ACTIVE BLOCK-MATRIX COMPLETION WITH ADAPTIVE CONFIDENCE SETS

Michal Valko

with Andrea Locatelli and Alexandra Carpentier

SequeL @ Inria Lille — Nord Europe

inventors for the digital world

MARKET SEGMENTATION

MARKET SEGMENTATION

Learn what sub-markets like

MARKET LEARNING/EVALUATION

Sub-markets are different!

- different size
- different complexity

complexity is unknown

Learning on the budget

- budget is tight
- small markets ➪ low budget
- small complexity ➪ low budget

ACTIVE MARKET LEARNING

How to adapt to the unknown structured complexity of the markets?

When is it possible and when it is not?

ACTIVE LEARNING

setting: **for t = 1:n**

we zoom on one market...

...and the natures gives us 1 sample

ACTIVE MATRIX BLOCK COMPLETION

OCCITANIE | PACA | HAUTS-DE-FRANCE

rank: **100**

rank: **?**

rank: **20**

rank: **?**

rank: **5**

rank: **?**

at time t we get a sample from the region we choose:

complexity = rank

OCCITANIE | PACA | HAUTS-DE-FRANCE

ACTIVE MATRIX BLOCK COMPLETION

at time t we get a sample from the region we choose:

Колчинский, Цыбаков (2011)

OCCITANIE

What does it mean <u>to sample</u>?

How do the entries arrive?

Bernoulli model

	A	B	C	D	E
F1	😡		😻		
F2		😡		😻	😻
F3			😡	😡	
F4	😻	😻			😡
F5		😻	😻		

$$Y_{i,j} = (f_{i,j} + \varepsilon_{i,j}) B_{i,j}, \quad (i,j) \in \{1, \ldots, d\}^2$$

$$B_{i,j} \sim_{iid} \mathcal{B}(n/d^2) \qquad |\varepsilon| \leq 1$$

Each of the entry is observed either 0x or 1x.

ACTIVE MATRIX BLOCK COMPLETION

informatics mathematics

at time we get a sample from the region we choose:

Колчинский, Цыбаков (2011)

OCCITANIE

What does it mean to sample?

How do the entries arrive?

Trace regression model

	A	B	C	D	E
F1		😻😻			😻
F2					
F3			😡		
F4	😡				
F5				😻😡	

$$Y_i = f_{X_i} + \varepsilon_i, \quad i = 1, \dots, n$$

$$X_i \sim_{iid} \mathcal{U}_{\{1,\dots,d\}^2} \quad |\varepsilon| \le 1$$

Each of the entry is observed either 0x, 1x, 2x, 3x, ...

realistic situation: $d^2 \gg n$ high-dimensional regime

ACTIVE MATRIX BLOCK COMPLETION

	A	B	C	D	E
F1		😻😻			😻
F2					
F3			😡		
F4	😡				
F5				😻😡	

Trace regression model

$$Y_i = f_{X_i} + \varepsilon_i, \quad i = 1, \ldots, n$$

$$X_i \sim_{iid} \mathcal{U}_{\{1,\ldots,d\}^2} \;\; |\varepsilon| \leq 1$$

Examples of multi-sampling:

- Naturally: **music recommendation**
  - several ratings (or skips)
- By design: **tasting experiments**
  - asking for the customer opinion second time
- Grouped data: **different episodes**, …

MATRIX COMPLETION ESTIMATORS

How to adapt to the unknown structured complexity of the markets?

- alternating least squares minimization
- gradient descend
- soft impute
- matrix lasso (convex relaxation)
- matrix square root lasso

HOW TO COMPLETE THE MATRIX

estimation

$$\frac{\left\|\widehat{\mathbf{M}}_n - \mathbf{M}_0\right\|_F^2}{d_1 d_2} \leq \rho(r, n, d)$$

#samples

rank

d=max(d_1,d_2)

Note: Same HD regime as Lenka yesterday: d ⩽ n ⩽ d^2

good estimator should be <u>adaptive to the rank</u> without knowing it

square-root lasso estimator

$$\widehat{\mathbf{M}}_n(\lambda) \in \underset{\mathbf{M}\in\mathbb{R}^{d_1\times d_2}}{\arg\min} \left\{ \sqrt{\frac{1}{n}\sum_{i=1}^{n}(Y_i - \langle \mathbf{X}_i, \mathbf{M}\rangle)^2} + \lambda \left\|\mathbf{M}\right\|_\star \right\}$$

HOW TO COMPLETE THE MATRIX

estimation

good estimator should be adaptive to the rank without knowing it

$$\widehat{\mathbf{M}}_n(\lambda) \in \underset{\mathbf{M}\in\mathbb{R}^{d_1\times d_2}}{\arg\min} \left\{ \sqrt{\frac{1}{n}\sum_{i=1}^{n}(Y_i - \langle \mathbf{X}_i, \mathbf{M}\rangle)^2} + \lambda \|\mathbf{M}\|_\star \right\}$$

square-root lasso estimator achieves is adaptive

Klopp (2014): $$\frac{||\widehat{\mathbf{M}}_n - \mathbf{M}||_F^2}{d^2} \leq C A^2 \cdot \frac{rd \log d}{n}$$

constant

noise

you cannot do much better

Колчинский, Lounici, Цыбаков's (2011): $$\mathbb{E}\left[\frac{||\widehat{\mathbf{M}}_n - \mathbf{M}||_F^2}{d_1 d_2}\right] \geq \frac{cA^2 rd}{n}$$

HOW TO COMPLETE THE MATRIX

adaptive estimation

$$1 \leq k_0 \leq k_1 \leq d \text{ and } h \in \{0,1\}$$

$$\mathcal{C}_h = \{f : \operatorname{rank}(f) \leq k_h, \|f\|_\infty \leq 1\}$$

Keshavan et al. (2009): **adaptive estimators of f exist** **for both models!**

$$\mathbb{E}\left[\|\tilde{f} - f\|_F\right] \leq Cd\sqrt{\frac{k_h d}{n}} \triangleq C r_h$$

= both models are equivalent

But what about the "error bars" = "confidence sets"?

Carpentier, Klopp, Löffler, Nickl (2016): **α-adaptive confidence sets sometimes exist**

HOW TO COMPLETE THE MATRIX

adaptive estimation: confidence sets

Trace regression model

α-adaptive confidence sets exist

Bernoulli model

α-adaptive confidence do not exist!

Bernoulli model with known noise variance

α-adaptive confidence sets exist

What gives?

adaptive estimation possible

adaptive confidence sets (often) impossible

What gives here?

noise variance!

adaptively bouding $||\widehat{\mathbf{M}} - \mathbf{M}||_F^2$

$$\left| \widehat{R}_N - \frac{||\widehat{\mathbf{M}} - \mathbf{M}||_F^2}{d^2} \right| \leq 8A^2 \sqrt{\frac{\log d}{N}} \Rightarrow \widehat{R}_N + 8A^2 \sqrt{\frac{\log d}{N}} \leq \mathcal{O}\left(\frac{rd \log d}{n}\right)$$

adaptive

honest

CREATING CONFIDENCE SETS FOR 1 EST

	A	B	C	D	E
F1		😻😻			😻
F2					
F3			😡	😡	
F4	😡				
F5		😻		😻😡	

$$\left| \widehat{R}_N - \frac{\|\widehat{\mathbf{M}} - \mathbf{M}\|_F^2}{d^2} \right| \leq 8A^2 \sqrt{\frac{\log d}{N}}$$

find doubly-sampled entries

$$\mathcal{D}' = \{(X_i, Y_i, Y_i')\}_{i=1,\ldots,N}$$

empirically estimate the variance

$$\widehat{R}_N = \frac{1}{N} \sum_{i=1}^{N} \left(Y_i - \langle X_i, \widehat{\mathbf{M}} \rangle \right) \left(Y_i' - \langle X_i, \widehat{\mathbf{M}} \rangle \right)$$

Will we have enough double-samples?

