Matrix factorization for time series analysis

Pierre Alquier

RIKEN

AIP
Center for
Advanced Intelligence Project

StatML – Imperial and Oxford
January 20, 2022

Co-authors

Alquier, P. and Marie, N. (2019). Matrix Factorization for Multivariate Time Series Analysis. *The Electronic Journal of Statistics*.

Alquier, P., Marie, N. and Rosier, A. (2021). Tight Risk Bound for High Dimensional Time Series Completion. *Preprint arXiv :2102.08178*.

Nicolas Marie

Université Paris Nanterre

Amélie Rosier

ESME Sudria and Université Paris Nanterre

Overview

- many statistical problems or ML problems involve the estimation of a high-dimensional, low-rank matrix.
- under an independence assumption on the observations, fast algorithms are known, with strong statistical guarantees.
- little theory in the case of time series / dependent observations.
- take-home message : the known algorithms can also be used for time series safely.

Contents

Contents

Some examples

Some stats/ML problems where the parameter is a large matrix :

1. denoising,
2. reduced rank regression (RRR),
3. matrix completion (e.g for recommender systems),
4. quantum tomography,
5. community detection in graphs...

Inzenman, A. J. (1975). Reduced-rank regression for the multivariate linear model. *Journal of Multivariate Analysis*.

Candès, E. and Tao, T. (2010). The power of convex relaxation : Near-optimal matrix completion. *IEEE Transactions on Information Theory*.

Gross, D., Liu, Y. K., Flammia, S. T., Becker, S. and Eisert, J. (2010). Quantum state tomography via compressed sensing. *Physical review letters*.

Paul, S. and Chen, Y. (2020). Spectral and matrix factorization methods for consistent community detection in multi-layer networks. *The Annals of Statistics*.

Dimension reduction

- in denoising, M is $d \times T$ and we observe exactly $n = d \times T$ entries (with noise),
- in matrix completion, M is $d \times T$ and we observe only n entries with $n \ll d \times T$ (possibly with noise)...

Estimation is only possible under dimension reduction.

The low-rank assumption is meaningful in many examples (and successful in applications).

Singular Value Decomposition (SVD)

Assume that M is $d \times T$ and $\mathrm{rank}(M) = r$.

$$M = \underbrace{\left(\begin{array}{c|c|c} & & \\ U_1 & \dots & U_r \\ & & \end{array}\right)}_{=U\ (d \times r)} \underbrace{\begin{pmatrix} \sigma_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \sigma_r \end{pmatrix}}_{=\Sigma\ (r \times r)} \underbrace{\begin{pmatrix} V_1^T \\ \hline \vdots \\ \hline V_r^T \end{pmatrix}}_{=V^T\ (r \times T)}$$

$$M = \underbrace{(U\Sigma)}_{d \times r} \underbrace{V^T}_{r \times T} = \underbrace{W}_{d \times r} \underbrace{V^T}_{r \times T}.$$

We have $\simeq r(d+T)$ parameters to estimate.

Existing results

Efficient algorithms exploiting a low-rank assumption were proposed.

A strong theory was developped so these algorithms come with guarantees. For example :

- in the noiseless case

Candès, E. and Tao, T. (2010). The power of convex relaxation : Near-optimal matrix completion. *IEEE Transactions on Information Theory*.

- when the noise is i.i.d

Koltchinskii, V., Lounici, K. and Tsybakov, A. (2011). Nuclear-norm penalization and optimal rates for noisy low-rank matrix completion. *The Annals of Statistics*.

Contents

Matrix denoising model

Matrix denoising model

We observe X where

$$X = M + \mathcal{E}$$

and

- M is a $d \times T$ matrix (to recover),
- $\mathcal{E}$ is some random noise.

Estimation

Reminder : Frobenius norm and nuclear norm

$$\|A\|_F = \sqrt{\sum_{i,j} A_{i,j}^2}, \ \|A\|_1 = \sum_i \sigma_i(A).$$

- Rank constrained estimator :

$$\widehat{M}_k = \operatorname*{arg\,min}_{\mathrm{rank}(A)\leq k} \|X - A\|_F^2.$$

Solution : in the SVD of X, replace each $\sigma_i(X)$ by $\sigma_i(X)\mathbf{1}_{\{i>k\}}$.

