Structured low-rank [matrix] approximation a useful tool for signal processing

Konstantin Usevich, CRAN (Nancy), CNRS/Univ. Lorraine

24.03.2022, RICOCHET Kick-off

Structured low-rank [matrix] approximation

a useful tool for  ?

Konstantin Usevich, CRAN (Nancy), CNRS/Univ. Lorraine

24.03.2022, RICOCHET Kick-off

Overview

Low-rank approximation

Hankel matrices

Sylvester matrices

Algorithms

Overview

Low-rank approximation

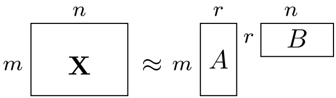
Hankel matrices

Sylvester matrices

Algorithms

Low-rank approximation

Given $\mathbf{X} \in \mathbb{R}^{m\times n}$ (or $\mathbb{C}^{m\times n}$) find its best rank-r approximation



Lots of applications, extensions...

Geometrically: project $\mathbf{X}$ on the set $\mathcal{M}_r$ of rank-$\leq r$ matrices

Singular value decomposition

(compact) singular value decomposition (SVD) ($m \leq n$):

$$\underset{m}{\overset{n}{\boxed{\boldsymbol{X}}}} = \sigma_1 \mathbf{u}_1 \mathbf{v}_1^T + \cdots + \sigma_m \mathbf{u}_m \mathbf{v}_m^T = \boxed{\boldsymbol{U}} \boxed{\begin{matrix}\sigma_1 & & \\ & \ddots & \\ & & \sigma_m\end{matrix}} \boxed{\boldsymbol{V}^T}$$

where

- $\boldsymbol{U}^T\boldsymbol{U} = \boldsymbol{V}^T\boldsymbol{V} = \boldsymbol{I}$ — semi-orthogonal matrices of singular vectors
  $\boldsymbol{U} = \begin{bmatrix}\mathbf{u}_1 & \cdots & \mathbf{u}_m\end{bmatrix}, \quad \boldsymbol{V} = \begin{bmatrix}\mathbf{v}_1 & \cdots & \mathbf{v}_m\end{bmatrix}$
- $\sigma_1 \geq \cdots \geq \sigma_m \geq 0$ — singular values

Properties:

- Exists and unique (except cases when $\sigma_k = \sigma_{k+1}$)
- Statistical interpretation: PCA
- Signal processing: subspace methods

SVD and low-rank approximation

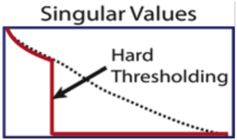
Theorem (Eckart-Young[-Mirsky[-Schmidt]]): best rank-r approximation

$$\min_{\widehat{\mathbf{X}}} \|\mathbf{X} - \widehat{\mathbf{X}}\| \quad \text{subject to} \quad \operatorname{rank} \widehat{\mathbf{X}} \leq r$$

in any unitarily invariant norm $\|\cdot\|$ is given by

$$\mathrm{tSVD}_r(\boldsymbol{X}) = \sigma_1 \cdot \begin{array}{c} | \\ \mathbf{u}_1 \\ | \end{array} \text{—}\, \mathbf{v}_1^T \,\text{—} + \cdots + \sigma_r \cdot \begin{array}{c} | \\ \mathbf{u}_r \\ | \end{array} \text{—}\, \mathbf{v}_r^T \,\text{—} = \boxed{\boldsymbol{U}_{1:r,:}} \boxed{\begin{matrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_r \end{matrix}} \boxed{\boldsymbol{V}_{1:r,:}^T}$$

truncated SVD



SVD: algorithms

Full SVD of $\mathbf{X} \in \mathbb{R}^{m \times n}$, $m \leq n$: $\mathcal{O}(m^2 n)$ ([Golub, Reinsh, 1970])

Truncated SVD (r first eigenvalues/vectors):

- Iterative (e.g., Lanzos) algorithms (e.g., [Simon, 1984])
- Randomized SVD [Halko, Martinsson, Tropp, 2011]

Bouth roughly $\boxed{\mathcal{O}(Mr)}$, where $M =$ cost of matrix-vector product $(\mathbf{X}\mathbf{v})$
:

- full matrices: $M = mn$
- for structured (sparse) matices
  - smaller M (e.g., $M = n\log(n)$ for Hankel)
  - do not need to store the whole $\mathbf{X}$