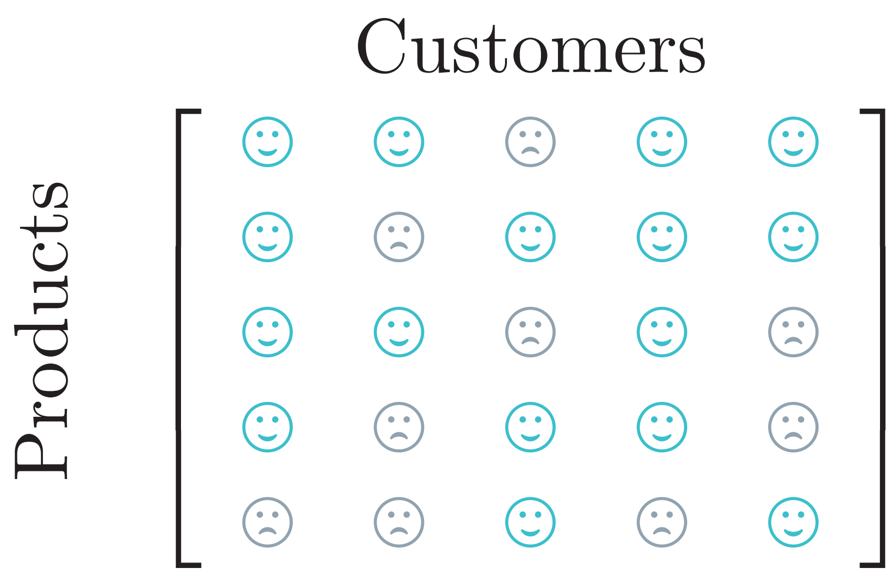
We have adaptive uncertainty!

Whp, for $n \leq d^2$ we get $N \geq \frac{Cn^2}{d^2}$ ⇨ $\widehat{R}_N + 8A^2 \sqrt{\frac{\log d}{N}} \leq \mathcal{O}\left(\frac{rd \log d}{n} \right)$

MATRIX BLOCK COMPLETION MODELS

credits: Alexandra Carpentier

H_0 : Random opinions!



H_1 : Rank one opinions.

H_0 : Random opinions!

H_1 : Rank one opinions.

MATRIX COMPLETION ESTIMATORS

How to adapt to the unknown structured complexity of the markets?

assumptions

$|Y_{k,t}| \leq A$

$||\mathbf{M}^k||_\infty \leq A$

At time t we get a sample from the region we choose:

OCCITANIE	PACA	HAUTS-DE-FRANCE
$\mathbf{M}^1$	$\mathbf{M}^2$	$\mathbf{M}^3$
$\widehat{\mathbf{M}}^1_n$	$\widehat{\mathbf{M}}^2_n$	$\widehat{\mathbf{M}}^3_n$

It depends ... What we want to achieve?

loss parameter: p

extension: weights

- max loss?
- average loss?
- average loss per entry?

$$\mathcal{L}^p_n = \left(\sum_{k \in [K]} \left\| \widehat{\mathbf{M}}^k_n - \mathbf{M}^k \right\|^{2p}_F \right)^{1/p}$$

OUR ALGORITHM: MALOCATE

main loop of MAlocate

- Pick matrix $\mathbf{M}^k$
- NewSamples from $\mathbf{M}^k$
- GetEstimator $\widehat{\mathbf{M}}_t^k$
- EstimateError
- Upper bound on the error
- Accept/Reject $\widehat{\mathbf{M}}_t^k$

output: $\{\widehat{\mathbf{M}}^k\}_{k\in[K]}$

OUR ALGORITHM: MALOCATE

main loop of MAlocate

- Pick matrix $\mathbf{M}^k$
- NewSamples from $\mathbf{M}^k$ ⇨
- GetEstimator $\widehat{\mathbf{M}}_t^k$
- EstimateError
- Upper bound on the error
- Accept/Reject $\widehat{\mathbf{M}}_t^k$

Algorithm 2 `NewSamples` (k, T)

Input: k, T
Sample (uniformly at random)
T new observations $\{(X_i, Y_i)\}_{i \le T}$ from $\mathbf{M}^k$
Output: New dataset $\{(X_i, Y_i)\}_{i \le T}$

output: $\{\widehat{\mathbf{M}}^k\}_{k \in [K]}$

OUR ALGORITHM: MALOCATE

main loop of MAlocate

- Pick matrix $\mathbf{M}^k$
- NewSamples from $\mathbf{M}^k$
- GetEstimator $\widehat{\mathbf{M}}_t^k$
- EstimateError
- Upper bound on the error
- Accept/Reject $\widehat{\mathbf{M}}_t^k$

Algorithm 3 `GetEstimator` $(k, \mathcal{D})$

Input: $k, \mathcal{D}$

$T = \frac{|\mathcal{D}|}{2}, \lambda = C\sqrt{\frac{\log(d_k)}{d_k T}}$

$\widehat{\mathbf{M}} = \underset{\|\mathbf{M}\|_\infty \leq A}{\arg\min} \sqrt{\frac{1}{T}\sum_{i=1}^{T}(Y - \langle X_i, \mathbf{M}\rangle)^2} + \lambda \|\mathbf{M}\|_*$

Output: Estimator $\widehat{\mathbf{M}}$

output: $\{\widehat{\mathbf{M}}^k\}_{k\in[K]}$

OUR ALGORITHM: MALOCATE

main loop of MAlocate

- Pick matrix $\mathbf{M}^k$
- NewSamples from $\mathbf{M}^k$
- GetEstimator $\widehat{\mathbf{M}}_t^k$
- EstimateError ⇨
- Upper bound on the error
- Accept/Reject $\widehat{\mathbf{M}}_t^k$

output: $\{\widehat{\mathbf{M}}^k\}_{k\in[K]}$

Algorithm 4 EstimateError $(\widehat{\mathbf{M}}, \mathcal{D})$

Input: $\widehat{\mathbf{M}}, \mathcal{D}$

$T = \frac{|\mathcal{D}|}{2}$

Find double-sampled entries

$\mathcal{D}' = \{(X_i, Y_i, Y_i')\}_{i=1,\dots,N}$ in $\mathcal{D}_{T+1,\dots,2T}$

$\widehat{R}_N = \frac{1}{N}\sum_{i=1}^N \left(Y_i - \langle X_i, \widehat{\mathbf{M}}\rangle\right)\left(Y_i' - \langle X_i, \widehat{\mathbf{M}}\rangle\right)$

Output: Number of double-sampled entries N and error estimate $\widehat{R}_N$

OUR ALGORITHM: MALOCATE

main loop of MAlocate

- Pick matrix $\mathbf{M}^k$
- NewSamples from $\mathbf{M}^k$
- GetEstimator $\widehat{\mathbf{M}}_t^k$
- EstimateError $N_t^k, \widehat{R}_{N_t^k}$
- Upper bound on the error ⇨
- Accept/Reject $\widehat{\mathbf{M}}_t^k$ **iff we improved**

output: $\{\widehat{\mathbf{M}}^k\}_{k \in [K]}$

upper bound on the error

$$\frac{||\widehat{\mathbf{M}_t^k} - \mathbf{M}^k||_F^2}{d^2}$$

⇩

$$\widehat{R}_{N_k^t} + 8A^2\sqrt{\frac{\log(d_k)}{N_t^k}}$$

OUR ALGORITHM: MALOCATE

main loop of MAlocate

output: $\{\widehat{\mathbf{M}}^k\}_{k\in[K]}$

How good are the estimates?

