Active multiple matrix completion with adaptive confidence sets

Andrea Locatelli and ***Alexandra Carpentier*** and ***Michal Valko***
andrea.locatelli@ovgu.de and *alexandra.carpentier@ovgu.de* and *michal.valko@inria.fr*

Motivation

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

setting: for t = 1:n

we zoom on one market...

...and the natures gives us 1 sample

À la une

- GOAL: **Actively complete multiple matrices**
- NEW: **Adapts to unknown complexity**
- OPTIMAL: **We give matching lower bound.**
- GENERALIZABLE: **When adaptive confidence sets exist.**

Matrix completion

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

Bernoulli model

$$Y_{i,j} = (f_{i,j} + \varepsilon_{i,j})B_{i,j}, \quad (i,j) \in \{1,\dots,d\}^2$$
$$B_{i,j} \sim_{iid} \mathcal{B}(n/d^2)$$

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

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

Trace regression model

$$Y_i = f_{X_i} + \varepsilon_i, \quad i = 1,\dots,n$$
$$X_i \sim_{iid} \mathcal{U}_{\{1,\dots,d\}^2} \; |\varepsilon| \le 1$$

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

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

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

square-root estimator
is adaptive (Klopp, 2014)

Adaptive confidence sets

$$\frac{\|\widehat{\mathbf{M}}_n - \mathbf{M}\|_F^2}{d^2} \le CA^2 \cdot \frac{rd\log d}{n}$$

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

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

find doubly-sampled entries

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

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

Setting

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

assumptions

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

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

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

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

It depends ... What we want to achieve?

loss parameter: p

extension: weights

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

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

Algorithm

main loop of **MAlocate**

- Pick matrix $\mathbf{M}^k \triangleq \arg\max_k d_k^2 B_k(t) T_k(t)^{-1/p}$
- 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 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 $\{(X_i, Y_i)\}_{i\le T}$

Algorithm 3 `GetEstimator` $(k, \mathcal{D})$
Input: $k, \mathcal{D}$
$T = \frac{|\mathcal{D}|}{2}, \lambda = C\sqrt{\frac{\log(d_k)}{d_k T}}$
$\widehat{\mathbf{M}} = \underset{\|\mathbf{M}\|_\infty \le A}{\arg\min} \sqrt{\frac{1}{T}\sum_{i=1}^{T}(Y - \langle X_i, \mathbf{M}\rangle)^2} + \lambda\|\mathbf{M}\|_*$
Output: Estimator $\widehat{\mathbf{M}}$

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 $\widehat{R}_N$

upper bound on the error

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

iff we improved

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

Guarantees

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

How good are the estimates?

Theorem: For max loss (p = ∞), whp:

How good is this?