- Nuclear-norm penalized estimator :

$$\widetilde{M}_\lambda = \operatorname*{arg\,min}_{A} \left\{ \|X - A\|_F^2 + \lambda \|A\|_1 \right\}.$$

Solution : in the SVD of X, replace each $\sigma_i(X)$ by $\max(\sigma_i(X) - \frac{\lambda}{2}, 0)$.

Theoretical analysis

We first focus on the independent case : the entries of $\mathcal{E}$ are i.i.d $\mathcal{N}(0, \sigma^2)$.

Theorem

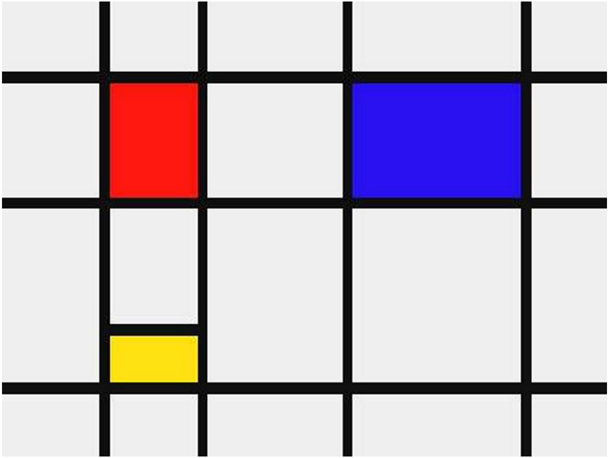
For any $s > 0$, with probability at least $1 - 2\exp(-s)$,

$$\|\widehat{M}_k - M\|_F^2 \leq C \left\{ \inf_{\mathrm{rank}(A) \leq k} \|A - M\|_F^2 + \sigma^2 k(d + T + s) \right\}$$

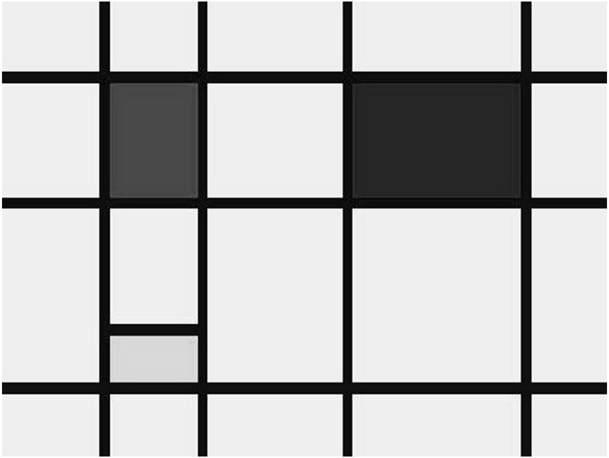
for some universal constant C. In particular, if $\mathrm{rank}(M) = r \leq k$,

$$\|\widehat{M}_k - M\|_F^2 \leq C\sigma^2 k(d + T + s).$$

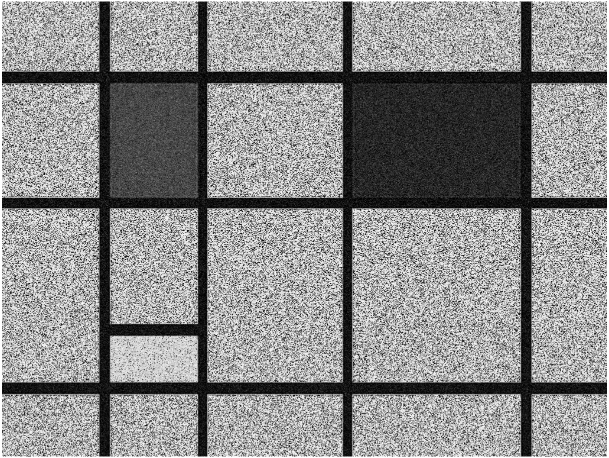
A toy example : denoising Mondrian's paintings