Theorem: For max loss (p = ∞), whp:

How good is this?

$$\max_{k\in[K]} \left\| \widehat{\mathbf{M}}_n^k - \mathbf{M}^k \right\|_F^2 \leq \mathcal{O}\left(\frac{\sum_{k=1}^K r_k d_k^3 \log(d_k)}{n} \right)$$

For general p:

How good is this?

$$\left(\sum_{k\in[K]} \left\| \widehat{\mathbf{M}}_n^k - \mathbf{M}^k \right\|_F^{2p} \right)^{1/p} \leq \mathcal{O}\left(\frac{\left(\sum_{k=1}^K (r_k d_k^3 \log(d_k))^{\frac{p}{p+1}} \right)^{\frac{p+1}{p}}}{n} \right)$$

OUR ALGORITHM: MALOCATE

For general p:

How good is this?

$$\left(\sum_{k\in[K]} \left\|\widehat{\mathbf{M}}_n^k - \mathbf{M}^k\right\|_F^{2p}\right)^{1/p} \leq \mathcal{O}\left(\frac{\left(\sum_{k=1}^K (r_k d_k^3 \log(d_k))^{\frac{p}{p+1}}\right)^{\frac{p+1}{p}}}{n}\right)$$

Complexity depends on the rank and the dimension of all subproblems!

complexity: $c_k = r_k d_k^3 \log(d_k)$

all: $\mathbf{c} = (c_1, \ldots, c_K)$

How good is this?

$$\frac{\|\mathbf{c}\|_{\frac{p}{p+1}}}{n} \leq \frac{K\,\|\mathbf{c}\|_p}{n}$$

Sampling UAR?

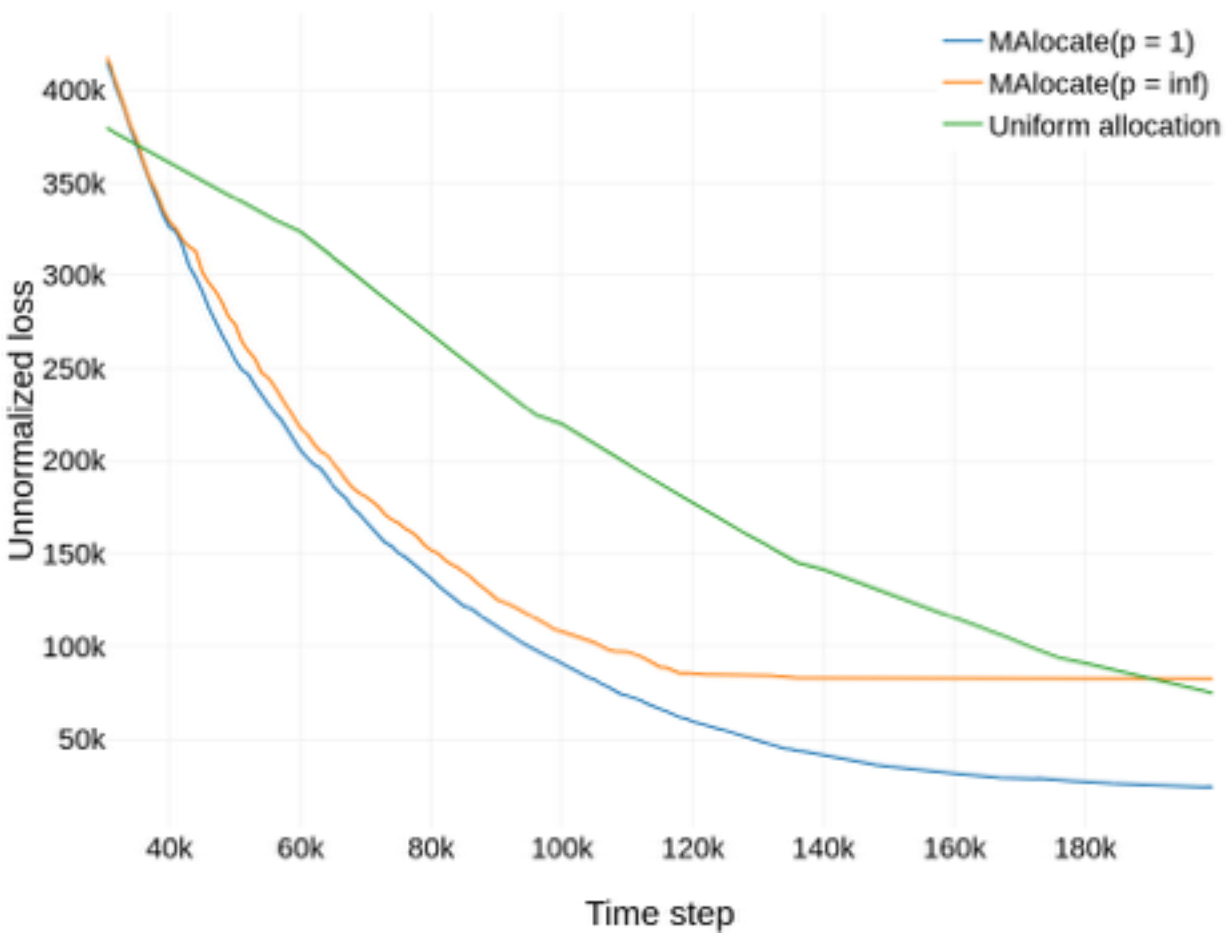
As good as it gets?

we have $\|\mathbf{x}\|_{q_1} \leq K^{1/q_1 - 1/q_2}\|\mathbf{x}\|_{q_2}$ for $0 < q_1 < q_2$

