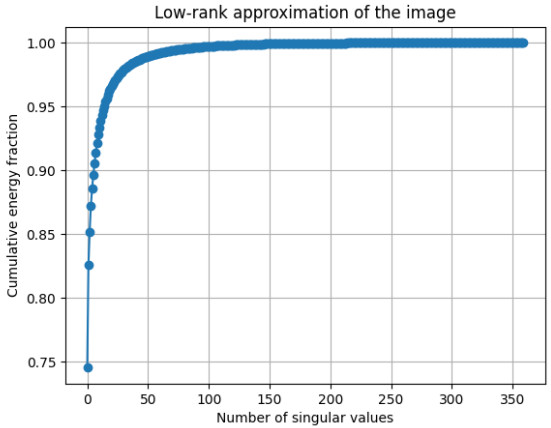
Project 3: Investigating Low-Rank Matrix Completion with Soft and Hard Imputation

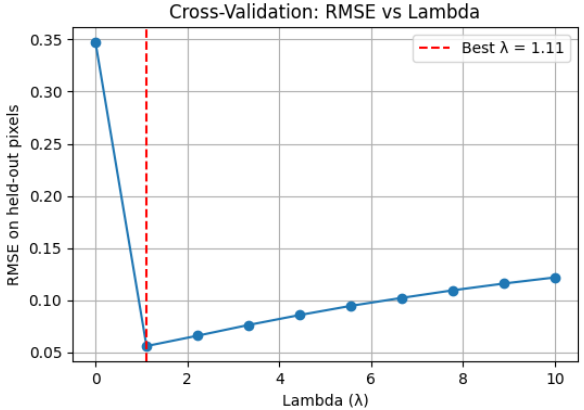
Course: Statistics and Machine Learning in High Dimensions (EEN100)

