IDENTIFYING FINANCIAL RISK FACTORS WITH A LOW-RANK/SPARSE DECOMPOSITION

Alexander D. Shkolnik

UC Berkeley

ads2@berkeley.edu

Risk Seminar. April 12th, 2016.

Joint work with Lisa Goldberg, UC Berkeley.

Factor models in finance

- Market model and CAPM
  - market equilibrium theory of asset prices under conditions of risk,
  - Sharpe (1964) and Treynor (1962).
- Arbitrage pricing theory and statistical models
  - the market portfolio plays no special role,
  - e.g. GNP and other factors (Ross 1976).
- Fundamental models
  - reverse the roles of the known and unknown variables to estimate factor returns; Rosenberg (1984) and Rosenberg (1984).

The factor model

- Returns to securities are driven by a relatively small number of risk factors, plus security specific returns: R is a $T \times N$ matrix

$$R = \psi Y + \epsilon \tag{1}$$

  - ψ is a $T \times K$ matrix of factor returns.
  - Y is a $K \times N$ matrix of factor exposures.
  - ϵ is a $T \times N$ matrix of security specific returns.

- N is the number of securities, T is the number of observations and K is (a relatively small) number of factors,
- To identify factors we rely on a universe of securities whose returns depend linearly on returns to factors.

Security covariance matrix

- In R the factor and specific returns ψ and ϵ are random while Y (the exposures) are to be estimated.

$$R = \psi Y + \epsilon \tag{2}$$

- Take ψ and ϵ to be mean zero and uncorrelated. Then the factor covariance matrix takes the form

$$\Sigma = Y^\top F Y + S \tag{3}$$

  - F is a factor covariance (diagonal if factors are uncorrelated),
  - S is the specific risk covariance matrix (diagonal if the specific returns are uncorrelated).

A taxonomy of risk factors

- **Broad:** market, equity styles, interest rates, credit
- **Thin:** industries, countries, currencies, credit
- Persistent: all of the above, in some cases
- Emerging or transient: new industries (internet), new sensitivities (housing bubble, climate), equity styles, liquidity, credit

Thin and broad risk factors

- If the factor and specific returns are uncorrelated:

$$\Sigma = L + S \tag{4}$$

  - $L = Y^\top F Y$ is a rank K matrix,
  - S is the specific risk covariance matrix.
- The K **broad factors** are contained in L,
  - i.e. factors that affect most/all of the securities.
- The **thin factors** may be found in S,
  - i.e. factors that affect a "thin" (smaller) number of securities.
  - (S diagonal means there are no thin factors)

Extracting risk factors

- Determine the matrices $L = Y^\top F Y$ and S in the decomposition

$$\Sigma = L + S\,. \tag{5}$$

- An elementary example (identifiability)
  - Suppose we observe R where $R = \psi + \epsilon$ and it is known that $\psi \sim N\left(0, \sigma_\psi\right)$ and $\epsilon \sim N\left(0, \sigma_\epsilon\right)$.
  - Is it possible to recover σ_ψ^2 and σ_ϵ^2 if we know only their sum?
- What conditions are required for L and S to be **identifiable**?
- There is **estimation error** (given sample covariance $\hat{\Sigma}$ only).

A number of approaches

- Fundamental models; requires many human analysts; factors are included as long as model designers think of them.
- Algebraic approach (Drton, Sturmfels & Sullivant 2007); has yet to yield scalable algorithms or addess practical issues.
- **ML factor analysis** (statistical model, latent factors):
  - Distribution assumptions address estimation errors.
  - Identifiability depends on initialization/constraints.
- **Dimensionality reduction** (PCA, distribution free):
  - Identifiability is inherently not an issue.
  - Estmation errors treated by regularization (asymptotic analysis).
- **Convex optimization:**
  - Low rank plus sparse decompositions

ML factor analysis

- We assume a distribution, e.g. let (R, ψ) be a Gaussian random vector in $\mathbb{R}^{N+K}$ with covariance matrix

$$\begin{bmatrix} \Sigma & Y^\top \\ Y & F \end{bmatrix} \tag{6}$$

  (R are observed variables; ψ are hidden (latent) variables).

- Σ may be decomposed as (Anderson 1968):

$$\Sigma = Y^\top F Y + S \qquad (S = \Sigma_{R\,|\,\psi})\,. \tag{7}$$

- Here, S is the conditional covariance (given the latent variables).
  - The fact that it is constant is a feature of the Gaussian distribution.
  - Plays the role of the specific covariance.

Some controversy

- Given the sample covariance $\hat{\Sigma}$, we maximize the likelihood

$$G(Y, F, S) = \log\det(Y^\top F Y + S)^{-1} - \mathrm{tr}\left((Y^\top F Y + S)^{-1}\hat{\Sigma}\right).$$

  Jöreskog (1967, grad), Rubin & Thayer (1982, EM algorithm), etc.

- Rubin & Thayer (1982) claim their EM algorithm found "multiple local maxima of the likelihood".

- Bentler & Tanaka (1983) point out "problems with EM algorithms for factor analysis":

  "Rather than highlight the limitations of ML factor analysis, the example demonstrates weaknesses in the EM algorithm ..."

- Rubin & Thayer (1983) respond that the EM algorithm can find local maxima, is complementary to other methods, etc ...

Decomposition multiplicity

- Letting $Q = (Y^\top F Y + S)^{-1}$ we may write $G(S, Y, F) = G(Q)$,

$$G(Q) = \log\det(Q) - \mathrm{tr}(Q\hat{\Sigma}).$$

- First order conditions (Petersen & Pedersen 2008) may be stated to prove the unique solution is $Q = \hat{\Sigma}^{-1}$.
- The identifiablity issues come in the multiplicity of decompositions

$$Q^{-1} = \hat{\Sigma} = L + S. \tag{8}$$

  - Multiplicity also occurs in $L = Y^\top F Y$, e.g. when F is diagonal $Y^\top F Y$ is unique upto multiplication by an orthogonal matrix.
  - Multiplicity of (8) persists even when $S = \Delta$, a diagonal matrix (setting of Rubin & Thayer (1982)).

Identifiability

- *"How much can we reduce the rank of a symmetric (positive definite) matrix by changing only its diagonal entries?"* (Shapiro 1982)

$$\hat{\Sigma} = L + \Delta \,. \tag{9}$$

  - No solution in the 1-dim case; a unique solution of rank 1 in the 2-dim case (4 equations and 4 unknowns).
  - Wilson & Worcester (1939) give example of a 6×6 correlation matrix and reduce it to rank 3 with two unique diagonals.
  - Basic analysis was carried out in Shapiro (1982), Shapiro (1985) and Shapiro & Ten Berge (2002).
  - Recently: Saunderson, Chandrasekaran, Parrilo & Willsky (2012).
- Even harder for $\hat{\Sigma} = L + S$ with non-zero off-diagonal entries in S.

ML factor analysis (summary)

- Under some assumed statistical model (typically Gaussian) we maximize a likelihood function.
  - In the Gaussian case the solution is the inverse of the sample covariance (the concentration or precision matrix).
- **Identification** issues arise when a certain structure (e.g. low rank plus diagonal) on the covariance matrix is assumed.
  - Hence, convergence may occur at multiple points.
  - More constraints are required to resolve the indeterminancy.
  - Implies assumptions on broad/thin factors.
- **Estimation error** issues are dealt with by assuming a statistical model (i.e. maximize the likelihood given observations).

