Probabilistic Low-Rank Matrix Completion with Adaptive Spectral Regularization Algorithms

Adrien.Todeschini@inria.fr, François.Caron@stats.ox.ac.uk & Marie.Chavent@u-bordeaux2.fr

Abstract

A novel class of algorithms for low rank matrix completion:

- Novel penalty functions on the singular values of the low rank matrix, using a mixture model representation.
- Suitable set of latent variables $\longrightarrow$ EM algorithm to obtain a MAP estimate of the completed matrix.

$\Rightarrow$ Iterative soft-thresholded SVD algorithm

$\Rightarrow$ Adapts the shrinkage coefficients associated to the singular values.

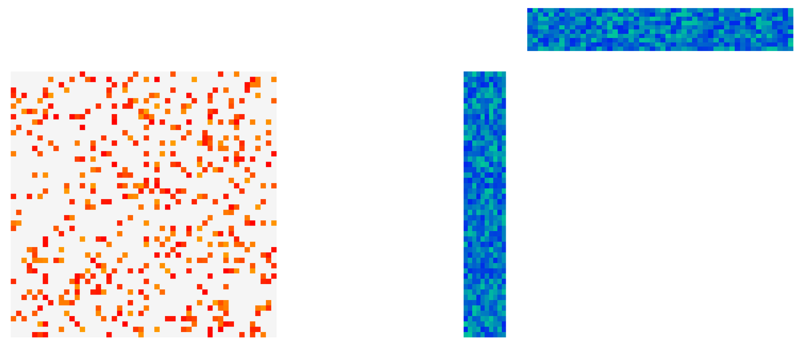
$\Rightarrow$ Simple to implement and can scale to large matrices.

- Good numerical results compared to recent alternatives.

Low-rank matrix completion

Objective: Complete the $m \times n$ matrix Z from a subset $(i,j) \in \Omega$ of noisy observations X_{ij} and assume Z can be approximated by a low rank factorization.

$$\underbrace{X}_{\text{sparse}} = P_\Omega(\underbrace{Z}_{\text{low rank}} + \underbrace{\varepsilon}_{\text{noise}})$$

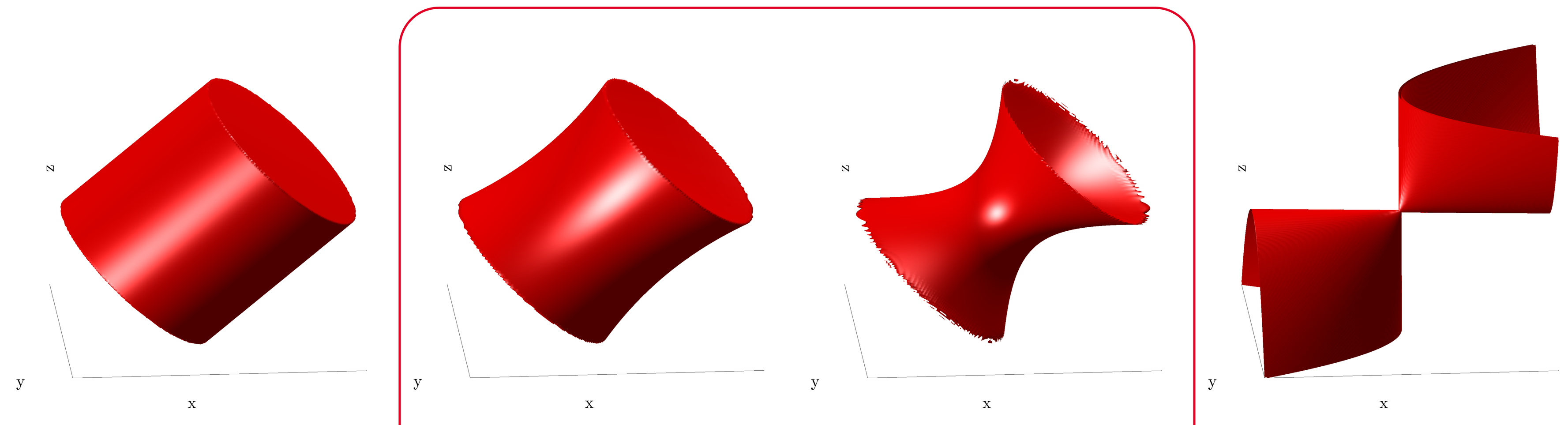


Hierarchical Adaptive Spectral Penalty (HASP)