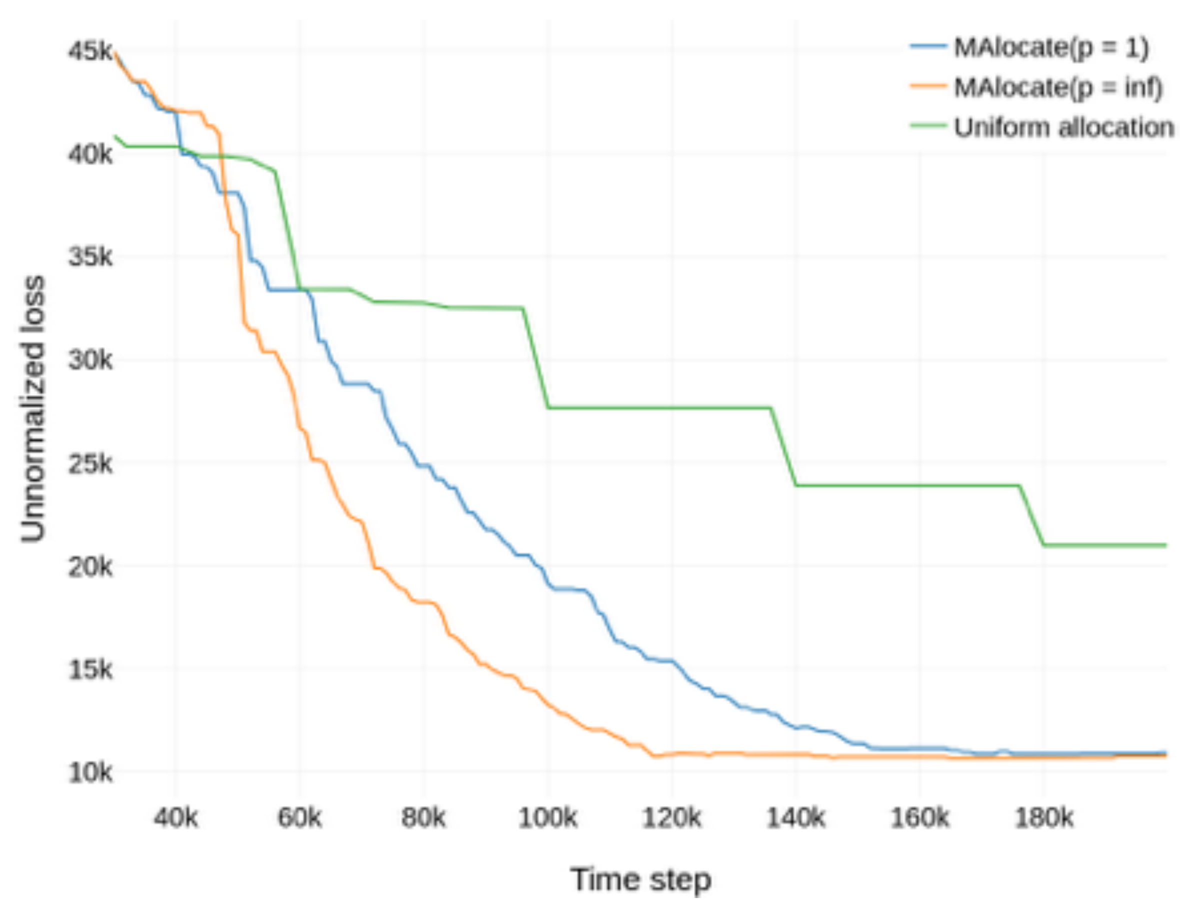
FIRST EXPERIMENT

Proof by picture:

sum loss



max loss



rank r, dimension d: $\mathbf{U} \in \mathbb{R}^{d \times r}$, $\mathbf{V} \in \mathbb{R}^{r \times d}$:

entries $\sim \mathcal{N}(0, \sigma_r^2 = r^{-1/2})$, noise $\mathcal{N}(0, \sigma = 0.1)$

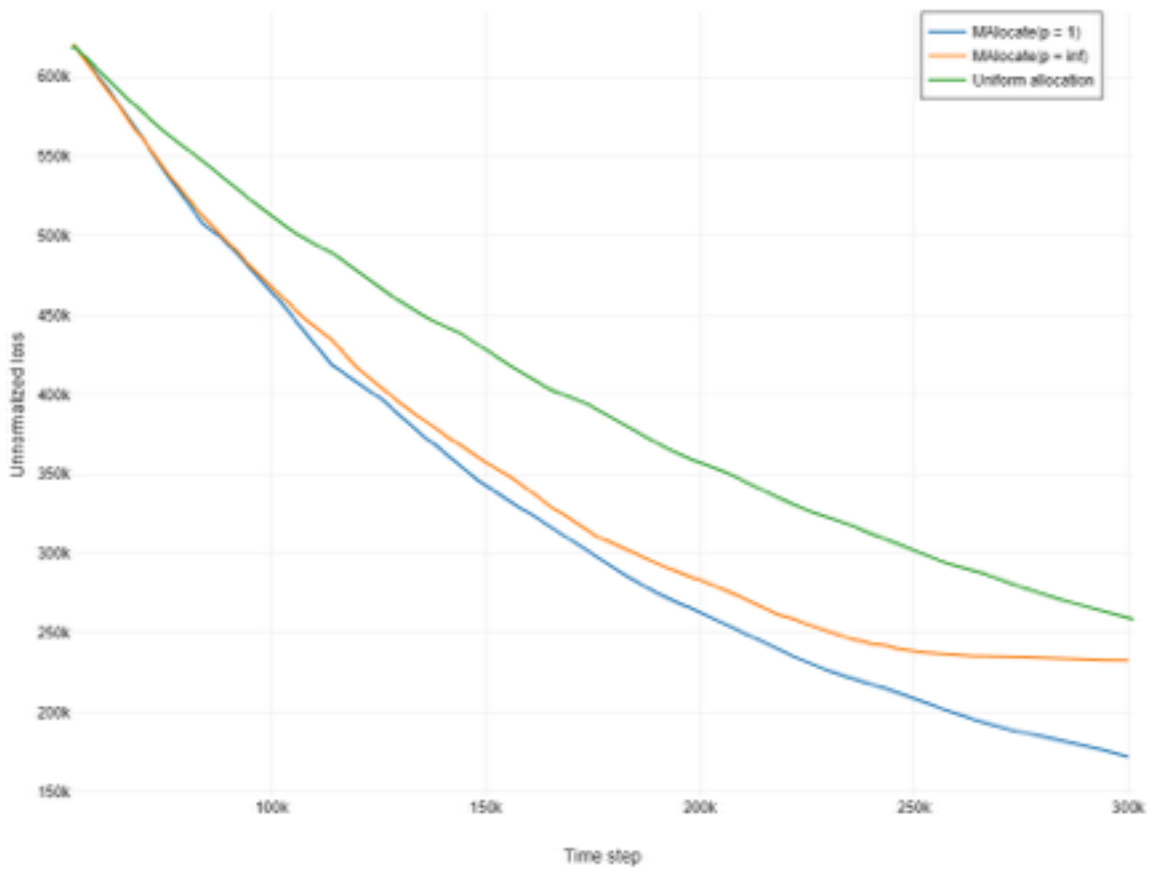
$d_k = d = 200$, $K = 10$, $r_1 = 40$ and $r_k = 10$ for the rest