Dimensionality reduction

- Seeks a low dimensional approximation of a given $\hat{\Sigma}$.
- Classical principal component analysis (PCA)

$$\begin{aligned} &\text{minimize} \quad \|\hat{\Sigma} - L\| && (10)\\ &\text{subject to} \quad \operatorname{rank}(L) \leq K && (11) \end{aligned}$$

- Solution given by truncated SVD

$$\hat{\Sigma} = U \Lambda U^\top; \qquad L = \sum_{i=1}^{K} \lambda_i u_i u_i^\top. \tag{12}$$

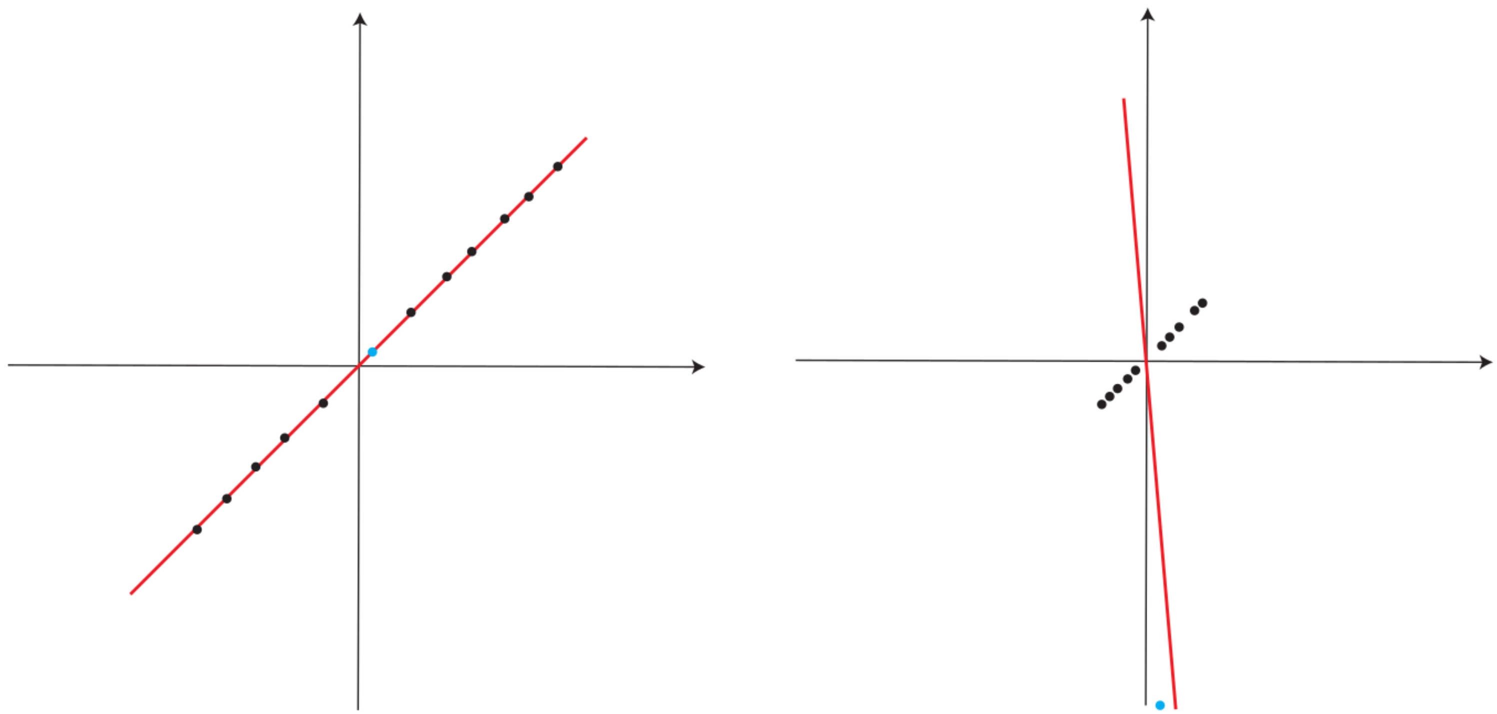
  extract components (factors) with largest variance.
- Many variants (LS, PFA, etc) (Jolliffe 2002).

Identifying factors (PCA)

- The output of PCA (truncated SVD) L retains the (broad) factors.
- The remainder may be diagonalized, i.e. $S = \text{diag}\,(\hat{\Sigma} - L)\,.$
  - Strict factor model (Ross 1976).
  - Justified (Gershgorin theorem) when off-diag entries are small.
- The remainder may be kept $S = \hat{\Sigma} - L$
  - Approximate factor model (Chamberlain & Rothschild 1982).
- No issues with **identification**:
  - The “largest” variance is always explained by the broad factors.
  - The thin factors (if any) always have smaller variance.

Estimation errors (PCA)

- PCA is very sensitive to outliers in $\hat{\Sigma}$.



Remedies for PCA sensitivity

- Robust PCA (not poly-time with performance guarantees)
  - multivariate trimming (Gnanadesikan & Kettenring 1972),
  - random sampling (Fischler & Bolles 1981),
  - influence function techniques (De La Torre & Black 2003)
  - alternating minimization (Ke & Kanade 2005),
- Covariance regularization (Bickel & Levina 2008):
  - approximate, thresholded covariance matrix for asymptotics:

$$\log(N)/T = o(1) .$$

- Random matrix theory (spectra) (Karoui 2008): $T \sim N$.

Traditional methods (summary)

- **ML factor analysis:**
  - address estimation error by requiring a statistical model,
  - do not deal well with identifiability (thin vs broad factors).
- **Dimentionality reduction (PCA):**
  - reduce identifiability to the claim that latent (broad) factors have the largest variance (picture for thin factors is unclear),
  - do not deal well with estimation error.
- It is possible to address both issues? (with or without distributional assumptions and imposing as little structure as possible.)

Low rank plus sparse decomposition

- Suppose the true covariance matrix satisfies

$$\Sigma = L + S \tag{13}$$

  - L is low rank (contains broad factors);
  - S is sparse (correlation due to thin factors).

- Given a sample covariance $\hat{\Sigma}$ we wish to recover estimates of (L, S).
  - recover (L, S) given the true covariance Σ?
  - approximate (L, S) given the sample covariance $\hat{\Sigma}$?

Is the problem well posed?

- Surprisingly, the answer is yes; but some assumptions are required on structure of the matrices L and S.
- Identifiability (eigenvectors of L and sparsity pattern of S)
  - deterministic conditions (Chandrasekaran, Sanghavi, Parrilo & Willsky 2011)
  - assumption on randomness (Candès, Li, Ma & Wright 2011)
- Both solve Principal Component Pursuit (PCP)

$$\text{minimize} \quad \|L\|_* + \gamma\|S\|_1 \tag{14}$$
$$\text{subject to} \quad \Sigma = L + S \tag{15}$$

- Ironically, γ is data-dependent in Chandrasekaran et al. (2011) and a universal constant ($1/\sqrt{N}$) in Candès et al. (2011).