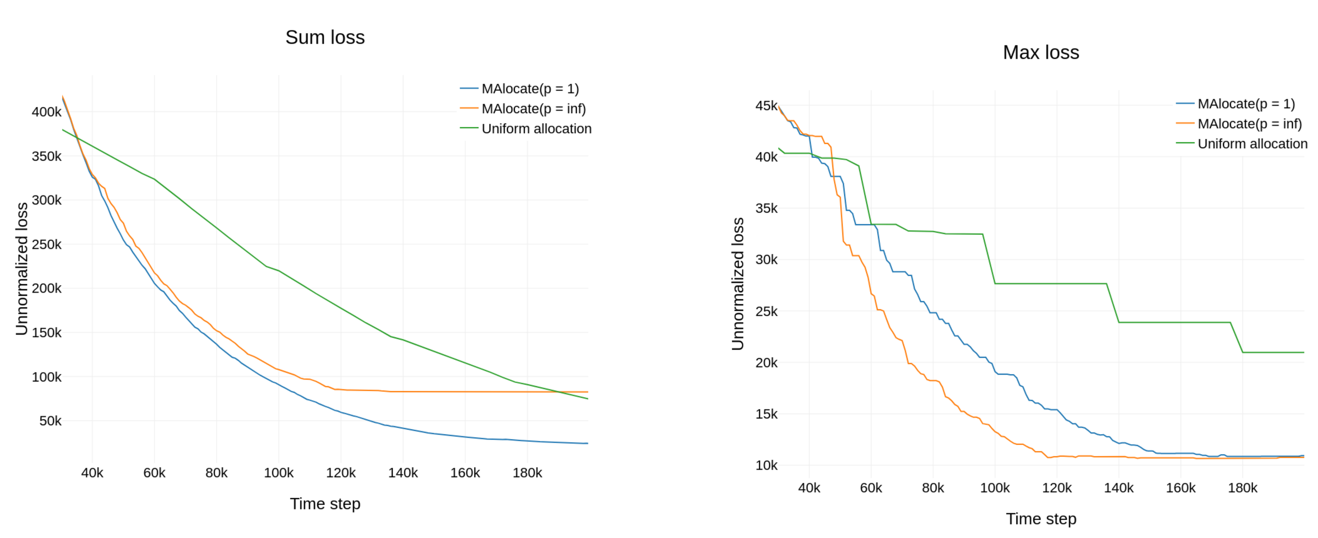
$$\max_{k\in[K]} \left\|\widehat{\mathbf{M}}_n^k - \mathbf{M}^k\right\|_F^2 \le \mathcal{O}\left(\frac{\sum_{k=1}^{K} r_k d_k^3 \log(d_k)}{n}\right)$$

For general p:

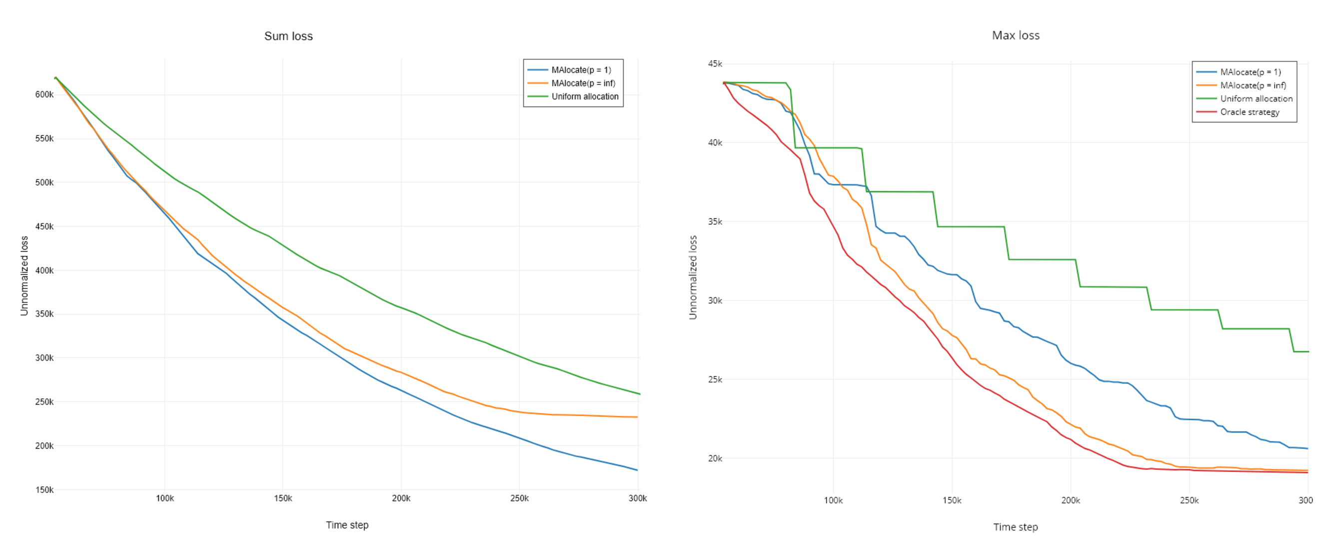
How good is this?

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

Experiments



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



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 $\mathcal{N}(0, \sigma = 0.1)$
$d_k = d = 200$, $K = 15$, $r_k = 18 + 0.0015k^4$