SECOND EXPERIMENT

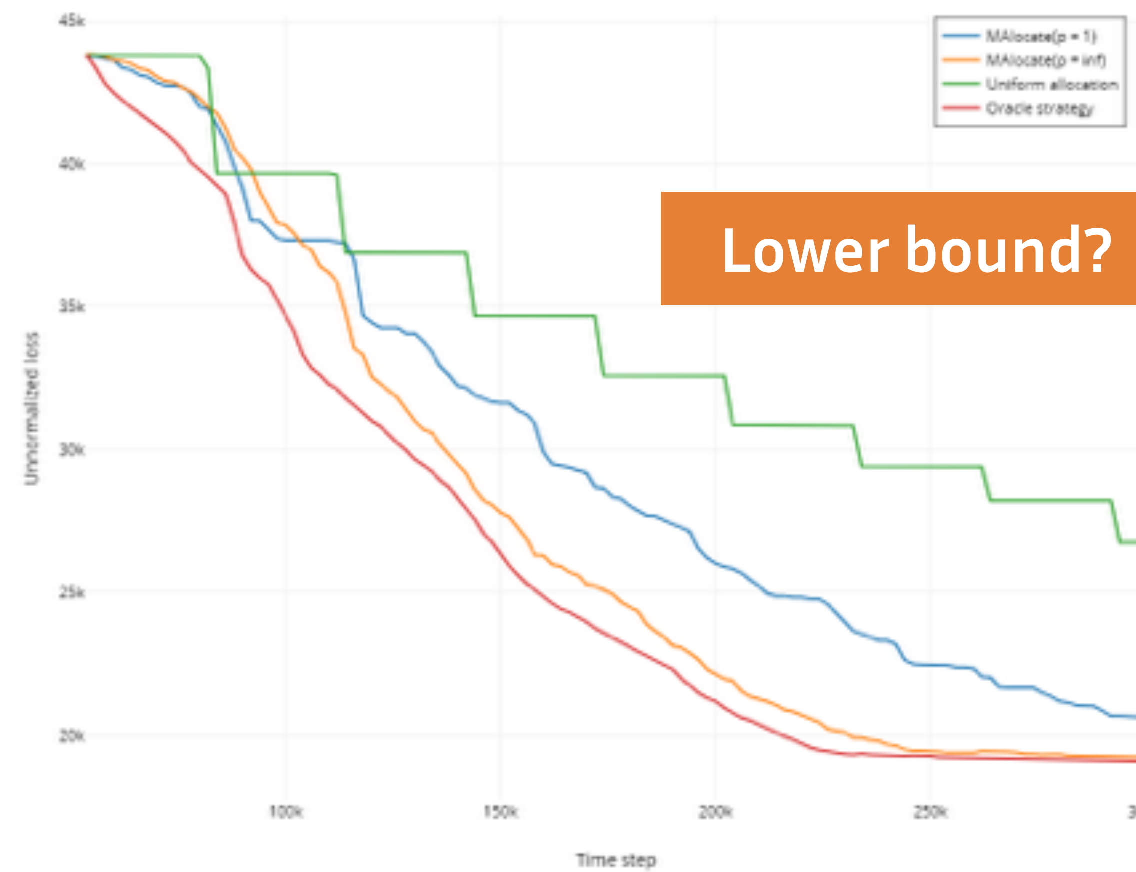
Proof by picture:

Better proof?

sum loss



max loss



Lower bound?

rank r, dimension d: $\mathbf{U} \in \mathbb{R}^{d\times r}$, $\mathbf{V} \in \mathbb{R}^{r\times d}$:
entries $\sim \mathcal{N}(0, \sigma_r^2 = r^{-1/2})$, noise $\mathcal{N}(0, \sigma = 0.1)$
$d_k = d = 200,\ K = 15,\ r_k = 18 + 0.0015k^4$

MALOCATE GUARANTEE: HIGH LEVEL

For general p:

$$\left(\sum_{k\in[K]} \left\|\widehat{\mathbf{M}}_n^k - \mathbf{M}^k\right\|_F^{2p}\right)^{1/p} \leq \mathcal{O}\left(\frac{\left(\sum_{k=1}^K (r_k d_k^3 \log(d_k))^{\frac{p}{p+1}}\right)^{\frac{p+1}{p}}}{n}\right)$$

Property 1: From sampling criterion ➪ for all k

standard analysis won't work

$$\left\|\widehat{\mathbf{M}}^k - \mathbf{M}^k\right\|_F^2 \leq \mathcal{O}\left(T_k(n)^{1/p}\left(\sum_k (r_k d_k^3 \log(d_k))^{p/(p+1)}\right)^{(p+1)/p} n^{-(p+1)/p}\right)$$

increases with $T_k(n)$

Property 2: Bound on the estimator ➪ for all k

$$\left\|\widehat{\mathbf{M}}^k - \mathbf{M}^k\right\|_F^2 \leq \mathcal{O}\left(\frac{r_k d_k^3 \log(d_k)}{T_k(n)}\right)$$

decreases with $T_k(n)$

MALOCATE GUARANTEE: HIGH LEVEL

For general p:

$$\left(\sum_{k\in[K]} \left\|\widehat{\mathbf{M}}_n^k - \mathbf{M}^k\right\|_F^{2p}\right)^{1/p} \leq \mathcal{O}\left(\frac{\left(\sum_{k=1}^K (r_k d_k^3 \log(d_k))^{\frac{p}{p+1}}\right)^{\frac{p+1}{p}}}{n}\right)$$

Property 1: From sampling criterion ➩ for all k

$$\left\|\widehat{\mathbf{M}}^k - \mathbf{M}^k\right\|_F^2 \leq \mathcal{O}\left(T_k(n)^{1/p}\left(\sum_k (r_k d_k^3 \log(d_k))^{p/(p+1)}\right)^{(p+1)/p} n^{-(p+1)/p}\right)$$

increases with $T_k(n)$

Property 2: Bound on the estimator ➩ for all k

$$\left\|\widehat{\mathbf{M}}^k - \mathbf{M}^k\right\|_F^2 \leq \mathcal{O}\left(\frac{r_k d_k^3 \log(d_k)}{T_k(n)}\right)$$

decreases with $T_k(n)$

Maximizing over $T_k(n)$?

MALOCATE GUARANTEE: DEEP DIVE

For general p:

$$\left(\sum_{k\in[K]}\left\|\widehat{\mathbf{M}}_n^k-\mathbf{M}^k\right\|_F^{2p}\right)^{1/p}\leq\mathcal{O}\left(\frac{\left(\sum_{k=1}^K(r_kd_k^3\log(d_k))^{\frac{p}{p+1}}\right)^{\frac{p+1}{p}}}{n}\right)$$

Step 1: Favorable event for a round t and matrix k

$$(1)\quad \frac{||\widehat{\mathbf{M}}_t^k-\mathbf{M}^k||_F^2}{d_k^2}\leq CA^2\cdot\frac{r_kd_k\log(d_k)}{t},$$

$$(2)\quad N_t^k\geq\frac{t^2}{64d_k^2},$$

$$(3)\quad \left|\widehat{R}_N-\frac{||\widehat{\mathbf{M}}_t^k-\mathbf{M}^k||_F^2}{d_k^2}\right|\leq 8A^2\sqrt{\frac{\log(d_k)}{N_t^k}}.$$

Conclusion: whp this holds for all k and all t we need

MALOCATE GUARANTEE: DEEP DIVE

For general p:

$$\left(\sum_{k\in[K]} \left\| \widehat{\mathbf{M}}_n^k - \mathbf{M}^k \right\|_F^{2p} \right)^{1/p} \leq \mathcal{O}\left(\frac{\left(\sum_{k=1}^K (r_k d_k^3 \log(d_k))^{\frac{p}{p+1}} \right)^{\frac{p+1}{p}}}{n} \right)$$