MAP: $\underset{Z}{\text{maximize}}$ Log-likelihood + Log-prior

$\Leftrightarrow$ $\underset{Z}{\text{minimize}}$ $\frac{1}{2\sigma^2}\sum_{(i,j)\in\Omega}(X_{ij} - Z_{ij})^2 + pen(d)$

Balls of constant penalty for a 2×2 matrix $[x, y; y, z]$

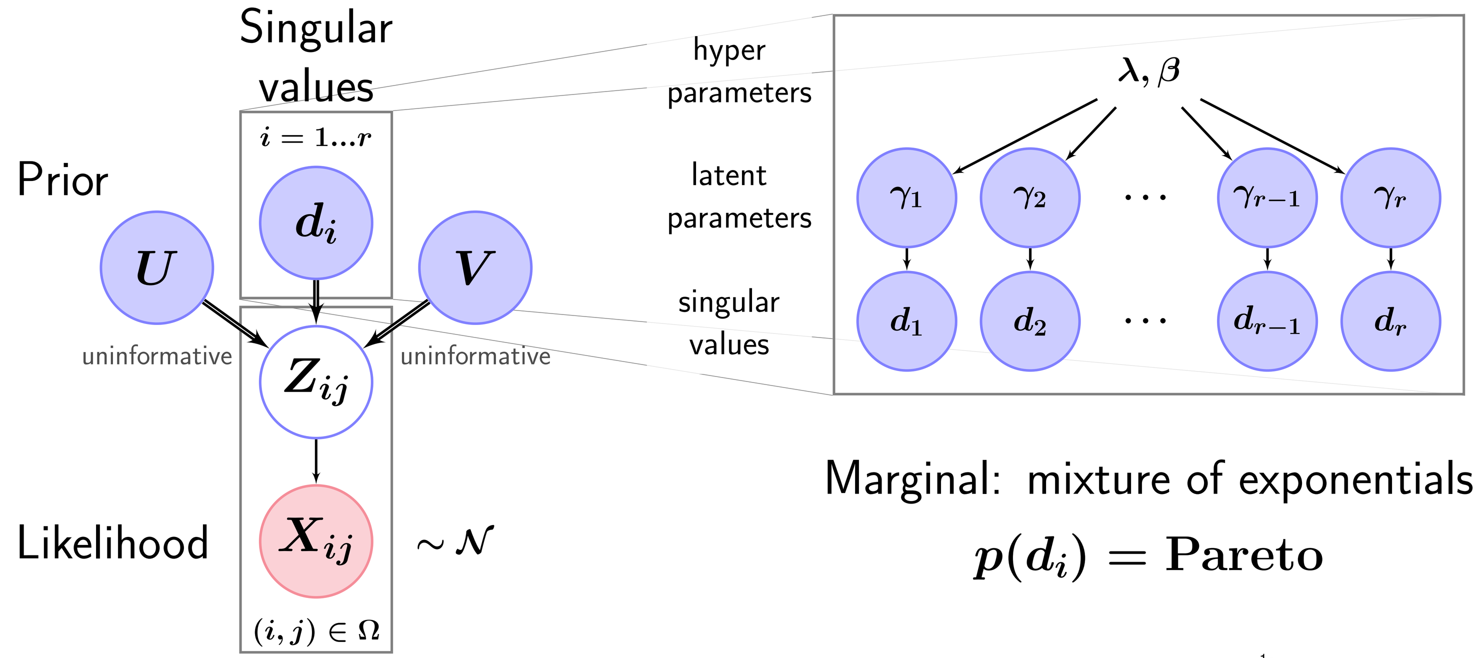


Nuclear norm	HASP ($\beta = 1$)	HASP ($\beta = 0.1$)	Rank penalty
$\lambda\Vert Z\Vert_*$	$(\lambda\beta + 1)\sum_{i=1}^r \log(\beta + d_i)$		$\lambda\text{rank}(Z)$
Convex $\rightarrow$ Soft thresholded SVD	Non convex $\rightarrow$ Weighted adaptive soft thresholded SVD		Non convex $\rightarrow$ Hard thresholded SVD

- Bridge the gap between the rank and the nuclear norm penalties.
- HASP recovers the nuclear norm when $\beta \rightarrow \infty$.