Convex optimization

- Goes back to Shapiro (1982): *"how much can we reduce the rank of a symmetric matrix by changing only its diagonal entries"*.

$$\begin{array}{ll} \text{minimize} & \operatorname{rank}(L) \\ \text{subject to} & \Sigma = L + \Delta, \\ & L \succeq 0, \;\; \Delta \text{ diagonal.} \end{array} \tag{16}$$

- Problem (16) is not convex; Shapiro (1982) considered the relaxation:

$$\begin{array}{ll} \text{minimize} & \operatorname{tr}(L) \\ \text{subject to} & \Sigma = L + \Delta, \\ & L \succeq 0, \;\; \Delta \text{ diagonal.} \end{array} \tag{17}$$

- Problem (17), minimum trace factor analysis (MTFA), is convex.

MTFA

- If MTFA is feasible, (L, Δ) it has a unique optimal solution.

$$\begin{aligned} \text{minimize} \quad & \operatorname{tr}(L) \\ \text{subject to} \quad & \Sigma = L + \Delta, \\ & L \succeq 0, \;\; \Delta \text{ (block) diagonal.} \end{aligned} \tag{18}$$

- Saunderson et al. (2012) study conditions under which the solution (L, Δ) is "correct" (they extend Δ to a block diagonal structure).
- **Factor extraction:** L contains the broad factors; blocks of Δ constitute thin factors (industries, countries, etc)
- If $\hat{\Sigma}$ is the input, we may not wish to preserve the equality constraint.

Principal component pursuit (PCP)

- Solves the convex optimization problem

$$\begin{aligned} \text{minimize} \quad & \|L\|_* + \gamma \|S\|_1 \\ \text{subject to} \quad & \Sigma = L + S \end{aligned} \tag{19}$$

- $\|L\|_*$ is the nuclear norm, the sum of singular values (equals the trace whenever L is symmetric):
  - attempts to reduce rank of L.
- $\|S\|_1$ is the ℓ_1-vector norm with S viewed as a vector.
  - encourages sparsity of S.
- Its simple to preserve symmetry; less so for positive definiteness.

LRPS recovery

- Suppose $\Sigma = L + S$ is $N \times N$ for (L, S) such that
  - support of S is uniformly distributed over sets of cardinality m.
  - L obeys *incoherence conditions* with parameter μ (Candès & Recht 2009); i.e. singular vectors of L are reasonably spread out.
  - For some constants ρ_s and ρ_r we have

$$\text{rank}(L) \leq \frac{\rho_r N}{\mu (\log N)^2} \quad \text{and} \quad m \leq \rho_s N^2. \tag{20}$$

- Then[a] (Candès et al. 2011) for $\gamma = 1/\sqrt{N}$ there exists constant c such that solving PCP recovers (L, S) exactly with probability

$$1 - cN^{-10}. \tag{21}$$

[a] see (Chandrasekaran et al. 2011) for a related result.

Violating the equality constraint

- Suppose $\hat{\Sigma}$ is the sample covariance of the observed Gaussian returns.
- Since $\hat{\Sigma}$ may not be close to Σ we may wish to violate the constraint

$$\hat{\Sigma} = L + S \tag{22}$$

and instead maximize the Gaussian likelihood

$$G(L, S) = \log\det(L+S)^{-1} - \mathrm{tr}\Big((L+S)^{-1}\hat{\Sigma}\Big).$$

Lemma. *Given the decompostion $\Sigma = L + S$ we have a decomposition $\Sigma^{-1} = \mathcal{S} - \mathcal{L}$ such that rank(L) = rank$(\mathcal{L})$ with*

$$\mathcal{S} = S^{-1} \qquad \textit{and} \qquad \mathcal{L} = S^{-1}L\Sigma. \tag{23}$$

Latent variable convex optimization

- Let (R, ψ) be Gaussian with R the observed security returns and ψ the (unobserved) latent factors.
- If we believe Σ has a LRPS decompostion $L + S$ and $\mathcal{S} = S^{-1}$ is sparse we may solve the convex optimization problem

$$\begin{aligned} \text{minimize} \quad & -G(\mathcal{L}, \mathcal{S}) + \lambda \left(\|\mathcal{L}\|_* + \gamma \|\mathcal{S}\|_1 \right) \\ \text{subject to} \quad & \mathcal{S} - \mathcal{L} \succ 0, \mathcal{L} \succeq 0 \end{aligned} \tag{24}$$

- For example, if S is block diagonal (or any permutation) so it $\mathcal{S}$.
- Problem (24) was proposed in Chandrasekaran, Parrilo & Willsky (2010) and analyzed in Chandrasekaran, Parrilo & Willsky (2012).

Parameters

- The problem parameters (λ, γ) appearing in

$$\begin{array}{ll} \text{minimize} & -G(\mathcal{L}, \mathcal{S}) + \lambda \left(\|\mathcal{L}\|_* + \gamma \|\mathcal{S}\|_1 \right) \\ \text{subject to} & \mathcal{S} - \mathcal{L} \succ 0, \mathcal{L} \succeq 0 \end{array} \tag{25}$$

  are data dependent and harder to select that for PCP.

- λ may be set in proportion to $\sqrt{N/T}$ and γ must be chose by trial and error (Chandrasekaran et al. 2012).

- The solution $(\mathcal{L}, \mathcal{S})$ depends on λ to the degree that

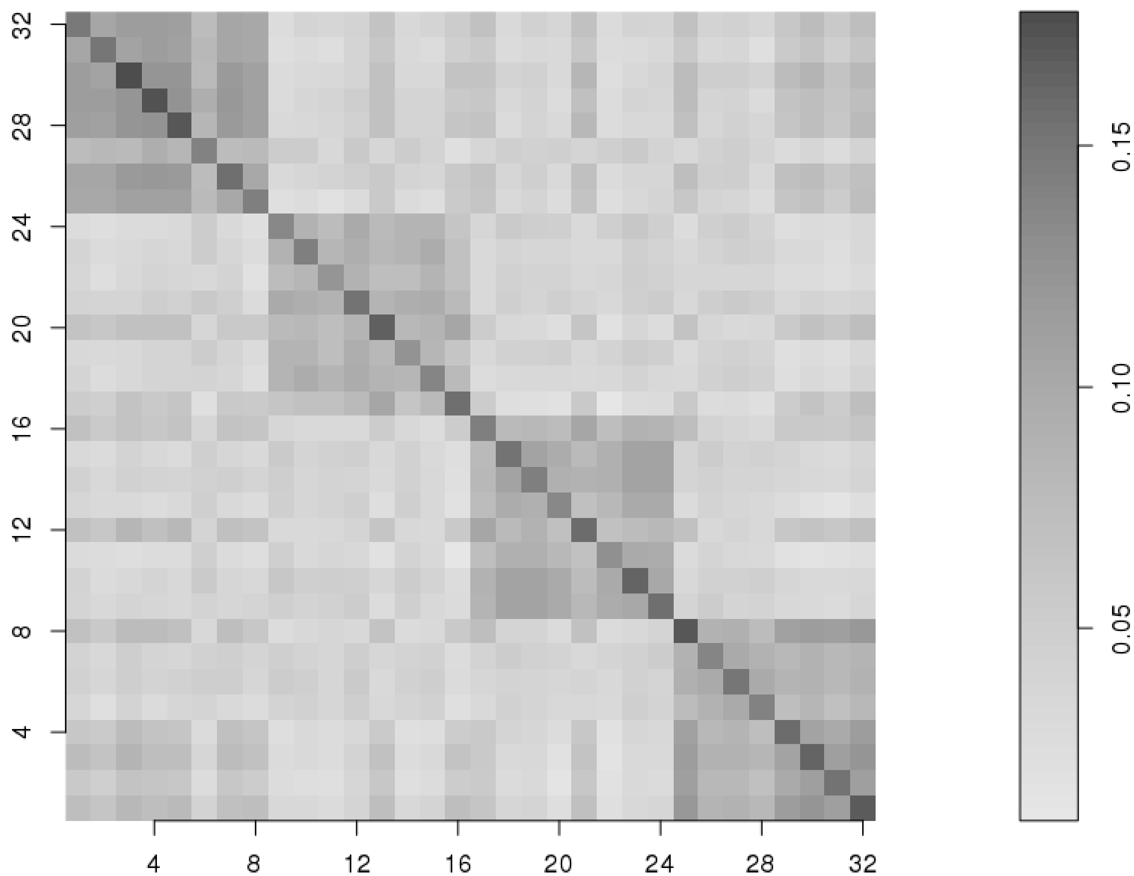
$$\hat{\Sigma}^{-1} \approx \mathcal{S} - \mathcal{L}. \tag{26}$$

- We solve (25) by using an algorithm of Ma, Xue & Zou (2013).