Step 2: After enough samples

... there exist a matrix m s.t.

$$n \geq 48 \sum_{k\in[K]} d_k \log(d_k) = 12 \sum_k T_k^I$$

$$T_m(n) - 6T_m^I \geq \frac{(r_m d_m^3 \log(d_m))^{\frac{p}{p+1}}}{\sum_{k\in[K]} (r_k d_k^3 \log(d_k))^{\frac{p}{p+1}}} \left(n - 6 \sum_{k\in[K]} T_k^I \right)$$

$$\geq \frac{(r_m d_m^3 \log(d_m))^{\frac{p}{p+1}}}{\sum_{k\in[K]} (r_k d_k^3 \log(d_k))^{\frac{p}{p+1}}} \left(\frac{n}{2} \right)$$

MALOCATE GUARANTEE: DEEP DIVE

For general p:

$$\left(\sum_{k \in [K]} \left\| \widehat{\mathbf{M}}_n^k - \mathbf{M}^k \right\|_F^{2p} \right)^{1/p} \leq \mathcal{O}\left(\frac{\left(\sum_{k=1}^K (r_k d_k^3 \log(d_k))^{\frac{p}{p+1}} \right)^{\frac{p+1}{p}}}{n} \right)$$

Step 2: After enough samples

$$n \geq 48 \sum_{k \in [K]} d_k \log(d_k) = 12 \sum_k T_k^I$$

... there exist a matrix m s.t.

Matrix m is picked at least 2x

$$T_m(n) \geq \frac{c_m}{\sum_k c_k} \left(\frac{n}{2} \right)$$

$$c_k \triangleq (r_k d_k^3 \log(d_k))^{\frac{p}{p+1}}$$

MALOCATE GUARANTEE: DEEP DIVE

For general p:

$$\left(\sum_{k\in[K]} \left\|\widehat{\mathbf{M}}_n^k - \mathbf{M}^k\right\|_F^{2p}\right)^{1/p} \leq \mathcal{O}\left(\frac{\left(\sum_{k=1}^K (r_k d_k^3 \log(d_k))^{\frac{p}{p+1}}\right)^{\frac{p+1}{p}}}{n}\right)$$

Step 2: After enough samples | **… there exist a matrix m s.t. for its last round: t_1**

$$\frac{d_m^2 B_m(t_1)}{T_m(t_1)^{1/p}} \leq A^2 \max(C, 128)\left(\frac{r_m d_m^3 \log(d_m)}{T_m(t_2) T_m(t_1)^{1/p}}\right)$$

$$= 2^{1/p} A^2 \max(C, 128)\left(\frac{r_m d_m^3 \log(d_m)}{T_m(t_2)^{\frac{p+1}{p}}}\right)$$

$$\leq 2^{1/p} 64 A^2 \max(C, 128)\left(\frac{\sum_k c_k}{n}\right)^{\frac{p+1}{p}}$$

MALOCATE GUARANTEE

For general p:

$$\left(\sum_{k\in[K]}\left\|\widehat{\mathbf{M}}_n^k-\mathbf{M}^k\right\|_F^{2p}\right)^{1/p}\leq\mathcal{O}\left(\frac{\left(\sum_{k=1}^K(r_kd_k^3\log(d_k))^{\frac{p}{p+1}}\right)^{\frac{p+1}{p}}}{n}\right)$$

Step 3: For all other matrices i

$$\frac{d_i^2B_i(t_1)}{T_i(t_1)^{\frac{1}{p}}}\leq\frac{d_m^2B_m(t_1)}{T_m(t_1)^{\frac{1}{p}}}<\infty$$

.... and from what we know about m ...

$$d_i^2B_i(t_1)\leq 2^{1/p}64A^2\max(C,128)T_i(t_1)^{\frac{1}{p}}\left(\frac{\sum_k c_k}{n}\right)^{\frac{p+1}{p}}$$

increases with $T_i(t_1)$

MALOCATE GUARANTEE: DEEP DIVE

For general p:

$$\left(\sum_{k\in[K]}\left\|\widehat{\mathbf{M}}_n^k-\mathbf{M}^k\right\|_F^{2p}\right)^{1/p}\leq\mathcal{O}\left(\frac{\left(\sum_{k=1}^K(r_k d_k^3\log(d_k))^{\frac{p}{p+1}}\right)^{\frac{p+1}{p}}}{n}\right)$$

Step 3: For all other matrices i…. for the round just before the round t_1

$$B_i(t_1)\leq B_i(t_i)\leq\widehat{R}_{N_i^{t_i}}+8A^2\sqrt{\frac{\log(d_i)}{N_i^{t_i}}}$$

$$\leq\left\|\widehat{\mathbf{M}}_i^{t_i}-\mathbf{M}^i\right\|_F^2+16A^2\sqrt{\frac{\log(d_i)}{N_i^{t_i}}}$$

$$\leq CA^2\left(\frac{r_i d_i\log(d_i)}{T_i(t_i)}\right)+16A^2\sqrt{\frac{\log(d_i)}{N_i^{t_i}}}$$

$$\leq CA^2\left(\frac{r_i d_i\log(d_i)}{T_i(t_i)}\right)+128A^2\frac{d_i\sqrt{\log(d_i)}}{T_i(t_i)}$$

$$\leq 2A^2\max(C,128)\left(\frac{r_i d_i\log(d_i)}{T_i(t_1)}\right)$$

decreases with $T_i(t_1)$

MALOCATE GUARANTEE: DEEP DIVE

For general p:

$$\left(\sum_{k\in[K]}\left\|\widehat{\mathbf{M}}_n^k-\mathbf{M}^k\right\|_F^{2p}\right)^{1/p}\leq\mathcal{O}\left(\frac{\left(\sum_{k=1}^K(r_kd_k^3\log(d_k))^{\frac{p}{p+1}}\right)^{\frac{p+1}{p}}}{n}\right)$$

Property 1: From sampling criterion ➩ for all i

$$\left\|\widehat{\mathbf{M}}_n^i-\mathbf{M}^i\right\|_F^2\leq 2^{1/p}64A^2\max(C,128)T_i(t_1)^{\frac{1}{p}}\left(\frac{\sum_k c_k}{n}\right)^{\frac{p+1}{p}}$$

increases with $T_i(t_1)$

Property 2: Bound on the estimator ➩ for all i

$$\left\|\widehat{\mathbf{M}}_n^i-\mathbf{M}^i\right\|_F^2\leq 2A^2\max(C,128)\left(\frac{r_id_i^3\log(d_i)}{T_i(t_1)}\right)$$

decreases with $T_i(t_1)$

Maximizing over $T_k(n)$

MALOCATE GUARANTEE: DEEP DIVE

For general p:

$$\left(\sum_{k\in[K]} \left\|\widehat{\mathbf{M}}_n^k - \mathbf{M}^k\right\|_F^{2p}\right)^{1/p} \leq \mathcal{O}\left(\frac{\left(\sum_{k=1}^K (r_k d_k^3 \log(d_k))^{\frac{p}{p+1}}\right)^{\frac{p+1}{p}}}{n}\right)$$