Cross approximation

If the matrix is huge or expensive to compute:
CUR (cross, or pseudo-skeleton) approximation: take size-r subsets $\mathcal{I}, \mathcal{J}$

$$\widehat{\mathbf{X}}_{\mathcal{I},\mathcal{J}} = \mathbf{X}_{:,\mathcal{J}} (\mathbf{X}_{\mathcal{I},\mathcal{J}})^{-1} \mathbf{X}_{\mathcal{I},:}$$

Advantages:

- Need a small portion (cross) of the matrix ($O(r(m+n))$)
- Quasi-optimality (thm. in [Goreinov, Tyrtyshnikov, 2001])

$$\|\widehat{\mathbf{X}}_{maxvol} - \mathbf{X}\|_{max} \leq (r+1)\sigma_r(\mathbf{X})$$

- iterative or randomized strategies to select $\mathcal{I}, \mathcal{J}$

Extensions of the basic problem

$$\min_{\substack{\widehat{\mathbf{X}}=\mathbf{A}\mathbf{B} \\ \mathbf{A}\in\mathbb{R}^{m\times r},\mathbf{B}\in\mathbb{R}^{r\times n}}} \|\mathbf{X}-\widehat{\mathbf{X}}\|_F$$

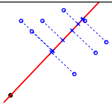
- Other norms (e.g. $\|\mathbf{X}\|_W^2 = \sum_{i,j} W_{ij} X_{ij}^2$), missing data
- Constraints on the factors (e.g., nonnegative A, B — NMF)
- Constraint on the matrix:
  structured $\widehat{\mathbf{X}}$ — structured low-rank approximation (next slide)

Structured low-rank approximation

Problem (SLRA). Given a structured matrix $\mathbf{X} \in \mathcal{S}$

$$\underset{\widehat{X}}{\text{minimize}} \ \|\mathbf{X} - \widehat{\mathbf{X}}\|_W^2 \quad \text{subject to} \quad \widehat{\mathbf{X}} \in \mathcal{S} \text{ and } \operatorname{rank} \widehat{\mathbf{X}} \leq r$$

Data $\approx$ low-complexity model

structure $\mathcal{S}$	approximation problem
unstructured	fit by r-dim. subspace 
Hankel	fitting by complex exponentials
block-Hankel	linear system identification model reduction
Sylvester	approx. greatest common divisor
generalized Vandermonde	fit set of points by algebraic hypersurfaces

Overview

Low-rank approximation

Hankel matrices

Sylvester matrices

Algorithms

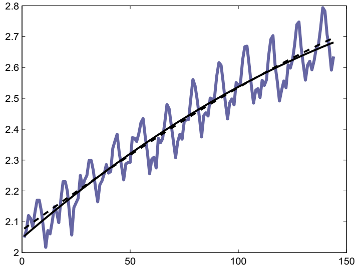
Hankel and Toeplitz matrices

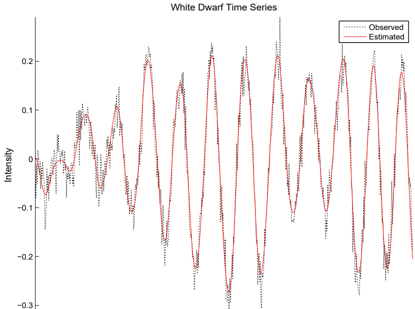
- $$\begin{bmatrix} p_0 & p_1 & \cdots & p_{K-1} & p_K \\ p_{-1} & p_0 & \ddots & \ddots & p_{K-1} \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ p_{1-K} & \ddots & \ddots & p_0 & p_1 \\ p_{-K} & p_{1-K} & \cdots & p_{-1} & p_0 \end{bmatrix}$$ Toeplitz
- $$\mathscr{H}_L(\mathbf{f}) := \begin{bmatrix} f_1 & f_2 & \cdots & f_K \\ f_2 & f_3 & \iddots & f_{K+1} \\ \vdots & \iddots & \iddots & \vdots \\ f_L & f_{L+1} & \cdots & f_N \end{bmatrix}$$ Hankel

  built from vector

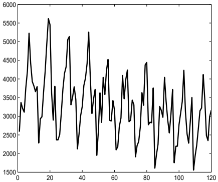
$$\mathbf{f} = \begin{bmatrix} f_1 & f_2 & \cdots & f_N \end{bmatrix}^T, \quad N = K + L - 1$$

