Award #: 1642385

High Performance Low Rank Approximation for Scalable Data Analytics

Grey Ballard, Srinivas Eswar, Koby Hayashi, Ramakrishnan Kannan, Haesun Park
Wake Forest University and Georgia Institute of Technology

Constrained Low Rank Approximation for modeling key data analytics problems of scientific unmixing and inverse problems

Problem focus:

- Nonnegative Matrix Factorization (NMF)

$$\min_{W\geq 0, H\geq 0} ||A - WH||_F$$

- Nonnegative Tensor Factorization (NTF)

$$\min_{H^{(n)}\geq 0} ||\boldsymbol{A} - \sum_r H_r^{(1)} \circ \cdots \circ H_r^{(N)}||_F$$

- Variants: Symmetric NMF, Hierarchical NMF, NMF with regularization

Software goals:

- Utilize advances in numerical linear algebra algorithms and software
- Allow for algorithmic flexibility
- Facilitate design of MPI-based parallelization for scalable solutions

Alternating-Updating Algorithms

Require: **A** is an $m \times n$ matrix, k is rank of approximation

1: Initialize **H** with a non-negative matrix
2: **while** stopping criteria not satisfied **do**
3: Update **W** using $\mathbf{HH}^T$ and $\mathbf{AH}^T$
4: Update **H** using $\mathbf{W}^T\mathbf{W}$ and $\mathbf{W}^T\mathbf{A}$
5: **end while**

Block Principal Pivoting (BPP)

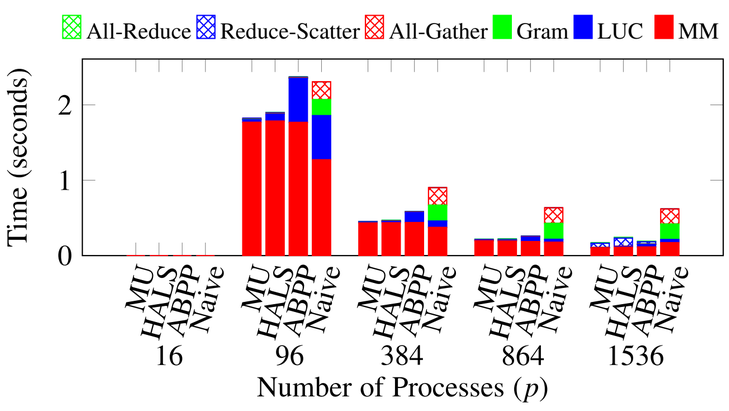
$$\mathbf{W} \leftarrow \underset{\tilde{\mathbf{W}}\geqslant 0}{\operatorname{argmin}} \left\|\mathbf{A} - \tilde{\mathbf{W}}\mathbf{H}\right\|_F, \quad \mathbf{H} \leftarrow \underset{\tilde{\mathbf{H}}\geqslant 0}{\operatorname{argmin}} \left\|\mathbf{A} - \mathbf{W}\tilde{\mathbf{H}}\right\|_F.$$

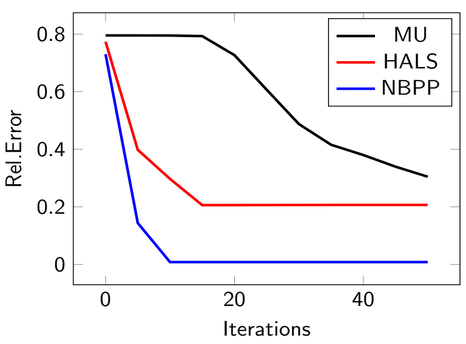
Hierarchical Alternating Least Squares (HALS)

$$\mathbf{w}^i \leftarrow \left[\mathbf{w}^i + \frac{(\mathbf{AH}^T)^i - \mathbf{W}(\mathbf{HH}^T)^i}{(\mathbf{HH}^T)_{ii}}\right]_+ \quad \mathbf{h}_i \leftarrow \left[\mathbf{h}_i + \frac{(\mathbf{W}^T\mathbf{A})_i - (\mathbf{W}^T\mathbf{W})_i\mathbf{H}}{(\mathbf{W}^T\mathbf{W})_{ii}}\right]_+$$

Multiplicative Update (MU)

$$w_{ij} \leftarrow w_{ij}\frac{(\mathbf{AH}^T)_{ij}}{(\mathbf{WHH}^T)_{ij}} \quad h_{ij} \leftarrow h_{ij}\frac{(\mathbf{W}^T\mathbf{A})_{ij}}{(\mathbf{W}^T\mathbf{WH})_{ij}}$$





