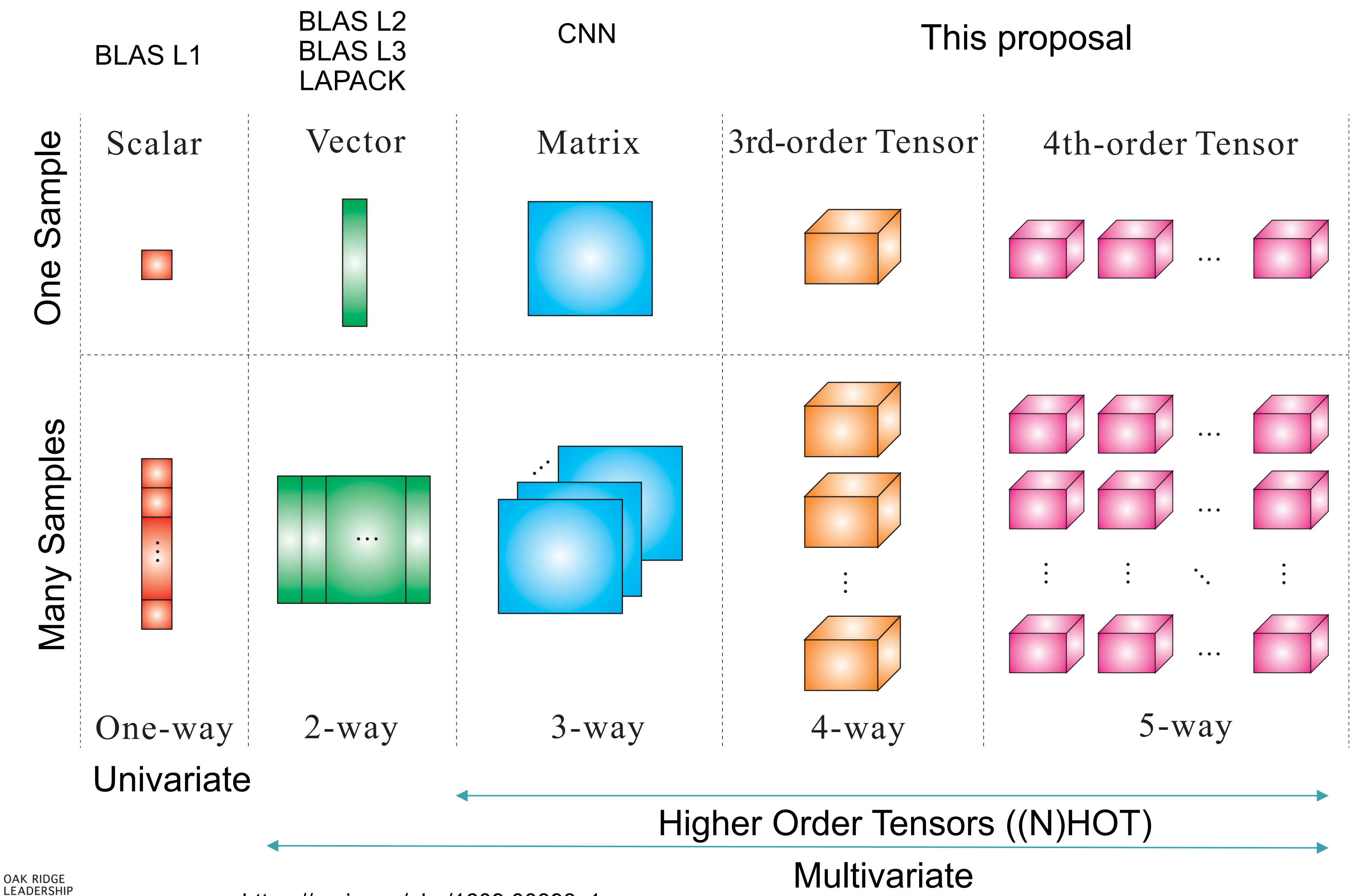
Non-Meaningful results for following reasons:

1. End-members and abundance maps are negative
2. Too many end-members participate in a particular location
3. Similar end-members and not distinctive enough
4. Ratio of end-members are not correct
5. Rotated end-members

Solution:
NLUM w/ Physical Constraints such as non-negativity, sparsity, spatial smoothness, sum to 1, orthogonal etc.

http://spectronet.de/story_docs/vortraege_2013/130306_ocm/130307_13_gro%C3%9F_fhg_iosb.pdf
http://www.sfpt.fr/hyperspectral/wp-content/uploads/2013/01/cours_Licciardi.pdf