Mondrian's paintings : black and white version



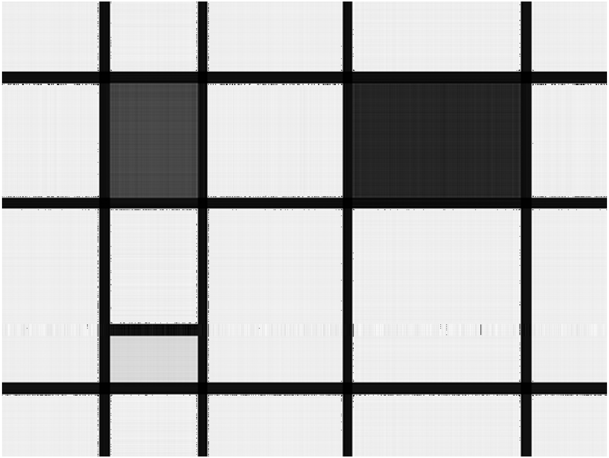
Mondrian's paintings : adding a Gaussian noise



Mondrian's paintings : SVD of X

```
> SVD = svd(X)
> SVD$d
  [1] 705.2966509 147.2653999  40.3910584   6.1156832
  [5]   5.9563710   5.8878358   5.8747333   5.8484616
 [10]   5.8154512   5.7746063   5.7607876   5.7512011
 [15]   5.7245453   5.7007277   5.6785695   5.6556892
 ...
[765]   0.4789888   0.4684564   0.4519168   0.4447403
> SVD$d[4:768] = 0
> Mhat = SVD$u %*% diag(SVD$d) %*% t(SVD$v)
```

Mondrian's paintings : adding a Gaussian noise



Rank selection

Theorem

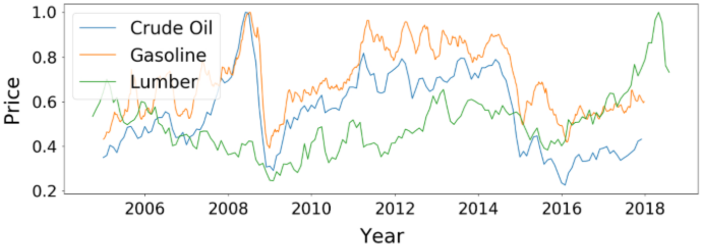
For some universal constant $c > 0$ (see the paper) we define

$$\widehat{r} = \arg\min_{k} \left\{ \|\widehat{M}_k - X\|_F^2 + c\sigma^2 k(d+T+s) \right\}.$$

Then, if $\mathrm{rank}(M) = r$ unknown, with probability at least $1 - 2\exp(-s)$,

$$\|\widehat{M}_{\widehat{r}} - M\|_F^2 \leq C\sigma^2 r(d+T+s).$$

Multivariate time series matrices



$$X = \begin{pmatrix} x_{1,1} & x_{1,2} & \dots \\ x_{2,1} & x_{2,1} & \dots \\ x_{3,1} & x_{3,1} & \dots \end{pmatrix}$$

Shih, S.-Y., Sun, F.-K. and Lee, H.-Y. (2019). Temporal pattern attention for multivariate time series forecasting. *Machine Learning*.

Theory for time series

This time, we allow the rows of $\mathcal{E}$ to be $\mathcal{N}(0, \Sigma)$.

Theorem

For any $s > 0$, with probability at least $1 - 2\exp(-s)$,

$$\|\widehat{M}_k - M\|_F^2 \leq C \left\{ \inf_{\mathrm{rank}(A) \leq k} \|A - M\|_F^2 + \|\Sigma\|_{\mathrm{op}} k(d + T + s) \right\}$$

where $\|\Sigma\|_{\mathrm{op}} = \sup_x \|\Sigma x\| / \|x\|$ is the operator norm of Σ.