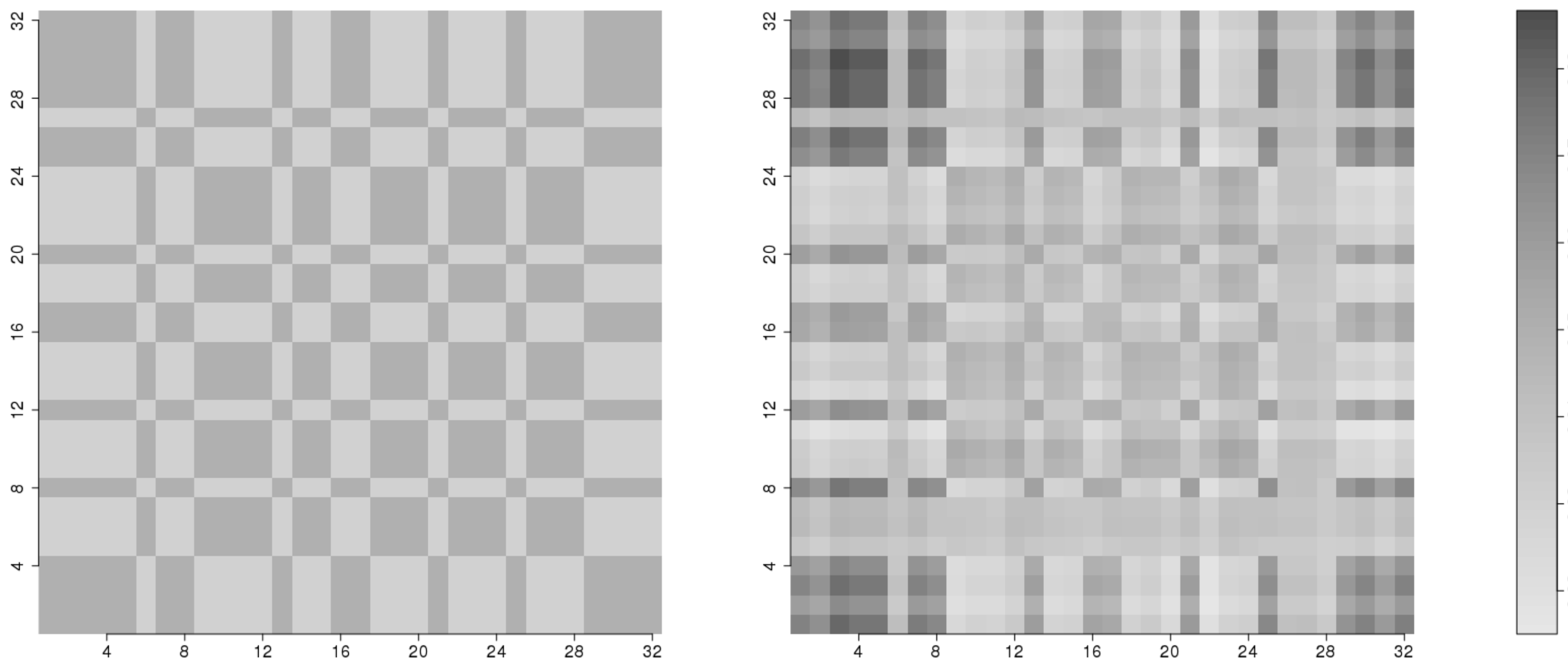
Numerical studies (synthetic data)

- $N = 32$ securities
- $T = 260$ observations (one year of daily data)
- $K = 2$ broad factors:
  - The market, with annualized volatility %20 (long only: all securities have positive exposure).
  - Creditworthiness, with annualized volatility of %10 (long/short factor: half creditworthy; half close to default)
- $\kappa = 4$ thin factors (e.g. countries)

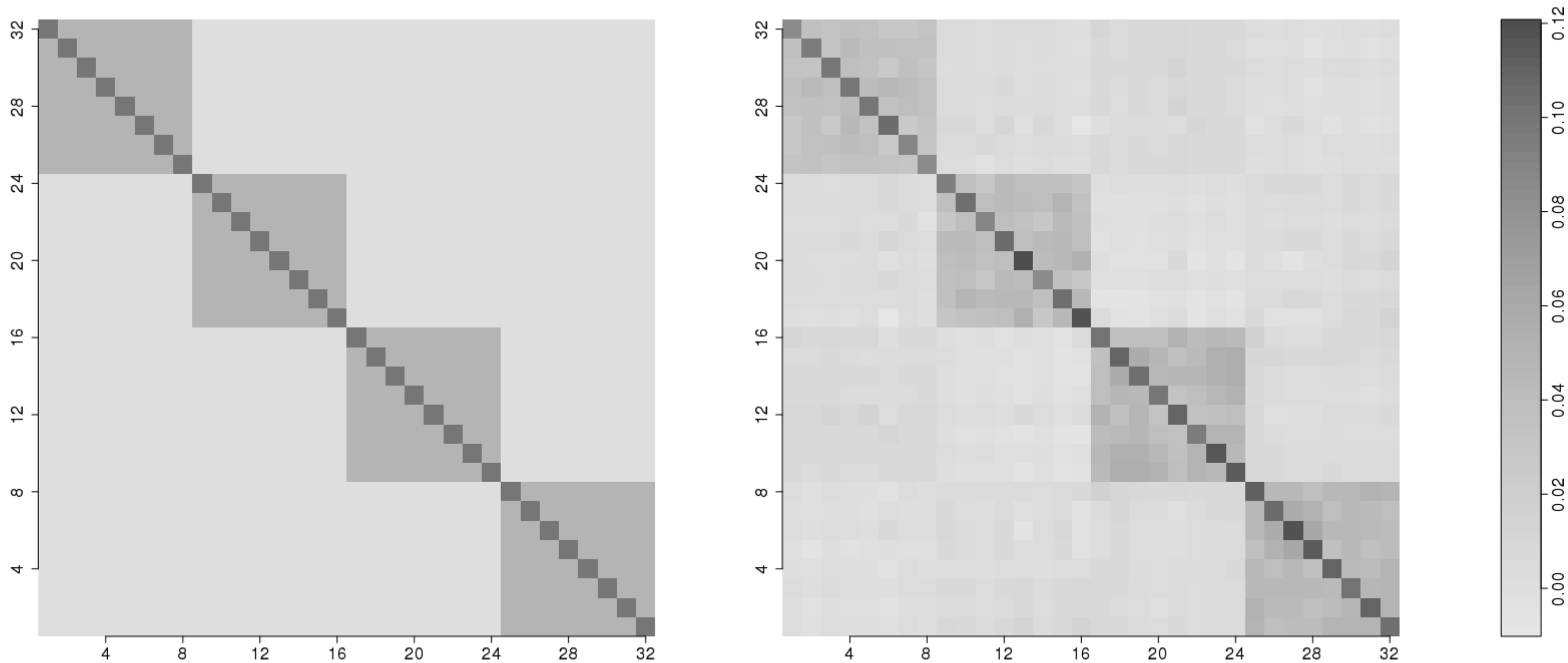
Algorithm input: sample covariance matrix