Bayesian model

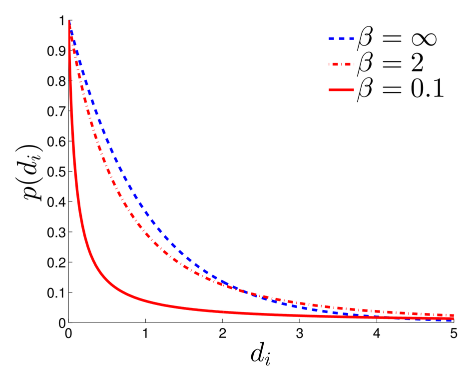
$$Z = UDV^T \text{ with } D = \begin{pmatrix} d_1 & & 0 \\ & \ddots & \\ 0 & & d_r \end{pmatrix}$$



Marginal: mixture of exponentials

$$p(d_i) = \text{Pareto}$$

- $\beta \rightarrow 0 \Rightarrow$ more concentrated at 0, heavier tails.
- $\beta \rightarrow \infty \Rightarrow$ recovers the exponential distribution.



$p(d_i)$ with $\lambda = 1$.

EM algorithm

- Exploit the mixture model representation.
- Use latent variables γ and the missing values $P_\Omega^\perp(X)$.

Algorithm 1 Hierarchical Adaptive Soft Impute (HASI)

Initialize $Z^{(0)}$ with Soft-Impute algorithm. At iteration $t \geq 1$:

- Impute the missing values: $X^* = P_\Omega(X) + P_\Omega^\perp(Z^{(t-1)})$
- Adapt the threshold coefficients of each singular value:

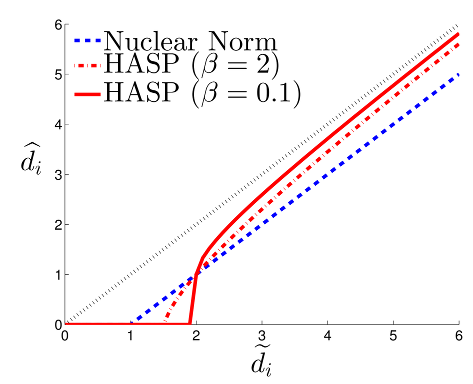
$$\text{For } i = 1, \ldots, r,\ \omega_i^{(t)} = \frac{\lambda\beta + 1}{\beta + d_i^{(t-1)}}$$

- Compute the weighted soft thresholded SVD of the completed matrix X^*:

$$Z^{(t)} = \text{S}_{\sigma^2\omega^{(t)}}(X^*) = \widetilde{U}\widetilde{D}_{\sigma^2\omega}\widetilde{V}^T \text{ with}$$

$$\widetilde{D}_\omega = \begin{pmatrix} (\widetilde{d}_1 - \omega_1)_+ & & 0 \\ & \ddots & \\ 0 & & (\widetilde{d}_r - \omega_r)_+ \end{pmatrix} \text{ and } X^* = \widetilde{U}\widetilde{D}\widetilde{V}^T \text{ is the SVD of } X^*.$$

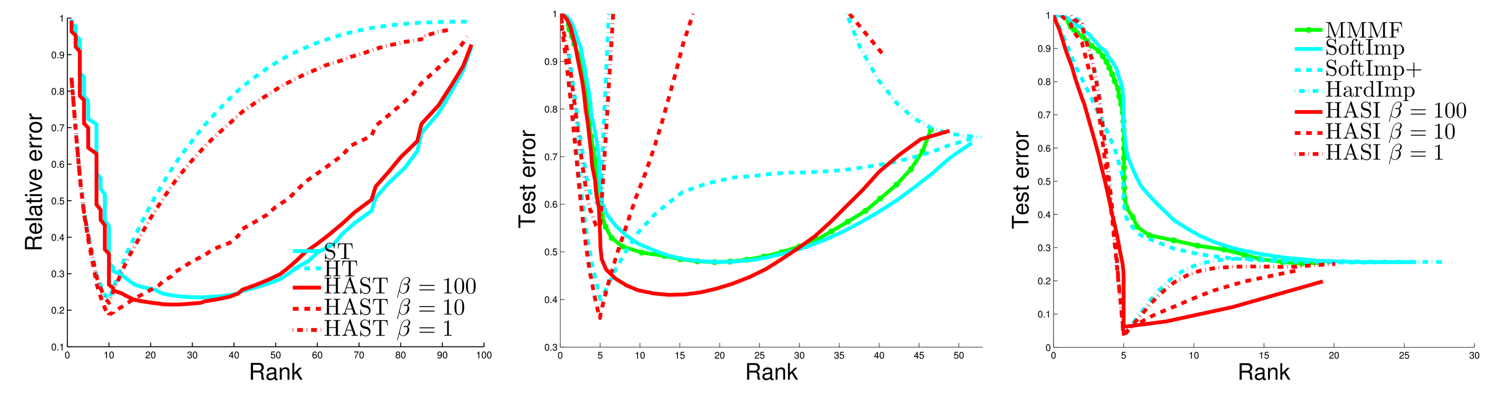
- HASP penalizes less heavily higher singular values $\Rightarrow$ Bias is reduced.
- HASI admits Soft-Impute as special case when $\beta \rightarrow \infty$.
- Initialization with Soft-Impute algorithm gives satisfactory results.
- Scaling: use PROPACK algorithm for computing the truncated SVD of large matrices.