Hankel SLRA in time series analysis: examples

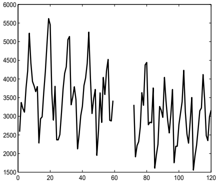




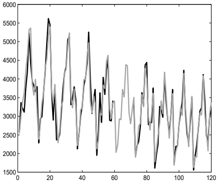
Hankel SLRA in time series analysis: examples



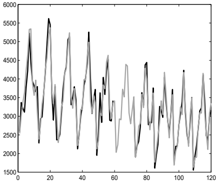
(a) Complete series



(b) Missing data



(c) SVD



(d) AP

Figure: Monthly volumes of fortified wine sales in Australia from January 1980 until January 1990 (black), with rank-11 approximations (grey)

Low-rank Hankel matrices and sums of exponentials

Theorem (Prony, Sylvester, ...)

$$f_n = \sum_{k=1}^{r} c_k \lambda_k^n, \qquad c_k, \lambda_k \in \mathbb{C} \quad \Rightarrow \quad \text{rank}\, \mathscr{H}(\mathbf{f}) \le r$$

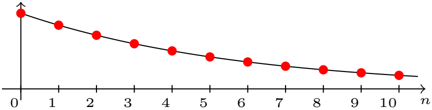
"Proof": For $F = (1, \lambda, \dots, \lambda^{N-1})$,

$$\mathscr{H}(\mathbf{f}) = \begin{bmatrix} 1 & \lambda & \cdots & \lambda^{N-L} \\ \lambda & \lambda^2 & \iddots & \lambda^{N-L+1} \\ \vdots & \iddots & \iddots & \vdots \\ \lambda^{L-1} & \lambda^L & \cdots & \lambda^{N-1} \end{bmatrix} = \begin{bmatrix} 1 \\ \lambda \\ \vdots \\ \lambda^{L-1} \end{bmatrix} \begin{bmatrix} 1 & \lambda & \cdots & \lambda^{N-L} \end{bmatrix}$$

Sums of complex exponentials (real case)

$r = 1, \quad c, \lambda \in \mathbb{R},$

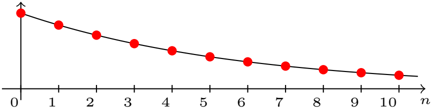
$f_n = c\lambda^n$



Sums of complex exponentials (real case)

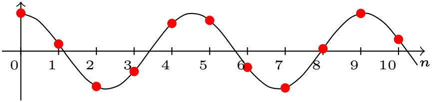
$r = 1, \quad c, \lambda \in \mathbb{R},$

$f_n = c\lambda^n$



$r = 2, \quad \lambda_1, \lambda_2 = e^{\pm i 2\pi\omega},$

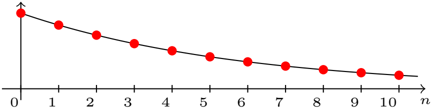
$f_n = \lambda_1^n + \lambda_2^n = \cos(2\pi(\omega n + \phi))$



Sums of complex exponentials (real case)

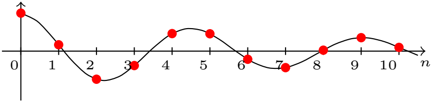
$r = 1, \quad c, \lambda \in \mathbb{R},$

$f_n = c\lambda^n$



$r = 2, \quad \lambda_1, \lambda_2 = \rho e^{\pm i 2\pi\omega},$

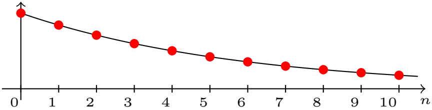
$f_n = \lambda_1^n + \lambda_2^n = \rho^n \cos(2\pi(\omega n + \phi))$



Sums of complex exponentials (real case)

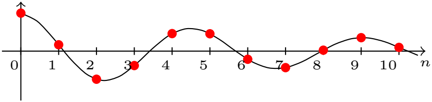
$r = 1, \quad c, \lambda \in \mathbb{R},$

$f_n = c\lambda^n$



$r = 2, \quad \lambda_1, \lambda_2 = \rho e^{\pm i 2\pi\omega},$

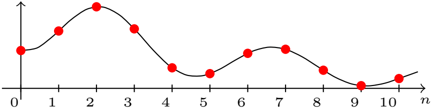
$f_n = \lambda_1^n + \lambda_2^n = \rho^n \cos(2\pi(\omega n + \phi))$