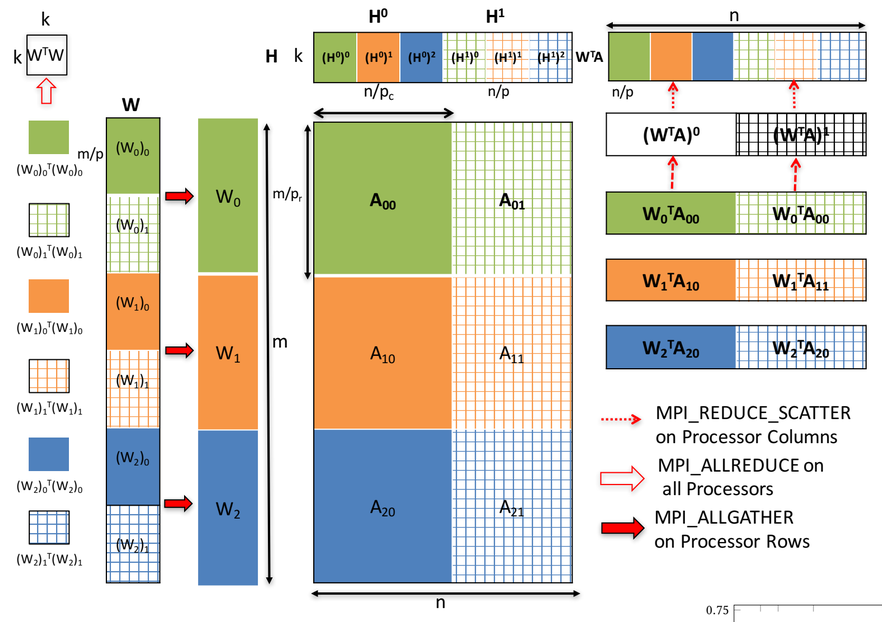
Parallel Nonnegative Matrix Factorization

- Parallelization strategy is independent of algorithm for enforcing nonnegativity
- Communication only of factor matrices
- Dense and sparse data yield different communication patterns

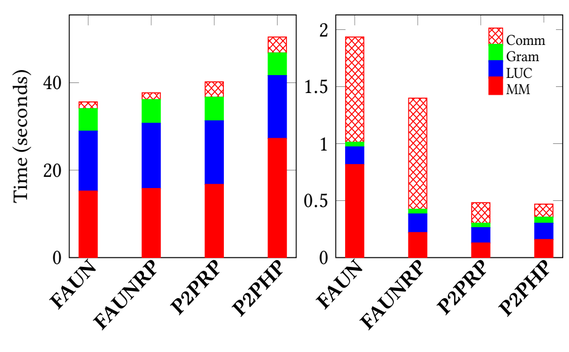
Top Keywords from Topics 1-25					Top Keywords from Topics 26-50				
word1	word2	word3	word4	word5	word1	word2	word3	word4	word5
refer	undefin	const	key	compil	echo	type=text	php	form	result
text	field	box	word	static	test	perform	fail	unit	result
imag	src	descript	alt=ent	size	tabl	key	queri	databas	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
function	call	event	work	variabl	object	properti	json	instanc	list
int	char	const	static	doubl	array	element	valu	key	index
public	overrid	virtual	static	extend	main	thread	program	frame	cout
return	param	result	def	boolean	type	field	properti	argument	resolv
info	thread	start	map	servic	select	item	queri	join	list
error	syntax	found	symbol	fail	sourc	target	except	java	fail
set	properti	virtual	default	updat	instal	version	packag	err	default
case	break	switch	default	cout	code	work	problem	chang	write
method	call	except	static	todo	void	overrid	protect	catch	extend
href	nofollow	src	link	work	true	requir	boolean	option	valid
end	def	dim	begin	properti	find	project	import	warn	referenc
debug	request	filter	match	found	view	control	item	overrid	posit
fals	boolean	fix	bool	autoincr	null	default	key	int(11	primari

TABLE 1
Top 5 words of 50 topics from *Stack Exchange* data set.

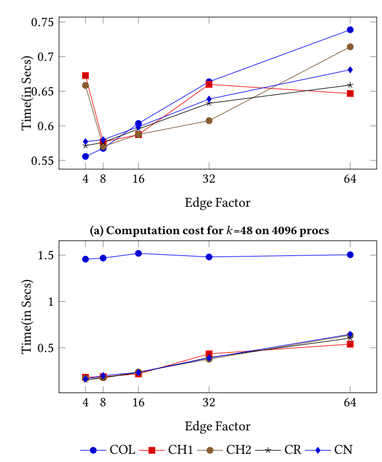
Results for Stack Exchange bag-of-words data
630K words x 12M docs with 365M nonzeroes



Sparse NMF



(a) Rhea $p = 16$ for $k = 48$ (b) Rhea $p = 3072$ for $k = 48$



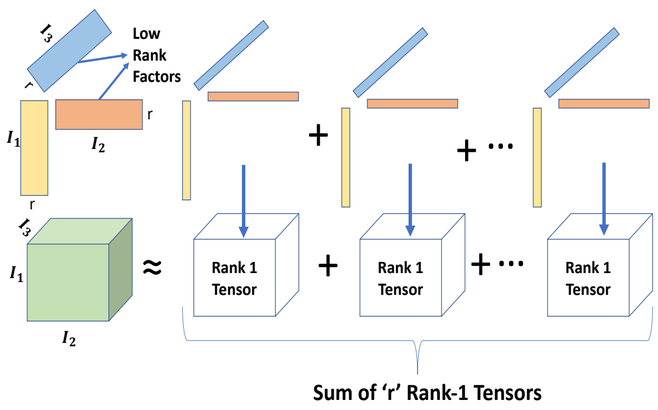
(a) Computation cost for k=48 on 4096 procs