Group members: Alireza Bordbar, Amandus Reimer, Kiarash Rezaei

A.Y. 25/26

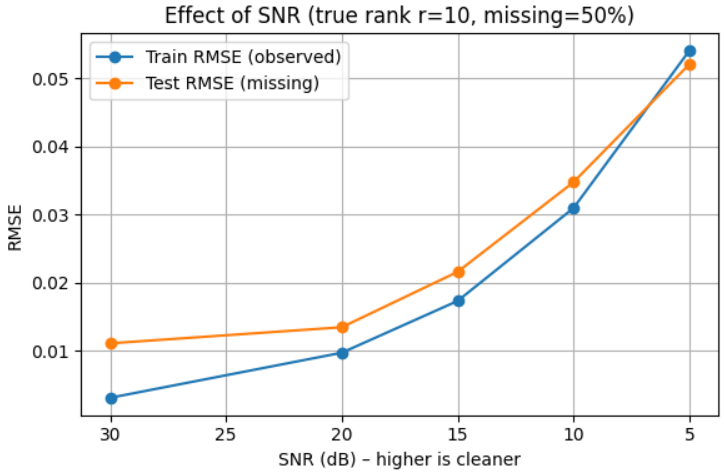
Soft Impute – Simple setting

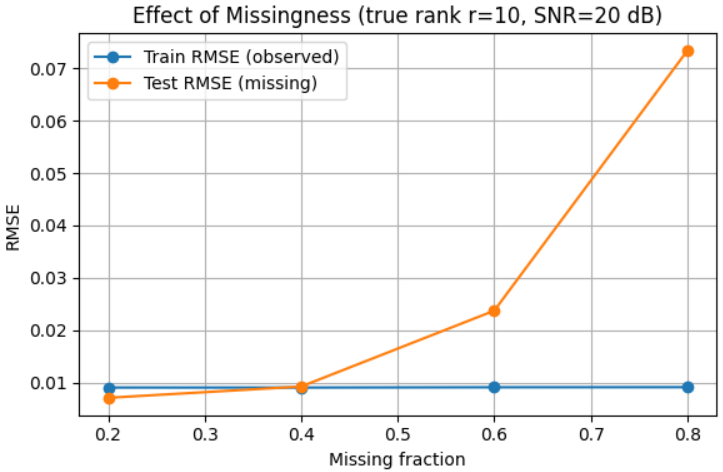




- SoftImpute recovers the main structure from 5% missing pixels and SNR=20 dB

SNR and Missingness Effect

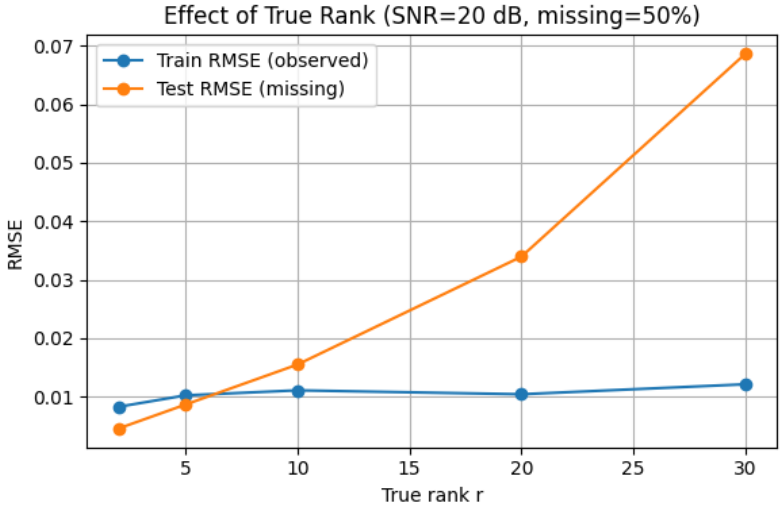


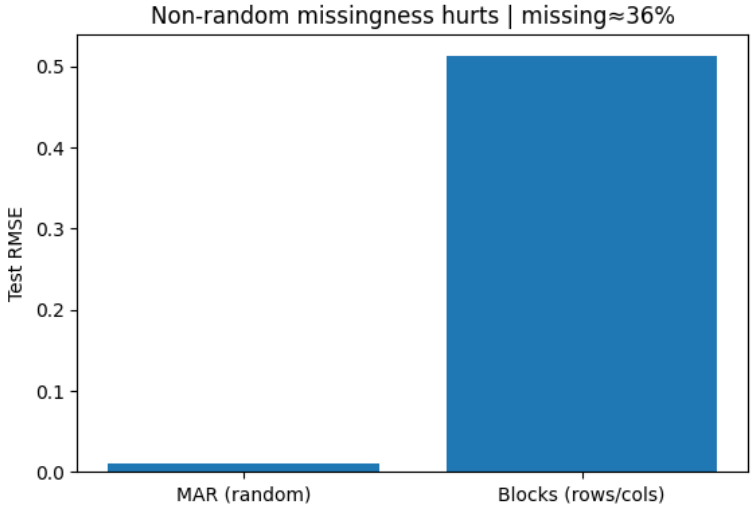


- SoftImpute works best when data are not dominated by noise.

- Performance degrades as missing rate increases.