$r = 3$

$\lambda_1, \lambda_2 = \rho e^{\pm i 2\pi\omega}, \lambda_3 \in \mathbb{R}$

$f_n = \lambda_1^n + \lambda_2^n + c\lambda_3^n$



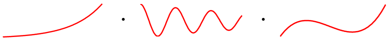
Low-rank Hankel matrices: general case

Theorem (Heinig, Rost, 1984).

$$\operatorname{rank} \begin{bmatrix} f_1 & f_2 & \cdots & f_K \\ f_2 & f_3 & \iddots & f_{K+1} \\ \vdots & \iddots & \iddots & \vdots \\ f_L & f_{L+1} & \cdots & f_{K+L-1} \end{bmatrix} \le r < \min(K, L)$$

if (and only if)

$$\mathbf{f} = \sum \Big(\quad \cdot \quad \cdot \quad \Big)$$



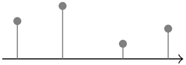
sum of products of exponentials/sines, polynomials $f_n = \sum_{k=1}^{m} p_k(n)\lambda_k^n$

follows from the **linear recurrence** (left kernel)

$$\theta_1 f_{k+1} + \cdots + \theta_L f_{k+L} = 0,$$

Unit-modulus case

Superresolution (Candès): recovery of sums of Diracs $\sum_{k=1}^{r} c_k \delta(t-\omega_k)$



from "Fourier" measurements:

$$p_n = \sum_{k=1}^{r} c_k \exp(i2\pi\omega_k n), c_k > 0 \qquad (1)$$

Theorem (Carathéodory): The Toeplitz matrix

$$\begin{bmatrix} p_0 & p_1 & \cdots & p_{K-1} & p_K \\ p_{-1} & p_0 & \ddots & \ddots & p_{K-1} \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ p_{1-K} & \ddots & \ddots & p_0 & p_1 \\ p_{-K} & p_{1-K} & \cdots & p_{-1} & p_0 \end{bmatrix}$$

is H.p.d. of rank $r < K \iff \mathbf{p}$ has the form (1)

Overview

Low-rank approximation

Hankel matrices

Sylvester matrices

Algorithms

Approximate GCD

Given two polynomials

$$A(z) = \sum_{k=0}^{L} a_k z^k, \quad B(z) = \sum_{k=0}^{L} b_k z^k$$

find their (approximate) greatest common divisor

Subproblem: for fixed degree D find the closest

$$\widehat{A}(z) = P(z)H(z), \widehat{B}(z) = Q(z)H(z),$$

such that $\deg H(z) = D$

Can be reformulated as a Sylvester low-rank approximation

Example # 1: blind deconvolution

Polynomial multiplication $\leftrightarrow$ convolution

$$C(z) = A(z)B(z) \leftrightarrow \begin{bmatrix} c_0 \\ c_1 \\ c_2 \\ \vdots \\ c_{N+D} \end{bmatrix} = \begin{bmatrix} a_0 & & \\ \vdots & \ddots & \\ a_N & & a_0 \\ & \ddots & \vdots \\ & & a_N \end{bmatrix} \begin{bmatrix} c_0 \\ \vdots \\ c_D \end{bmatrix} = \mathbf{a} * \mathbf{b}.$$

Blind deconvolution: recover $\mathbf{u}$ convolved by two different (FIR) filters

$$\begin{matrix} \mathbf{u} * \mathbf{a} = \mathbf{y}_1 \\ \mathbf{u} * \mathbf{b} = \mathbf{y}_2 \end{matrix} \iff \gcd(Y_1(z), Y_2(z)) = U(z)$$

Example # 2: polarized phase retrieval (PPR)

- Classic (1D) phase retrieval = recovery of a signal $\mathbf{x} \in \mathbb{C}^N$ from

$$y_m = \left|\mathbf{a}_m^H \mathbf{x}\right|, m = 0, 1, \dots M-1, \quad \text{Fourier measurements}$$

$\mathbf{a}_m \in \mathbb{C}^N$ — discrete Fourier vector for frequency $\frac{2\pi m}{M}$, $M \geq 2N-1$

solution essentially **nonunique**

Example # 2: polarized phase retrieval (PPR)