Final step: **summing the errors**

$$\begin{aligned}
\mathcal{L}_n^p &= \left(\sum_{k\in[K]} \left\|\widehat{\mathbf{M}}_n^k - \mathbf{M}^k\right\|_F^{2p}\right)^{1/p} \\
&\leq \mathcal{O}\left(\frac{\left(\sum_k c_k\right)}{n}\left(\sum_{k=1}^K c_k\right)^{1/p}\right) \\
&\leq \mathcal{O}\left(\frac{\left(\sum_k c_k\right)^{\frac{p+1}{p}}}{n}\right) \\
&\leq \mathcal{O}\left(\frac{\left(\sum_k (r_k d_k^3 \log(d_k))^{\frac{p}{p+1}}\right)^{\frac{p+1}{p}}}{n}\right)
\end{aligned}$$

LOWER BOUND

Theorem: **For max loss (p = ∞), whp:**

How good is this?

$$\max_{k\in[K]} \left\| \widehat{\mathbf{M}}_n^k - \mathbf{M}^k \right\|_F^2 \leq \mathcal{O}\left(\frac{\sum_{k=1}^K r_k d_k^3 \log(d_k)}{n} \right)$$

Theorem: **For max loss (p = ∞) and any active strategy S, there exists a problem**

$$\mathbb{E}_{P,\mathcal{S}}\left[\max_{k\in[K]} \left(\left\| \widehat{\mathbf{M}}^k - \mathbf{M}^k \right\|_F^2 \right) \right] \geq \frac{A^2}{2048} \frac{\sum_{k=1}^K r_k d_k^3}{n}$$

minimax optimal up to logs

works even for S that knows ranks in advance

- **logs coming from the estimators**
- **Klopp (2015) gives sharp bounds for Bernoulli**

LOWER BOUND PROOF IDEA

Theorem: For max loss (p = ∞) and any active strategy S, there exists a problem

$$\mathbb{E}_{P,\mathcal{S}}\left[\max_{k\in[K]}\left(\left\|\widehat{\mathbf{M}}^k - \mathbf{M}^k\right\|_F^2\right)\right] \geq \frac{A^2}{2048}\frac{\sum_{k=1}^K r_k d_k^3}{n}$$

By Direchlet pigeonhole principle there exists $m \in [K]$

such that $\mathbb{E}_{P,\mathcal{S}}[T_k(n)] \leq \frac{r_m d_m^3}{\sum_k r_k d_k^3} n$ and we show that

Цыбаков's LB on many hypotheses + Gilbert-Varshamov LB

$$\inf_{\widehat{P}} \sup_{P\in\mathcal{P}} \mathbb{E}_P\left(\max_k\left(||\widehat{\mathbf{M}}^k - \mathbf{M}^k||_F^2\right)\right) \geq \inf_{\widehat{P}} \max_{k\in[K]} \sup_{P\in\mathcal{P}_k} \mathbb{E}_P\left(\max_i(||\widehat{\mathbf{M}}^i - \mathbf{M}^i||_F^2)\right)$$

$$\geq \inf_{\widehat{P}} \max_{k\in[K]} \sup_{P\in\mathcal{P}_k} \mathbb{E}_P(||\widehat{\mathbf{M}}^k - \mathbf{M}^k||_F^2)$$

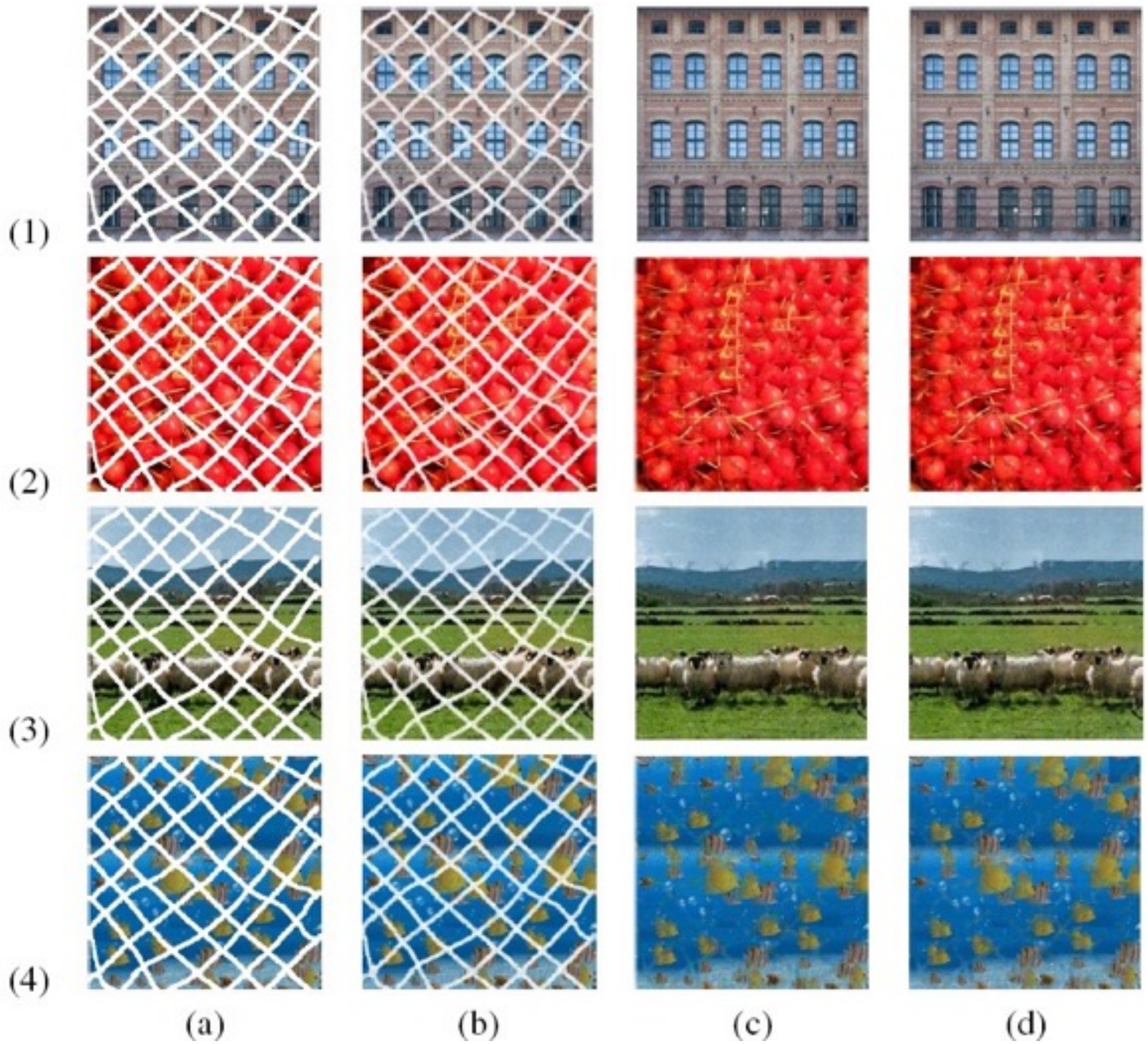
$$\geq \max_{k\in[K]} \frac{A^2}{2048} \cdot \frac{r_k d_k^3}{\tau_k},$$