Examples :

- i.i.d noise, $\Sigma = \sigma^2 I$, then $\|\Sigma\|_{\mathrm{op}} = \sigma^2$.
- autoregressive noise $\mathcal{E}_{i,t+1} = \rho \mathcal{E}_{i,t} + \eta_{i,t}$, $\mathrm{Var}(\eta_{i,t}) = \sigma^2$, then

$$\Sigma = \sigma^2 \begin{pmatrix} 1 & \rho & \dots & \rho^{T-1} \\ \rho & 1 & \dots & \rho^{T-2} \\ \vdots & \ddots & \ddots & \vdots \\ \rho^{T-1} & \dots & \rho & 1 \end{pmatrix} \text{ and } \|\Sigma\|_{\mathrm{op}} = \sigma^2 \frac{1 + |\rho|}{1 - |\rho|}.$$

Contents

A matrix completion model

Matrix completion model

We observe, for $\ell = 1, \ldots, n$

- (i_ℓ, j_ℓ) drawn uniformly on $\{1, \ldots, d\} \times \{1, \ldots, T\}$,
- $X_\ell = M_{i_\ell, j_\ell} + \mathcal{E}_\ell$,

where

- M is a $d \times T$ matrix (to recover),
- $\mathcal{E}_\ell$ is some noise.

Estimators

- Rank constrained estimation :

$$\widehat{M}_k = \underset{\mathrm{rank}(A) \leq k}{\arg\min} \sum_{\ell=1}^{n} (X_\ell - A_{i_\ell, j_\ell})^2 .$$

- Nuclear-norm penalized estimator :

$$\widetilde{M}_\lambda = \arg\min_{A} \left\{ \sum_{\ell=1}^{n} (X_\ell - A_{i_\ell, j_\ell})^2 + \lambda \|A\|_1 \right\} .$$

Both estimators are implemented in R in the SoftImpute package.

Collaborative filtering / recommender systems

				
Claire	4	?	3	$\ldots$
Nial	?	4	?	$\ldots$
Brendon	2	?	4	$\ldots$
Andrew	?	4	?	$\ldots$
Adrian	1	?	?	$\ldots$
Pierre	?	5	?	$\ldots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$

Netflix Prize

Home | Rules | Leaderboard | Update | Download

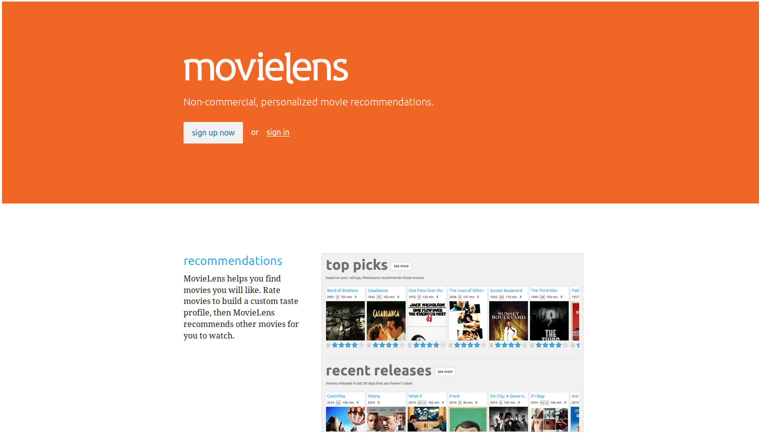
Leaderboard

Showing Test Score. Click here to show quiz score

Display top 20 leaders.

Rank	Team Name	Best Test Score	% Improvement	Best Submit Time
Grand Prize - RMSE = 0.8567 - Winning Team: BellKor's Pragmatic Chaos				
1	BellKor's Pragmatic Chaos	0.8567	10.06	2009-07-26 18:18:28
2	The Ensemble	0.8567	10.06	2009-07-26 18:38:22
3	Grand Prize Team	0.8582	9.90	2009-07-10 21:24:40
4	Opera Solutions and Vandelay United	0.8588	9.84	2009-07-10 01:12:31
5	Vandelay Industries !	0.8591	9.81	2009-07-10 00:32:20
6	PragmaticTheory	0.8594	9.77	2009-06-24 12:06:56
7	BellKor in BigChaos	0.8601	9.70	2009-05-13 08:14:09
8	Dace_	0.8612	9.59	2009-07-24 17:18:43