- Classic (1D) phase retrieval = recovery of a signal $\mathbf{x} \in \mathbb{C}^N$ from

$$y_m = \left|\mathbf{a}_m^H \mathbf{x}\right|, m = 0, 1, \ldots M-1, \quad \text{Fourier measurements}$$

  $\mathbf{a}_m \in \mathbb{C}^N$ — discrete Fourier vector for frequency $\frac{2\pi m}{M}$, $M \geq 2N-1$

  solution essentially **nonunique**

- PPR: recover $\mathbf{X} \in \mathbb{C}^{N\times 2}$ from measurements

$$y_{m,p} = \left|\mathbf{a}_m^H \mathbf{X}\mathbf{b}_p\right|^2, m = 0, 1, \ldots M-1, \quad p = 0, 1, \ldots P-1,$$

  where $\mathbf{b}_p \in \mathbb{C}^2$ are P projection vectors ($P \geq 4$)

  (Flamant, U., Clausel, Brie, 2022): unique solution (under mild conditions)

Example # 2: polarized phase retrieval (PPR, contd.)

Measurement

$$y_{m,p} = \left|\mathbf{a}_m^H \mathbf{X} \mathbf{b}_p\right|^2 = \langle \mathfrak{F}[m], \mathbf{b}_p \mathbf{b}_p^{\mathsf{H}} \rangle$$

where

$$\mathfrak{F}[m] = \mathbf{X}^H \mathbf{a}_m \mathbf{a}_m^H \mathbf{X} \quad \text{Fourier covariances}$$

recovery of $\mathbf{X} = \begin{bmatrix} \mathbf{x}_1 & \mathbf{x}_2 \end{bmatrix}$ from $y_{m,p} \leftrightarrow$ recovery of $X_1(z), X_2(z)$ from:

$$\Gamma_{ij}(z) = z^{N-1} X_i(z) \overline{X}_j(z^{-1}), \quad i,j \in \{1,2\} \quad \text{autocorrelation polynomials}$$

$$X_1(z) = \gcd(\Gamma_{11}(z), \Gamma_{12}(z)), \quad X_2(z) = \gcd(\Gamma_{21}(z), \Gamma_{22}(z))$$

Sylvester matrices

Theorem. (Sylvester)

$$\operatorname{rank} \underbrace{\left[\begin{array}{cccc|cccc} a_0 & & & & b_0 & & & \\ \vdots & \ddots & & & \vdots & \ddots & & \\ a_L & & & a_0 & b_L & & & b_0 \\ & \ddots & & \vdots & & \ddots & & \vdots \\ & & & a_L & & & & b_L \end{array}\right]}_{\mathcal{S}(A,B)\in\mathbb{C}^{2L\times 2L} \text{ Sylvester matrix}} = 2L - \underbrace{\deg\gcd(A(z),B(z))}_{\text{rank defect}}$$

In particular, $\mathcal{S}(A,B)$ nonsingular $\iff$ $A(z)$, $B(z)$ are coprime.

$\Rightarrow$ GCD can be found from left or right kernel of $\mathcal{S}(A,B)$

Right kernel of reduced Sylvester matrix

Theorem. Let $A(z) = F(z)H(z), B(z) = G(z)H(z)$ ($H(z)$ is gcd of degree D). Then all the solutions $(\mathbf{u}, \mathbf{v}) \in (\mathbb{C}^{D+1})^2$ to

$$\underbrace{\left[\begin{array}{ccc|ccc} a_0 & & & b_0 & & \\ \vdots & \ddots & & \vdots & \ddots & \\ a_L & & a_0 & b_L & & b_0 \\ & \ddots & \vdots & & \ddots & \vdots \\ & & a_L & & & b_L \end{array}\right]}_{\mathcal{S}_D(A,B) \in \mathbb{C}^{(2L-D+1)\times 2(L-D+1)}} \begin{bmatrix} \mathbf{u} \\ \mathbf{v} \end{bmatrix} = 0;$$

are given (up to scaling c) by

$$U(z) = -cG(z), \quad V(z) = cF(z),$$

Simplest (SVD) algorithm: use last right singular vector of $\mathcal{S}_D(A, B)$.