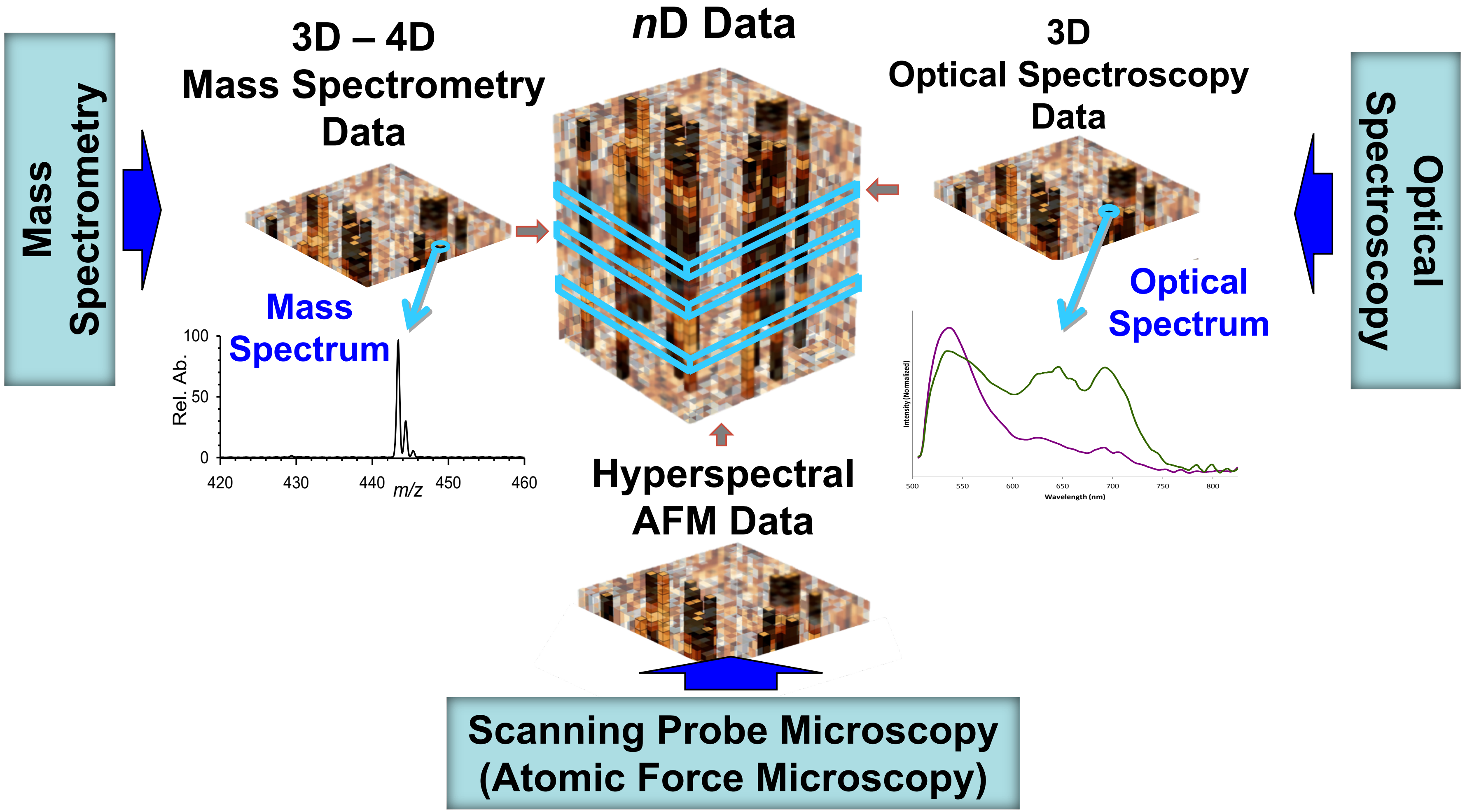
Higher Order Tensors



https://arxiv.org/abs/1609.00893v1

Dimensionality Reduction in Scientific Data

- Multimodal characterization of materials – *comprehensive characterization from chemical composition to functional properties on the nanoscale*



Thanks Anton Ievlev


рахмат
danke
謝謝
ngiyabonga
شكراً جزيلاً
Баярлалаа
mersi
kia ora
barka
welalin
tack
spas
teşekkür ederim
спасибо
faafetai lava
vinaka
спасибі
blagodaram
dank je
misaotra
matondo
paldies
grazzi
mahalo
tapadh leat
kiitos
dankie
nanni
nandri
dhanyavad
mauruuru
thank you
gracias
хвала
asante
manana
bedankt
bayarlalaa
gràçie
hvala
köszönöm
akun
dankon
aciū
obrigada
murakoze
tenki
enkosi
djiere dieuf
tau
mochchakkeram
дякую
mamnun
dziękuję
chnorakaloutioun
gratias ago
gràcies
sulpáy
go raibh maith agat
chokrane
sobodi
děkuji
sagolun
sukriya
kop khun krap
taiku
arigatô
takk
dakujem
trugarez
obrigado
mèsi
didi madloba
ありがとう
tanemirt
grazie
najis tuke
terima kasih
rahmet
kam sah hamnida
rahmat
ευχαριστώ
diolch
dhanyavadagalu
shukriya
mercé
мерси
তোমাকে ধন্যবাদ
감사합니다
xièxie
merci