http ://movielens.org



Movies in MovieLens 100K

1682 movies

```
1|Toy Story (1995)|01-Jan-1995||http://us.imdb.com/M/title-exact?Toy%20Story%20(1995)|0|0|0|1|1|1|0
2|GoldenEye (1995)|01-Jan-1995||http://us.imdb.com/M/title-exact?GoldenEye%20(1995)|0|1|1|0|0|0|0|0
3|Four Rooms (1995)|01-Jan-1995||http://us.imdb.com/M/title-exact?Four%20Rooms%20(1995)|0|0|0|0|0|0
4|Get Shorty (1995)|01-Jan-1995||http://us.imdb.com/M/title-exact?Get%20Shorty%20(1995)|0|1|0|0|0|1
5|Copycat (1995)|01-Jan-1995||http://us.imdb.com/M/title-exact?Copycat%20(1995)|0|0|0|0|0|0|1|0|1|0
6|Shanghai Triad (Yao a yao yao dao waipo qiao) (1995)|01-Jan-1995||http://us.imdb.com/Title?Yao+a+
7|Twelve Monkeys (1995)|01-Jan-1995||http://us.imdb.com/M/title-exact?Twelve%20Monkeys%20(1995)|0|0
8|Babe (1995)|01-Jan-1995||http://us.imdb.com/M/title-exact?Babe%20(1995)|0|0|0|0|1|1|0|0|1|0|0|0|0
9|Dead Man Walking (1995)|01-Jan-1995||http://us.imdb.com/M/title-exact?Dead%20Man%20Walking%20(199
10|Richard III (1995)|22-Jan-1996||http://us.imdb.com/M/title-exact?Richard%20III%20(1995)|0|0|0|0|
11|Seven (Se7en) (1995)|01-Jan-1995||http://us.imdb.com/M/title-exact?Se7en%20(1995)|0|0|0|0|0|0|1|
12|Usual Suspects, The (1995)|14-Aug-1995||http://us.imdb.com/M/title-exact?Usual%20Suspects,%20The
13|Mighty Aphrodite (1995)|30-Oct-1995||http://us.imdb.com/M/title-exact?Mighty%20Aphrodite%20(1995
14|Postino, Il (1994)|01-Jan-1994||http://us.imdb.com/M/title-exact?Postino,%20Il%20(1994)|0|0|0|0|
15|Mr. Holland's Opus (1995)|29-Jan-1996||http://us.imdb.com/M/title-exact?Mr.%20Holland's%20Opus%2
16|French Twist (Gazon maudit) (1995)|01-Jan-1995||http://us.imdb.com/M/title-exact?Gazon%20maudit%
17|From Dusk Till Dawn (1996)|05-Feb-1996||http://us.imdb.com/M/title-exact?From%20Dusk%20Till%20Da
18|White Balloon, The (1995)|01-Jan-1995||http://us.imdb.com/M/title-exact?Badkonake%20Sefid%20(199
19|Antonia's Line (1995)|01-Jan-1995||http://us.imdb.com/M/title-exact?Antonia%20(1995)|0|0|0|0|0|0
20|Angels and Insects (1995)|01-Jan-1995||http://us.imdb.com/M/title-exact?Angels%20and%20Insects%2
21|Muppet Treasure Island (1996)|16-Feb-1996||http://us.imdb.com/M/title-exact?Muppet%20Treasure%20
22|Braveheart (1995)|16-Feb-1996||http://us.imdb.com/M/title-exact?Braveheart%20(1995)|0|1|0|0|0|0|
```

Users in MovieLens 100K

943 users

```
1|24|M|technician|85711
2|53|F|other|94043
3|23|M|writer|32067
4|24|M|technician|43537
5|33|F|other|15213
6|42|M|executive|98101
7|57|M|administrator|91344
8|36|M|administrator|05201
9|29|M|student|01002
10|53|M|lawyer|90703
11|39|F|other|30329
12|28|F|other|06405
13|47|M|educator|29206
```

Reading the data with SoftImpute

```
> X = read.table("u.data",header=FALSE)
> X
     V1   V2 V3        V4
1   196  242  3 881250949
2   186  302  3 891717742
3    22  377  1 878887116
4   244   51  2 880606923
5   166  346  1 886397596
....
> A = Incomplete(i=X$V1,j=X$V2,x=X$V3)
```