Note: Castro+Nowak (2018)

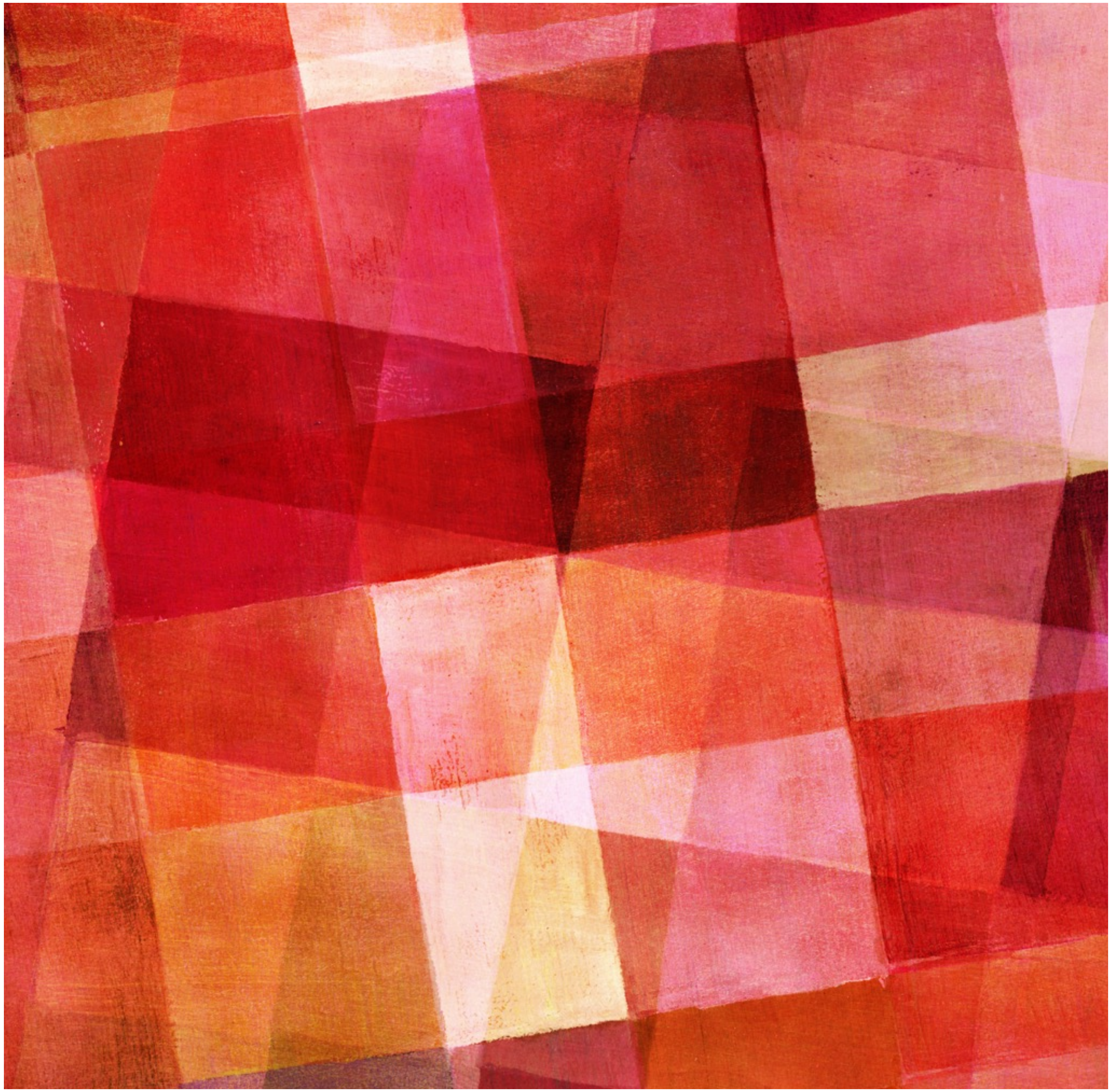
DISCUSSION AND WHAT'S NEXT

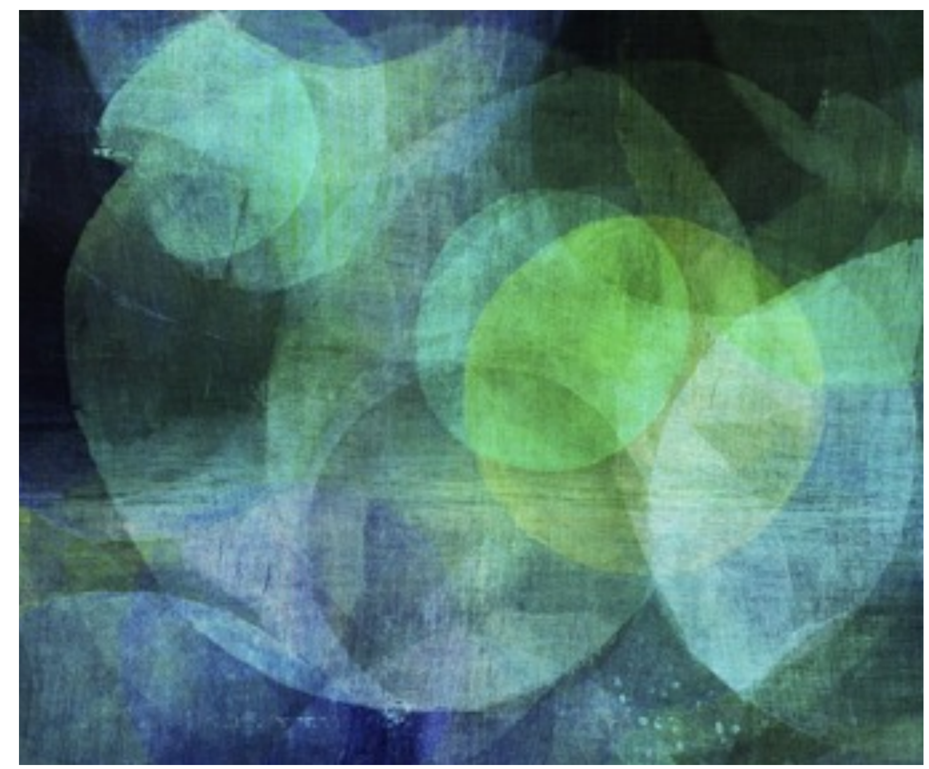
- active adaptation to (sub)-structure of different complexities
- **adaptive inference is necessary**
  - inference = estimation + uncertainty quantification
  - with matrix completion it is sometimes possible
- square-root lasso estimator
  - the approach takes the estimator + guarantees as blackbox
- contrib: **MAlocate** and the guarantees
  - anytime, fixed budget (fixed confidence - δ,ε - possible)
- numerical simulations: adaptive and close to the oracle
  - next step: recommendation datasets
- results are **generalizable** for knowing whether this **active adaptation** is possible or not

BEYOND RECOMMENDER SYSTEMS?



credits: Junzhou Huang





Michal Valko, SequeL, Inria Lille - Nord Europe, michal.valko@inria.fr
http://researchers.lille.inria.fr/~valko/hp/