Example of recovery

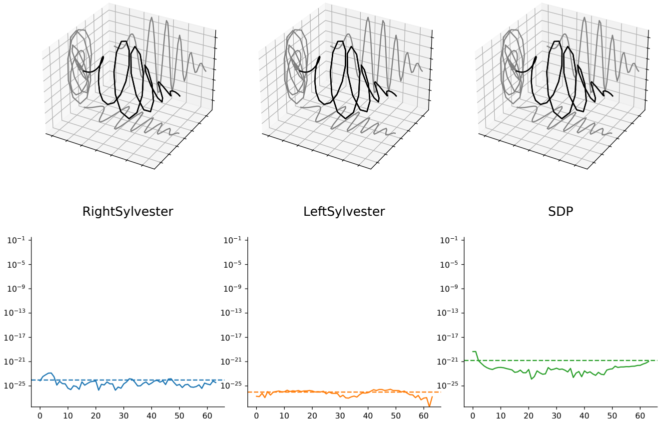
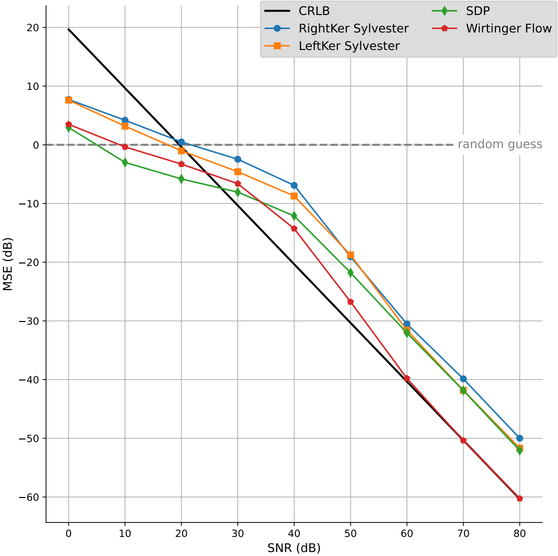


Figure: Reconstruction of a bivariate pulse ($N = 64$) from noiseless PPR measurements ($M = 2N - 1, P = 4$)

Noisy PPR reconstruction ($N = 32,\ M = 2N-1,\ P = 4$)



Overview

Low-rank approximation

Hankel matrices

Sylvester matrices

Algorithms

Weighted (unstructured) LRA

$$\min_{\widehat{\mathbf{X}}} \|\mathbf{X} - \widehat{\mathbf{X}}\|_W^2 \quad \text{subject to} \quad \widehat{\mathbf{X}} \in \mathcal{M}_r$$

where

$$\|\mathbf{X}\|_W^2 = \sum_{i,j=1}^{m,n} \mathbf{X}_{i,j}^2 W_{i,j}, \quad W_{i,j} \in [0; +\infty] \quad \text{weighted semi-norm}$$

Extreme cases:

- fixed values: $W_{i,j} = \infty \longleftrightarrow$ constraint $\mathbf{X}_{i,j} = \widehat{\mathbf{X}}_{i,j}$
- missing values: $W_{i,j} = 0 \longleftrightarrow \widehat{\mathbf{X}}_{i,j}$ is not important

Weighted (unstructured) LRA

$$\min_{\widehat{\mathbf{X}}} \|\mathbf{X} - \widehat{\mathbf{X}}\|_W^2 \quad \text{subject to} \quad \widehat{\mathbf{X}} \in \mathcal{M}_r$$

where

$$\|\mathbf{X}\|_W^2 = \sum_{i,j=1}^{m,n} \mathbf{X}_{i,j}^2 W_{i,j}, \quad W_{i,j} \in [0; +\infty] \quad \text{weighted semi-norm}$$

Extreme cases:

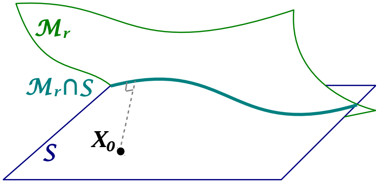
- fixed values: $W_{i,j} = \infty \longleftrightarrow$ constraint $\mathbf{X}_{i,j} = \widehat{\mathbf{X}}_{i,j}$
- missing values: $W_{i,j} = 0 \longleftrightarrow \widehat{\mathbf{X}}_{i,j}$ is not important

Complexity

- $W_{ij} \equiv 1$ (or $\operatorname{rank} W = 1$) $\rightarrow$ solution by SVD
- In general case, no closed form solution: N. Gillis, F. Glineur, **Low-Rank Matrix Approximation with Weights or Missing Data is NP-hard**, *SIMAX*, 2011.