Effect of True Rank and structured missing

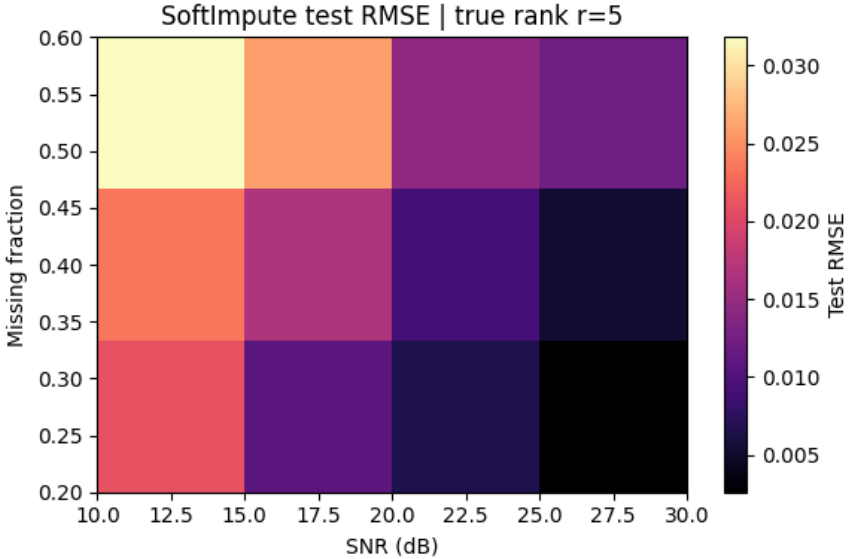




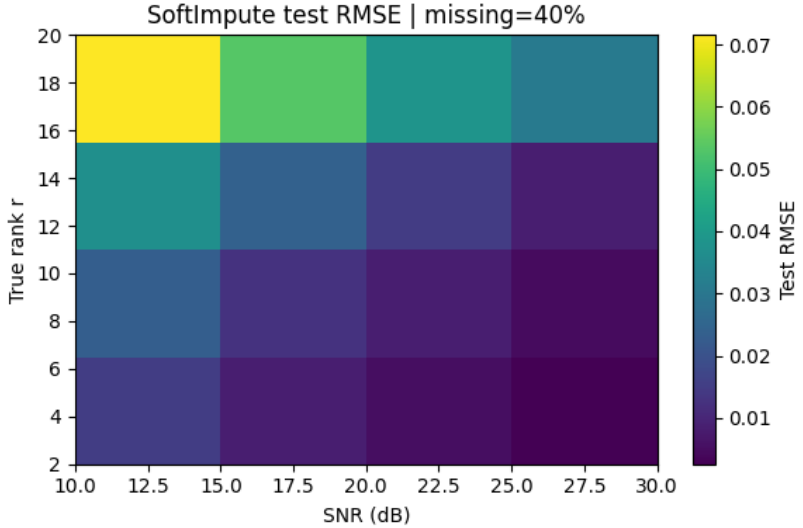
- We generated synthetic matrices with known rank r
- Confirms theory: low-rank assumption is crucial.

- when entire rows or columns vanish, there's no neighboring information to "borrow."

Soft Impute test RMSE heatmaps

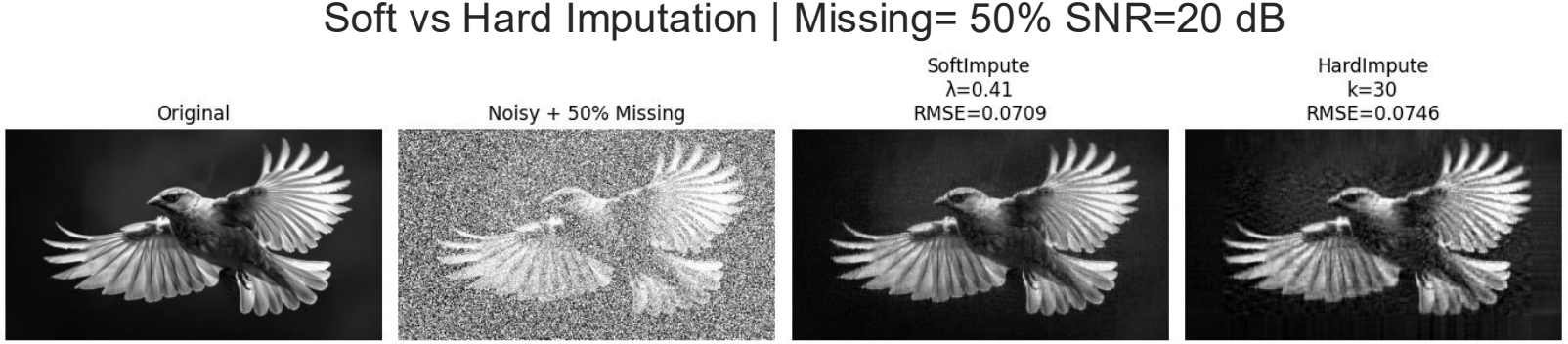


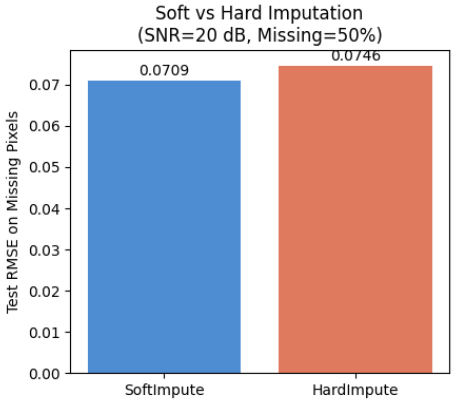
- This directly illustrates that SoftImpute's accuracy depends jointly on both the amount of missingness and the noise level.



- This confirms the "world is low-rank" idea: recovery is feasible only when structure is simple and data are clean.

Soft vs Hard Imputation: Who Wins?





- SoftImpute (λ-regularized) produces smoother denoising and slightly lower RMSE.
- HardImpute (top-k SVD) recovers crisper detail but can amplify small.

CHALMERS

UNIVERSITY OF TECHNOLOGY