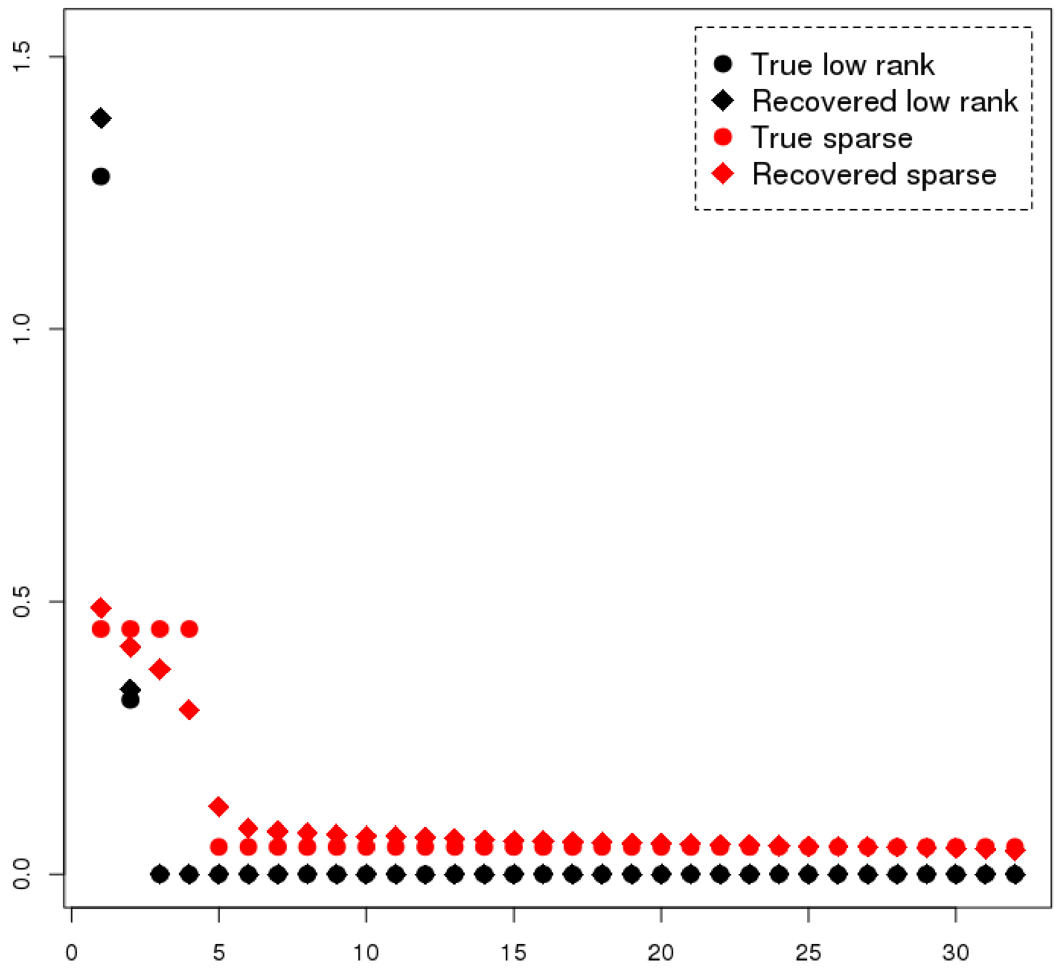
Low-rank part (true vs recovered)



Sparse part (true vs recovered)



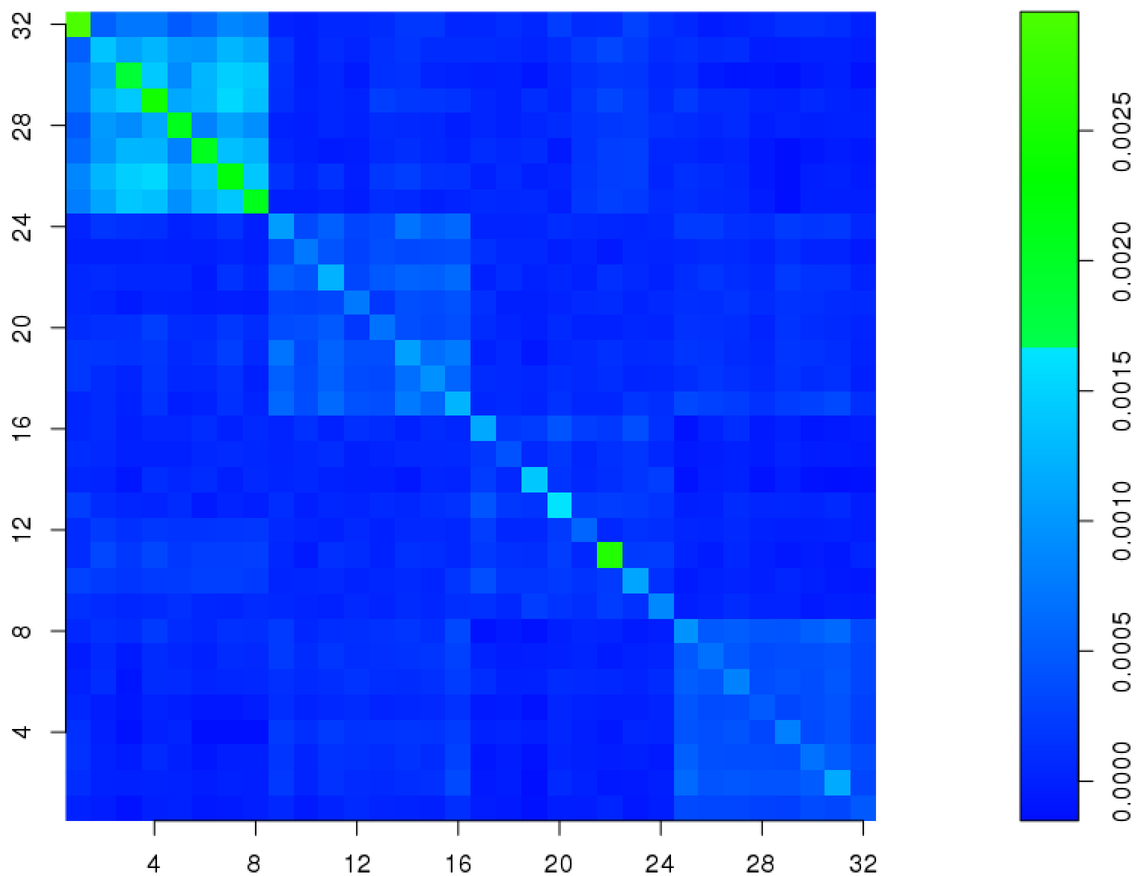
Eigenvalues (true vs recovered)