Thresholding rules on the singular values with $\lambda = 1$

Experiments

- Simulated data

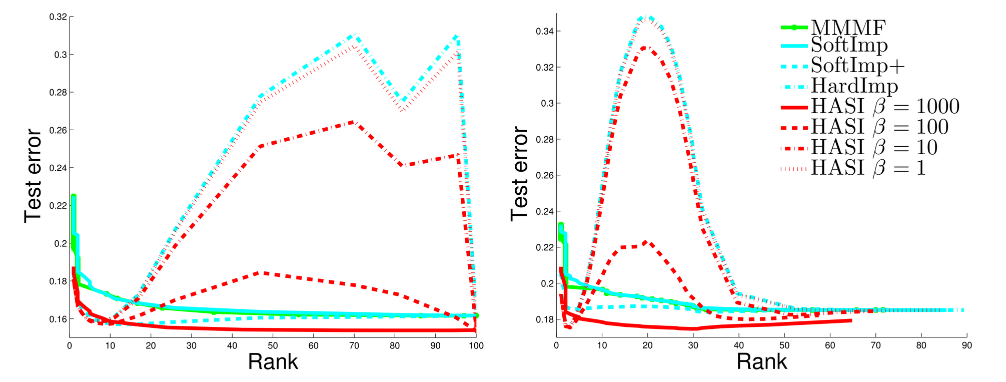


(a) SNR=1; Complete; rank=10 (b) SNR=1; 50% missing; rank=5 (c) SNR=10; 80% missing; rank=5

Figure : Test error w.r.t. the rank obtained by varying the value of the regularization parameter λ.

- Collaborative filtering examples

	Jester 1		Jester 2		Jester 3		MovieLens 100k		MovieLens 1M	
	24983×100		23500×100		24938×100		943×1682		6040×3952	
	27.5% miss.		27.3% miss.		75.3% miss.		93.7% miss.		95.8% miss.	
Method	NMAE	Rank	NMAE	Rank	NMAE	Rank	NMAE	Rank	NMAE	Rank
MMMF	0.161	95	0.162	96	0.183	58	0.195	50	0.169	30
Soft Imp	0.161	100	0.162	100	0.184	78	0.197	156	0.176	30
Soft Imp+	0.169	14	0.171	11	0.184	33	0.197	108	0.189	30
Hard Imp	0.158	7	0.159	6	0.181	4	0.190	7	0.175	8
HASI	0.153	100	0.153	100	0.174	30	0.187	35	0.172	27



(a) Jester 1 (b) Jester 3

Figure : NMAE on the test set of the (a) Jester 1 and (b) Jester 3 datasets.

- Low values of β: bimodal with modes at low rank and full rank.
- $\beta = 1000$: unimodal, outperforms Soft-Impute at any given rank.

Extensions

- Using a 3 parameters Generalized inverse Gaussian prior distribution for the parameters $\gamma_i \longrightarrow$ additional degree of freedom.
- Extension to binary matrices using a probit model.

References

- Gaïffas, S. and Lecué, G. (2011). Weighted algorithms for compressed sensing and matrix completion. arXiv preprint arXiv:1107.1638.
- Larsen, R. M. (2004). Propack-software for large and sparse svd calculations. *Available online. URL http://sun. stanford. edu/rmunk/PROPACK.*
- Mazumder, R., Hastie, T., and Tibshirani, R. (2010). Spectral regularization algorithms for learning large incomplete matrices. *The Journal of Machine Learning Research*, 11:2287–2322.
- Rennie, J. and Srebro, N. (2005). Fast maximum margin matrix factorization for collaborative prediction. In *Proceedings of the 22nd international conference on Machine learning*, pages 713–719. ACM.
- Srebro, N. and Jaakkola, T. (2003). Weighted low-rank approximations. In *NIPS*, volume 20, page 720.
- Todeschini, A., Caron, F., and Chavent, M. (2013). Probabilistic low-rank matrix completion with adaptive spectral regularization algorithms. In *Advances in Neural Information Processing Systems*, pages 845–853.

▷ https://sites.google.com/site/adrientodeschini/software/hasi