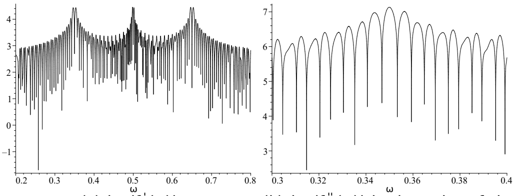
(Hankel) structured low-rank approximation

$$\boxed{\begin{aligned} &\underset{\widehat{X}}{\text{minimize}} \ \|\widehat{X} - X_0\|_W^2 \\ &\text{subject to } \ \widehat{X} \in \mathcal{M}_r \cap \mathcal{S} \end{aligned}}$$



Nasty nonconvex problem:

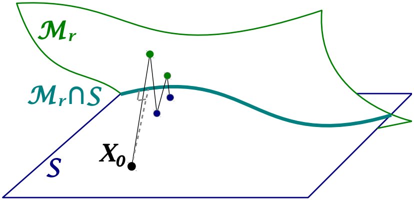
- [Ottaviani, Spaenlehauer, Sturmfels, 2013]: $\#$ of stationary points $= O(N^r)$
- [Gillard, Zhigljavsky, 2013]: Lipschitz constant of $f(\omega) = \sum_{n=1}^{N} (f_n - sin(2\pi\omega n))$



(a) $\log|f'(\omega)|$ (b) $\log|f''(\omega)|$ in the region of the global minimizer $\omega^{(0)} = 0.35$

Cadzow's algorithm

minimize $_{\widehat{X}}$ $\|\widehat{X} - X_0\|_W^2$
subject to $\widehat{X} \in \mathcal{M}_r \cap \mathcal{S}$



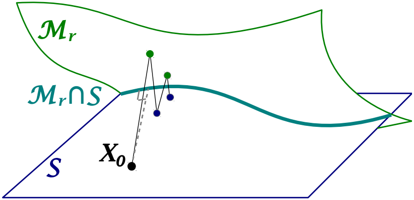
Alternating projections (Cadzow, 1988):

1. $X_{k+1} = \mathrm{proj}_{\mathcal{M}_r}(X_k)$ — truncated singular value decomposition
2. $X_{k+2} = \mathrm{proj}_{\mathcal{S}}(X_{k+1})$ — projection on a linear subspace

Properties:

- Linear convergence rate to $\mathcal{M}_r \cap \mathcal{S}$
- Does not minimize $\|\widehat{X} - X_0\|_W$

Cadzow's algorithm: more details



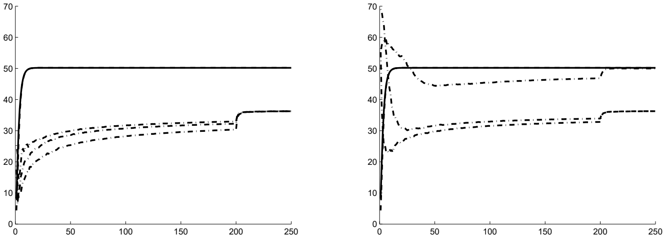
- $X_{k+1} = \mathrm{proj}_{\mathcal{M}_r}(X_k)$
  truncated singular value decomposition
- $X_{k+2} = \mathrm{proj}_{\mathcal{S}}(X_{k+1})$
  projection on a linear subspace (diagonal averaging)

  both steps $\boxed{O(rN \log N)}$
- one step of SLRA: singular spectrum analysis (SSA) (Golyandina et al, 2001), subspace-based methods: good guess thanks to separability (signal/noise, signal/signal)

Modifications of Cadzow

to try to minimize $\|\mathbf{X} - \widehat{\mathbf{X}}\|_W$

- [Condat, Hirabayashi, 2015]: proximal methods
- [Andersson et al, 2014]: ADMM-type methods
- [Schost, Spaelenhauer, 2016]: Newton-type methods
- [Gillard, Zhigljavsky, 2015]: stochastic optimization (randomization + backtracking)



(a) $s_0 = 0.25$ (b) $s_0 = 0.9$

Figure: (from [GZ2015]): plots of the Frobenius distances $\| - X_0 - \widehat{X}_k\|_F$ as functions of k for AP (bold) and three Multistart APBR iterations (dashed).