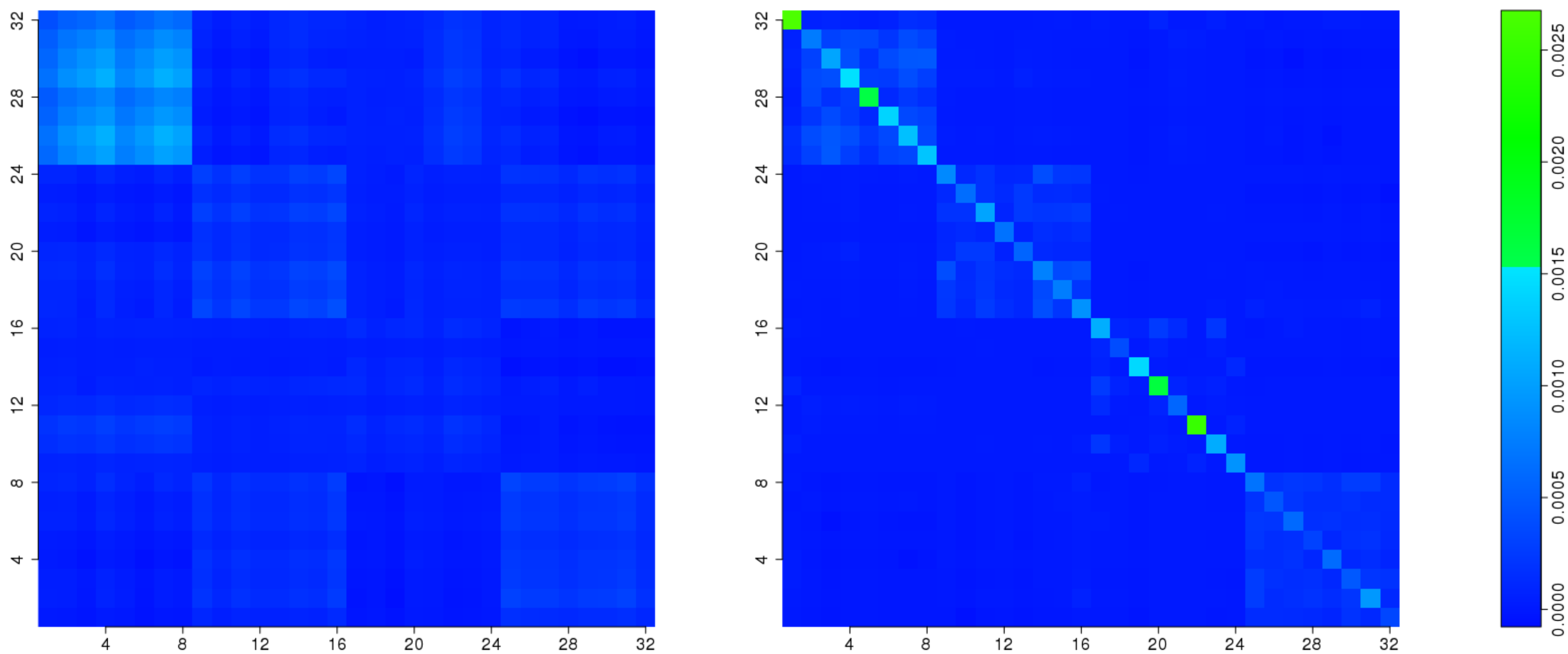
Numerical studies (real data)

- $N = 32$ securities
- $T = 260$ observations (one year of daily data)
- $K = ?$ broad factors
- Securities drawn from $\kappa = 4$ countries
  - China
  - Argentina
  - India
  - Saudi Arabia

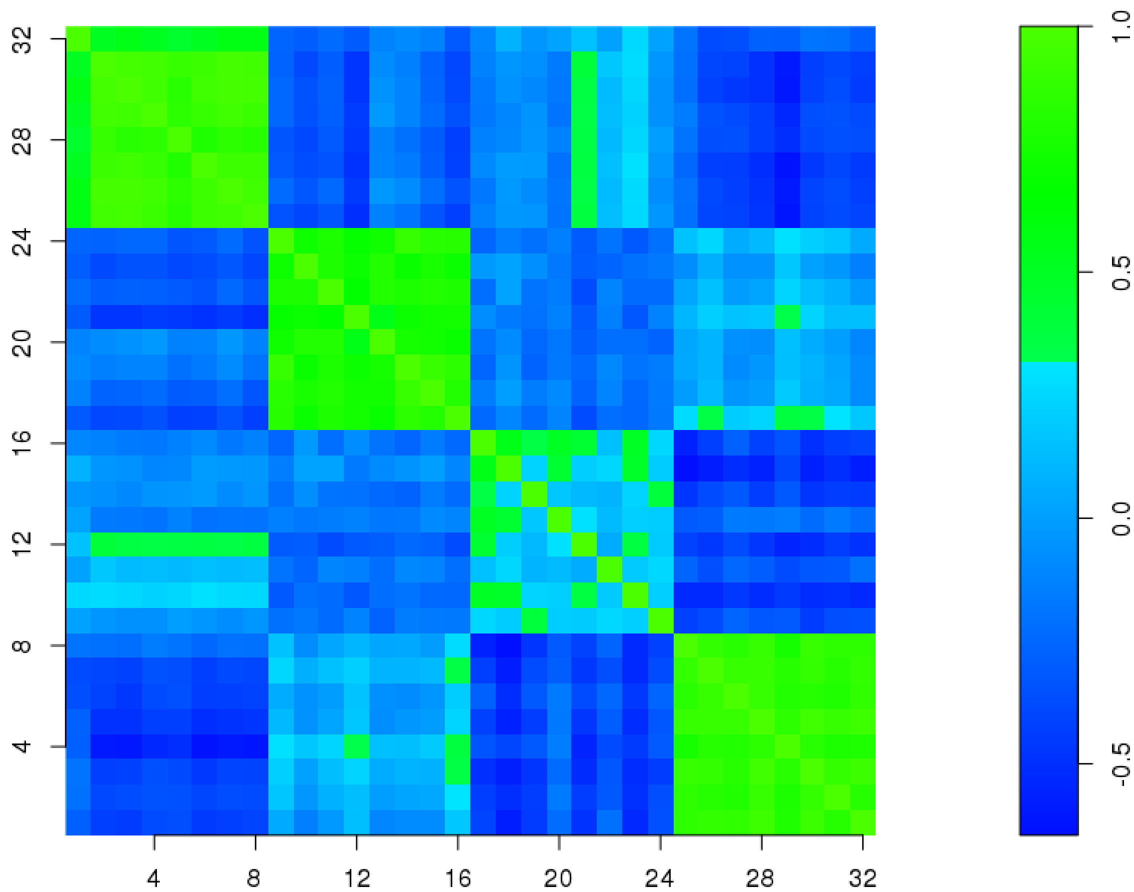
Algorithm input: sample covariance, Oct. 2015