Incomplete matrices

```
> A
 [1,] 5 3 4 3 3 5 4 1 5 3 2 5 5 5 5 5 3 4 5 4
 [2,] 4 . . . . . . . . 2 . . 4 4 . . . . 3 .
 [3,] . . . . . . . . . . . . . . . . . . . .
 [4,] . . . . . . . . . . 4 . . . . . . . . .
 [5,] 4 3 . . . . . . . . . . . . . . 4 . . .
 [6,] 4 . . . . . 2 4 4 . . 4 2 5 3 . . . 4 .
 [7,] . . . 5 . . 5 5 5 4 3 5 . . . . . . . .
 [8,] . . . . . . 3 . . . 3 . . . . . . . . .
 [9,] . . . . . 5 4 . . . . . . . . . . . . .
[10,] 4 . . 4 . . 4 . 4 . 4 5 3 . . 4 . . . .
```

The *softImpute* function

$$\widehat{M} \leftarrow \min_{A} \left\{ \sum_{i,j} (A_{i,j} - X_{i,j})^2 + \lambda \|A\|_1 \right\}$$

How it works :

```
> B = softImpute(A, rank.max = ..., lambda = ...,
                 type = x("als","svd"), ... )
> i = ...
> j = ...
> impute(B,i,j)
```

Example

```
> A
[1,] 5 3 4 3 3 5 4 1 5 3 2 5 5 5 5 5 3 4 5 4
[2,] 4 . . . . . . . . 2 . . 4 4 . . . . 3 .
[3,] . . . . . . . . . . . . . . . . . . . .
[4,] . . . . . . . . . . 4 . . . . . . . . .

> B=softImpute(A,rank.max=5,lambda=0,type="svd")
> impute(B,3,1)
[1] 1.361311
> impute(B,4,1)
[1] 2.48563
```

How to get many entries

```
> A
[1,] 5 3 4 3 3 5 4 1 5 3 2 5 5 5 5 5 3 4 5 4
[2,] 4 . . . . . . . . 2 . . 4 4 . . . . 3 .
[3,] . . . . . . . . . . . . . . . . . . . .
[4,] . . . . . . . . . . 4 . . . . . . . . .

> B=softImpute(A,rank.max=5,lambda=0,type="svd")
> i = c(3,3,3,4,4,4)
> j = c(1,2,3,1,2,3)
> impute(B,i,j)
[1] 1.3432932 0.3631449 1.0265744 2.7306180
    0.1185635 1.8732696
```

How to get the full matrix

```
> B=softImpute(A,rank.max=5,lambda=0,type="svd")
> Y = complete(A,B)
```

Be careful, it can be larger than the memory of your laptop !

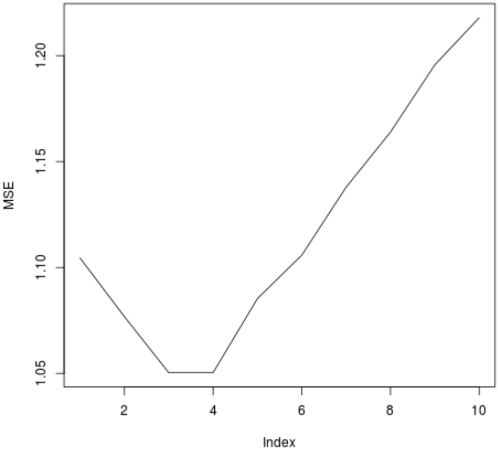
Completed MovieLens matrix

```
> round(Y)
     [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] ...
[1,]    5    3    4    3    3    5    4    1    5
[2,]    4    1    2    2    0    3    4    3    5
[3,]    1    0    1    2    1    1    2    1    3
[4,]    3    0    2    2    1    3    4    2    5
...
> A
[1,] 5 3 4 3 3 5 4 1 5 3 2 5 5 5 5 5 3 4 5 4
[2,] 4 . . . . . . . . 2 . . 4 4 . . . . 3 .
[3,] . . . . . . . . . . . . . . . . . . . .
[4,] . . . . . . . . . . 4 . . . . . . . . .
```