Kernel representation and variable projection

rank constraint $\leftrightarrow$ kernel form

$$\operatorname{rank} \widehat{\mathbf{X}} \leq r \quad \Longleftrightarrow \quad \begin{array}{c} \\ d \end{array}\overset{m}{\boxed{R}} \cdot \begin{array}{c} \\ m \end{array}\overset{n}{\boxed{\quad \widehat{\mathbf{X}} \quad}} = 0, \qquad d := m - r$$

full row rank

$$\underset{\substack{R \in \mathbb{R}^{d \times m} \\ \text{rank } R = d}}{\text{minimize}} \; f(R), \quad f(R) := \underbrace{\left(\min_{\widehat{\mathbf{X}} \in \mathcal{S}} \|\mathbf{X} - \widehat{\mathbf{X}}\|_w^2 \;\; \text{subject to } R\widehat{\mathbf{X}} = 0 \right)}_{\text{least-norm problem}}$$

Variable projection: [Markovsky et al., 2006], [U., Markovsky, 2014]

- Fast $O(d^3 m^2 n)$ Gauss-Newton/Levenberg-Marquardt iterations for mosaic Hankel matirces and $w \in (0; +\infty]$
- Allows for various constraints on kernel (known poles, etc.)
- Software: `http://slra.github.io/`

Rank minimization and convex relaxation

Convex relaxation:

$$\min_{\widehat{X}\in\mathcal{S}} \|\widehat{X}\|_* + \lambda\|X - \widehat{X}\|_F^2, \quad \overbrace{\|\widehat{X}\|_* := \sum \text{ of singular values of } \widehat{X}}^{\textbf{nuclear norm}}$$

[Recht et al, 2008]: theoretical guarantees but for random structure!

Rank minimization and convex relaxation

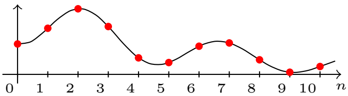
Convex relaxation:

$$\min_{\widehat{X} \in \mathcal{S}} \|\widehat{X}\|_* + \lambda \|X - \widehat{X}\|_F^2, \quad \overbrace{\|\widehat{X}\|_* := \sum \text{ of singular values of } \widehat{X}}^{\textbf{nuclear norm}}$$

[Recht et al, 2008]: theoretical guarantees but for random structure!

Example: time series forecasting $(h_0, h_1, \ldots, h_{N-1}, ?, ..., ?)$, $\quad h_k \in \mathbb{R}$

$$\min \left\| \begin{bmatrix} h_0 & h_1 & \cdots & \cdots & h_K \\ h_1 & h_2 & \iddots & \iddots & \vdots \\ \vdots & \iddots & \iddots & \iddots & h_{N-1} \\ \vdots & \iddots & \iddots & \iddots & \vdots \\ h_L & \cdots & h_{N-1} & \cdots & h_{N-1+p} \end{bmatrix} \right\|_*$$



[U, Comon, 2016], [Gillard, U., 2018]: works only if exponentials sufficiently damped

Conclusions

- SLRA - flexible framework (hopefully useful for RICOCHET!)
- Plenty of algorithms and things we can do

References:

- `http://www.gipsa-lab.fr/summerschool/slra2015/`
- SLRA: J. Gillard, K. U., *Hankel low-rank approximation and completion in time series analysis and forecasting: a brief review*, Statistics and Its Interface, 2022.
- Phase retrieval + Sylvester: J. Flamant, K. U., M. Clausel, D. Brie, *Polarimetric phase retrieval: uniqueness and algorithms*, `hal-03613352v1` + GRETSI.

Conclusions

- SLRA - flexible framework (hopefully useful for RICOCHET!)
- Plenty of algorithms and things we can do

References:

- `http://www.gipsa-lab.fr/summerschool/slra2015/`
- SLRA: J. Gillard, K. U., *Hankel low-rank approximation and completion in time series analysis and forecasting: a brief review*, Statistics and Its Interface, 2022.
- Phase retrieval + Sylvester: J. Flamant, K. U., M. Clausel, D. Brie, *Polarimetric phase retrieval: uniqueness and algorithms*, `hal-03613352v1` + GRETSI.

GRETSI 2022 is in Nancy (submission still open!)