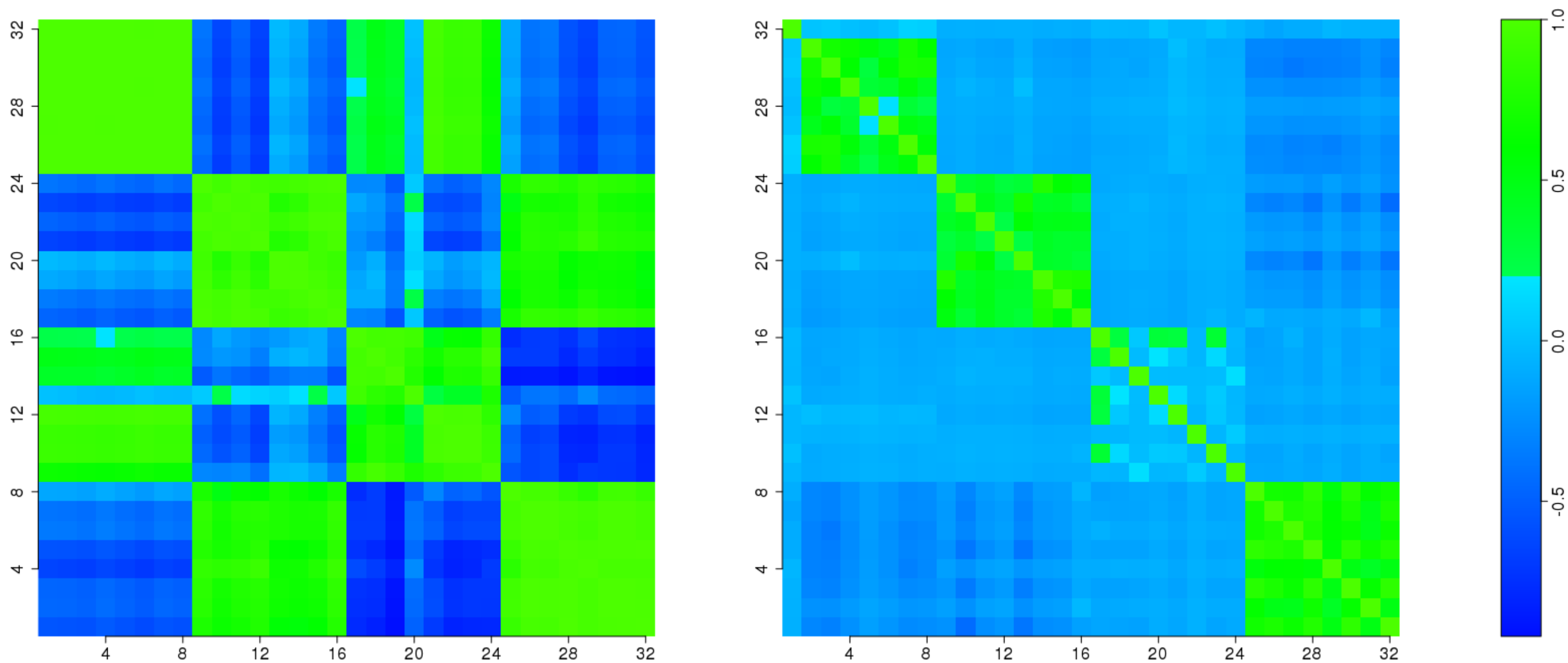
LRPS decomposition (covariance)