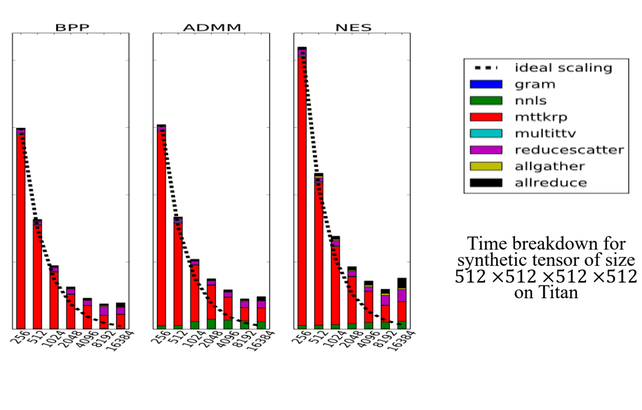
(b) Communication cost for k=48 on 4096 procs

Figure 5: Flickr dataset Time Breakdown for $k = 48$ on Rhea

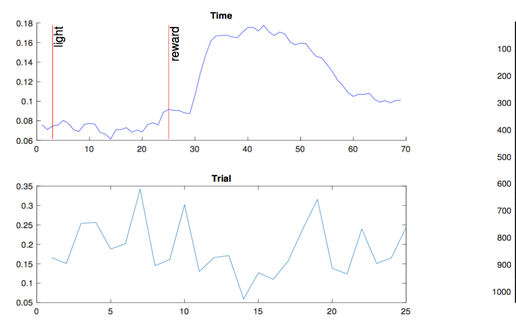
Figure 2: Communication and computation costs of different partitioning strategies on Kronecker graphs on 4096 processors Rhea for $k = 48$

Parallel Nonnegative (CP) Tensor Factorization

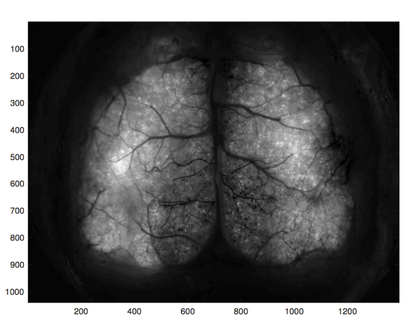




Time breakdown for synthetic tensor of size 512 ×512 ×512 ×512 on Titan



(a) Time and trial factors



(b) Brain image factor

Component of NNCP of Mouse Brain Imaging Tensor
1.5M x 69 x 25 (20GB)

PLANC Software: Parallel Low Rank Approximations with Non-negativity Constraints
https://github.com/ramkikannan/nmflibrary

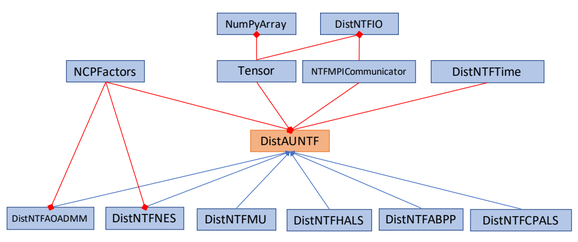
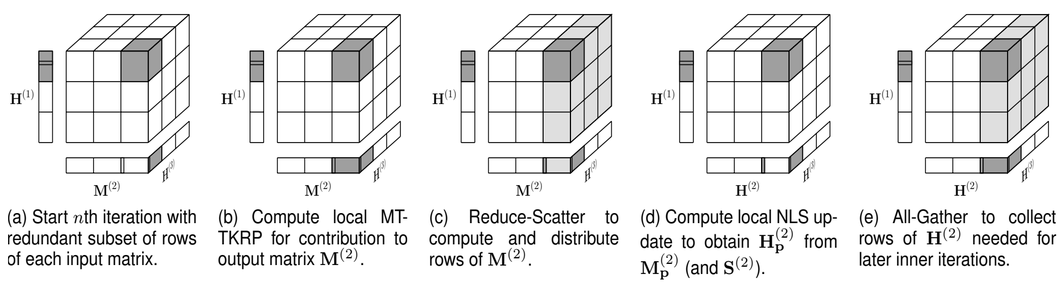


Fig. 6. PLANC class diagram. Utility classes are at the top of the diagram, and the algorithm classes at the bottom of the diagram all derive from the abstract class `DistAUNTF` in orange. The blue arrows denote an "is-a" relationship and the red diamond arrows denote a "has-a" relationship.



(a) Start nth iteration with redundant subset of rows of each input matrix.
(b) Compute local MTTKRP for contribution to output matrix $\mathbf{M}^{(2)}$.
(c) Reduce-Scatter to compute and distribute rows of $\mathbf{M}^{(2)}$.
(d) Compute local NLS update to obtain $\mathbf{H}_\mathbf{p}^{(2)}$ from $\mathbf{M}_\mathbf{p}^{(2)}$ (and $\mathbf{S}^{(2)}$).
(e) All-Gather to collect rows of $\mathbf{H}^{(2)}$ needed for later inner iterations.