MSE by rank

```
> data = X[1:80000,]
> A = Incomplete(i=data$V1,j=data$V2,x=data$V3)
> test = X[80001:100000,]
> MSE = c()
> for (k in 1:10)
> {
>  B = softImpute(A,rank.max=k,lambda=0,type="svd")
>  pred = impute(object=B,i=test$V1,j=test$V2)
>  MSE = c(MSE,mean((pred-test$V3)^2))
> }
```

MSE by rank



Minimax rate of estimation

Theorem

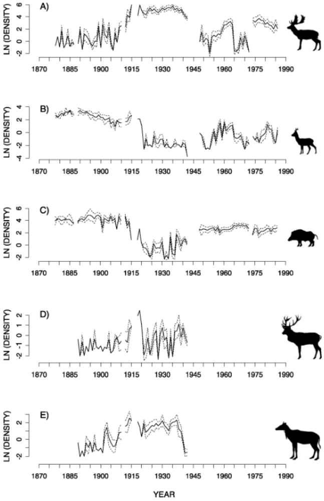
Assume the noise (ε_ℓ) is i.i.d $\mathcal{N}(0, \sigma^2)$ (sub-exponential also works). For a well chosen λ that does not depend on $r = \mathrm{rank}(M)$, and under minimal assumptions on M, with large probability

$$\frac{1}{dT} \left\| \widetilde{M}_\lambda - M \right\|_F^2 \leq C \frac{\sigma^2 r(d+T) \log(d+T)}{n}.$$

Moreover, this rate is minimax-optimal (up to the log term).

Koltchinskii, V., Lounici, K. and Tsybakov, A. (2011). Nuclear-norm penalization and optimal rates for noisy low-rank matrix completion. *The Annals of Statistics*.

Multivariate time series



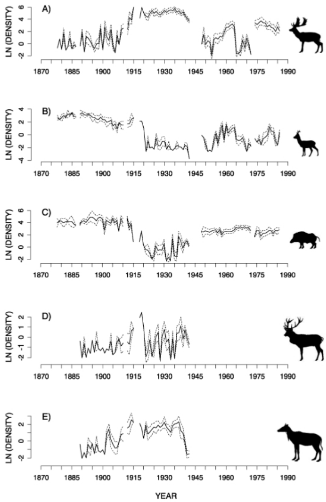
Imperio, S. *et al.* (2010). Investigating population dynamics in ungulates : Do hunting statistics make up a good index of population abundance ? *Wildlife Biology*.

- multivariate series
- correlations
- noisy observations
- missing entries

Examples

- econometrics : panel data with missing entries,
- industry : data from sensors at multiple locations,
- ecology : spatial data with observations from a few sites only at each date,
- ...
- more generally, any situation where we have multivariate time series and each measurement is expensive.

Assumptions



- low-rank trend :

$$M = \underbrace{U}_{d \times k} \underbrace{V}_{k \times T}$$

- temporal correlated noise ε :

$$\varepsilon_{i,t} \text{ indep. } \varepsilon_{j,t'} \ (i \neq j)$$

$$\varepsilon_{i,t} \text{ not indep. } \varepsilon_{i,t'}$$

- (i_ℓ, t_ℓ) i.i.d uniform, ξ_ℓ observation noise :

$$X_\ell = M_{i_\ell, t_\ell} + \varepsilon_{i_\ell, t_\ell} + \xi_\ell.$$

Reminder on ϕ-mixing

Given two σ-algebras $\mathcal{A}$ and $\mathcal{B}$,

$$\phi(\mathcal{A},\mathcal{B}) = \sup\Big\{ |\mathbb{P}(A) - \mathbb{P}(A|B)| : (A,B) \in \mathcal{A}\times\mathcal{B},\ \mathbb{P}(B) > 0 \Big\}.$$

Given a stationary time series $S = (S_t)_{t\in\mathbb{Z}}$,

$$\phi_S(h) = \phi\Big(\sigma(\dots, S_{t-2}, S_{t-1}, S_t), \sigma(S_{t+h}, S_{t+h+1}, S_{t+h+2}, \dots)\Big).$$