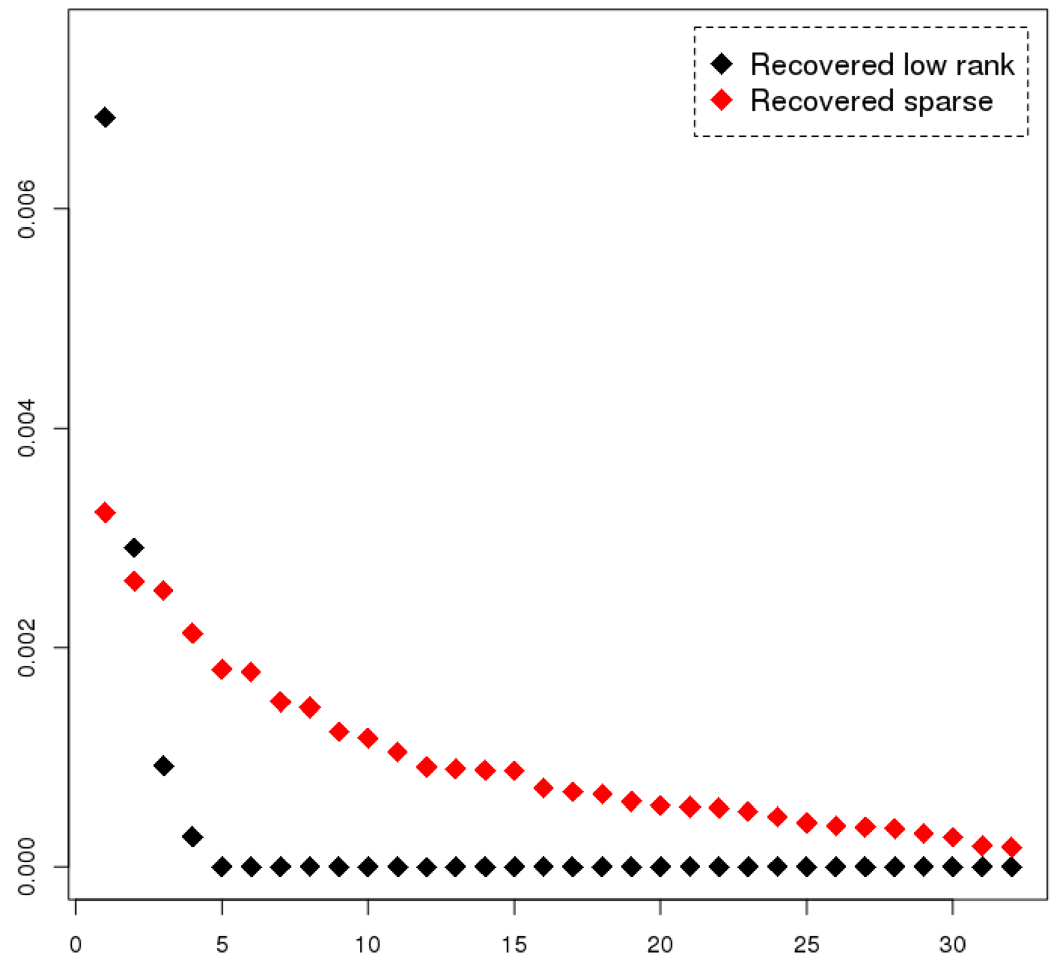
Sample correlation matrix, Oct. 2015



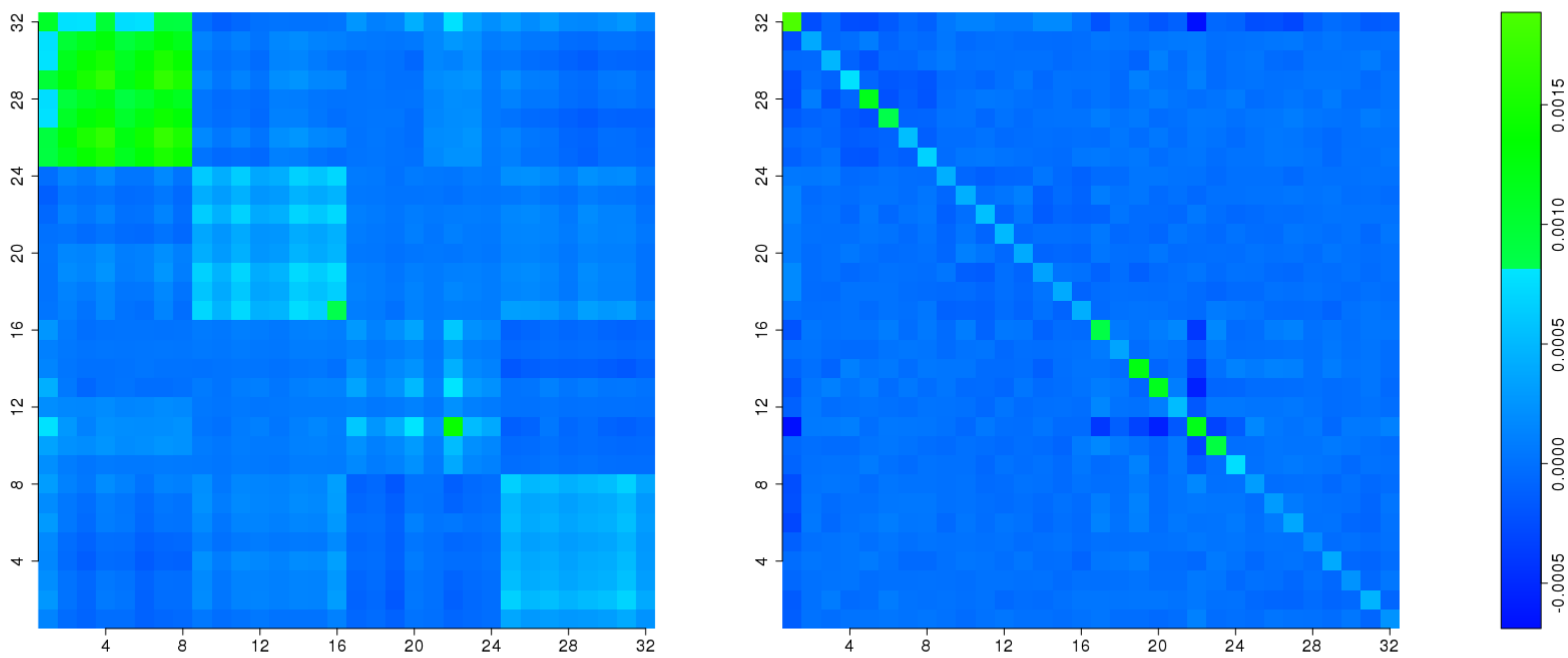
LRPS decomposition (correlations)



Recovered eigenvalues



PCA solution (K=4) and remainder



Questions.

(ads2@berkeley.edu)

References

Anderson, TW (Theodore Wilbur) (1968), *An introduction to multivariate statistical analysis*, John Wiley & Sons.

Bentler, Peter M & Jeffrey S Tanaka (1983), 'Problems with em algorithms for ml factor analysis', *Psychometrika* **48**(2), 247–251.

Bickel, Peter J & Elizaveta Levina (2008), 'Covariance regularization by thresholding', *The Annals of Statistics* pp. 2577–2604.

Candès, Emmanuel J & Benjamin Recht (2009), 'Exact matrix completion via convex optimization', *Foundations of Computational mathematics* **9**(6), 717–772.

Candès, Emmanuel J, Xiaodong Li, Yi Ma & John Wright (2011), 'Robust principal component analysis?', *Journal of the ACM (JACM)* **58**(3), 11.

Chamberlain, Gary & Michael Rothschild (1982), 'Arbitrage, factor structure, and mean-variance analysis on large asset markets'.

Chandrasekaran, Venkat, Pablo A Parrilo & Alan S Willsky (2010), Latent variable graphical model selection via convex optimization, *in* 'Communication, Control, and Computing (Allerton), 2010 48th Annual Allerton Conference on', IEEE, pp. 1610–1613.

Chandrasekaran, Venkat, Pablo A Parrilo & Alan S Willsky (2012), 'Latent variable graphical model selection via convex optimization', *The Annals of Statistics* **40**(4), 1935–1967.

Chandrasekaran, Venkat, Sujay Sanghavi, Pablo A Parrilo & Alan S Willsky (2011), 'Rank-sparsity incoherence for matrix decomposition', *SIAM Journal on Optimization* **21**(2), 572–596.

De La Torre, Fernando & Michael J Black (2003), 'A framework for robust subspace learning', *International Journal of Computer Vision* **54**(1-3), 117–142.

Drton, Mathias, Bernd Sturmfels & Seth Sullivant (2007), 'Algebraic factor analysis: tetrads, pentads and beyond', *Probability Theory and Related Fields* **138**(3-4), 463–493.

Fischler, Martin A & Robert C Bolles (1981), 'Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography', *Communications of the ACM* **24**(6), 381–395.

Gnanadesikan, Ramanathan & John R Kettenring (1972), 'Robust estimates, residuals, and outlier detection with multiresponse data', *Biometrics* pp. 81–124.

Jolliffe, Ian (2002), *Principal component analysis*, Wiley Online Library.

Jöreskog, Karl G (1967), 'A general approach to confirmatory maximum likelihood factor analysis', *ETS Research Bulletin Series* **1967**(2), 183–202.

Karoui, Noureddine El (2008), 'Spectrum estimation for large dimensional covariance matrices using random matrix theory', *The Annals of Statistics* pp. 2757–2790.

Ke, Qifa & Takeo Kanade (2005), Robust l 1 norm factorization in the presence of outliers and missing data by alternative convex programming, *in* 'Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on', Vol. 1, IEEE, pp. 739–746.

Ma, Shiqian, Lingzhou Xue & Hui Zou (2013), 'Alternating direction methods for latent variable gaussian graphical model selection', *Neural computation* **25**(8), 2172–2198.

Petersen, Kaare Brandt & Michael Syskind Pedersen (2008), 'The matrix cookbook', *Technical University of Denmark* **7**, 15.

Rosenberg, Barr (1984), 'Prediction of common stock investment risk', *The Journal of Portfolio Management* **11**(1), 44–53.

Ross, Stephen A (1976), 'The arbitrage theory of capital asset pricing', *Journal of economic theory* **13**(3), 341–360.

Rubin, Donald B & Dorothy T Thayer (1982), 'Em algorithms for ml factor analysis', *Psychometrika* **47**(1), 69–76.

Rubin, Donald B & Dorothy T Thayer (1983), 'More on em for ml factor analysis', *Psychometrika* **48**(2), 253–257.

Saunderson, James, Venkat Chandrasekaran, Pablo A Parrilo & Alan S Willsky (2012), 'Diagonal and low-rank matrix decompositions, correlation matrices, and ellipsoid fitting', *SIAM Journal on Matrix Analysis and Applications* **33**(4), 1395–1416.

Shapiro, Alexander (1982), 'Rank-reducibility of a symmetric matrix and sampling theory of minimum trace factor analysis', *Psychometrika* **47**(2), 187–199.

Shapiro, Alexander (1985), 'Identifiability of factor analysis: Some results and open problems', *Linear Algebra and its Applications* **70**, 1–7.

Shapiro, Alexander & Jos MF Ten Berge (2002), 'Statistical inference of minimum rank factor analysis', *Psychometrika* **67**(1), 79–94.

Sharpe, William F (1964), 'Capital asset prices: A theory of market equilibrium under conditions of risk', *The journal of finance* **19**(3), 425–442.

Treynor, Jack L (1962), Toward a theory of market value of risky assets. Presented to the MIT Finance Faculty Seminar.

Wilson, Edwin B & Jane Worcester (1939), 'The resolution of six tests into three general factors', *Proceedings of the National Academy of Sciences of the United States of America* **25**(2), 73.