The series S is said to be ϕ-mixing iff

$$\sum_{h=0}^{\infty} \phi_S(h) < +\infty.$$

More assumptions

$$X_\ell = M_{i_\ell,t_\ell} + \varepsilon_{i_\ell,t_\ell} + \xi_\ell.$$

- $M = \underbrace{U}_{d\times r}\underbrace{V}_{r\times T}$ and $|U_{i,h}|, |V_{h,t}| \leq c_{U,V}/\sqrt{r}$.
- (i_ℓ, t_ℓ) i.i.d uniform on $\{1,\dots,d\} \times \{1\dots,T\}$;
- $(\varepsilon_{i,t})_{t=1,\dots,T}$ is a bounded, ϕ-mixing time series :

$$|\varepsilon_{i,t}| \leq m_\varepsilon \text{ and } \sum_{t=1}^{\infty} \phi_{\varepsilon_{i,\cdot}}(t) \leq \Phi_\varepsilon.$$

- (ξ_ℓ) are i.i.d, sub-exponential variables : for $k \geq 2$,

$$\mathbb{E}(|\xi_\ell|^q) \leq \frac{v_\xi c_\xi^{q-2} q!}{2}.$$

Estimator and risk bound

$$\widehat{M}_k = \arg\min_{\underbrace{A}_{d\times T} = \underbrace{W}_{d\times k}\underbrace{Z}_{k\times T}} \sum_{\ell=1}^{n} (X_\ell - A_{i_\ell, j_\ell})^2.$$

Theorem

With probability at least $1 - \exp(-s)$, as soon as $k \geq r$,

$$\frac{1}{dT} \left\| \widehat{M}_k - M \right\|_F^2 \leq C \frac{k(d+T)\log(n) + s}{n}$$

where $C = C(c_{U,V}, m_\varepsilon, \Phi_\varepsilon, v_\xi, c_\xi)$ is known.

Remarks on the proof

1. decompose the difference between *empirical risk* and *expected risk* $\frac{1}{n}\sum_{\ell=1}^{n}(Y_\ell - X_{i_\ell,j_\ell})^2 - \frac{1}{dT}\sum_{i,j}(M_{i,j} - X_{i,j})^2$ in elementary terms.
2. some of these terms are sums of i.i.d variables. Bound them via Bernstein inequality. Some are sums of ϕ-mixing variables, use :

Samson, P.-M. (2000). Concentration of measure inequalities for Markov chains and Φ-mixing processes. *The Annals of Probability*.

3. union bound.

REMARK : if the $\varepsilon_{i,\cdot}$ satisfy another notion of mixing or weak-dependence, we can use alternative versions of Bernstein inequality but this lead to slower rates of convergence, in $1/\sqrt{n}$.

Rank selection

$$\hat{r} = \mathop{\arg\min}_{1 \leq k \leq \min(d,T)} \left\{ \frac{1}{n} \sum_{\ell=1}^{n} ((\widehat{M}_k)_{i_\ell, j_\ell} - X_\ell)^2 + c \frac{k(d+T)\log(n)}{n} \right\}$$

where $c = c(c_{U,V}, m_\varepsilon, \Phi_\varepsilon, v_\xi, c_\xi)$ is **known** but too large.

In practice : we use the slope heuristic to calibrate a better c.

Theorem

With probability at least $1 - \exp(-s)$,

$$\frac{1}{dT} \left\| \widehat{M}_{\hat{r}} - M \right\|_F^2 \leq C \frac{r(d+T)\log(n) + s}{n}.$$

What about Bayes ?

Review of low-rank inducing priors, and theoretical study for denoising / matrix regression :

Alquier, P. (2013). Bayesian Methods for Low-rank Matrix Estimation : Short Survey and Theoretical Study. *International conference on Algorithmic Learning Theory (ALT)*.

For matrix completion :

Mai, T. T. and Alquier, P. (2015). A Bayesian Approach for Matrix Completion : Optimal Rate under General Sampling Distribution. *Electronic Journal of Statistics*.