

Identifying Small Mean Reverting Portfolios

Alexandre d'Aspremont
ORFE, Princeton University

Support from NSF, DHS and Google.

Introduction

Mean reversion:

- Classic case of statistical arbitrage.
- Highlights long-term structural relationships in the data.
- We could replace mean-reversion by momentum throughout the talk.

Sparse portfolios:

- Better interpretability.
- Less transaction costs.

Mean reversion

- Let S_{ti} be the value at time t of an asset S_i for $i = 1, \dots, n$ and $t = 1, \dots, m$.
- We form portfolios P_t of these assets with coefficients x_i , modeled by an Ornstein-Uhlenbeck process:

$$dP_t = \lambda(\bar{P} - P_t)dt + \sigma dZ_t \quad \text{with } P_t = \sum_{i=1}^n x_i S_{ti}$$

where Z_t is a standard Brownian motion.

- **Objective:** maximize the mean reversion coefficient λ of P_t by adjusting the coefficients x , while imposing $\|x\| = 1$ and $\mathbf{Card}(x) \leq k$.

Outline

- Small Mean Reverting Portfolios
 - **Canonical decomposition**
 - Sparse generalized eigenvalue problems
 - Estimation and trading
 - Numerical results
- Covariance Selection
 - Introduction
 - Robust Maximum Likelihood Estimation
 - Algorithms
 - Numerical Results

Canonical decomposition

- In a discrete setting, we assume that the asset prices follow a (stationary) autoregressive process with:

$$S_t = AS_{t-1} + Z_t \quad (1)$$

where S_{t-1} is the lagged portfolio process, $A \in \mathbf{R}^{n \times n}$ and Z_t is a vector of i.i.d. Gaussian noise with zero mean and covariance $\Sigma \in \mathbf{S}^n$, independent of S_{t-1} .

- Take $n = 1$ in equation (1):

$$\mathbf{E}[S_t^2] = \mathbf{E}[(AS_{t-1})^2] + \mathbf{E}[Z_t^2]$$

which can be rewritten as $\sigma_t^2 = \sigma_{t-1}^2 + \Sigma$.

- Box & Tiao (1977) then measure the **predictability** of stationary series by:

$$\lambda = \frac{\sigma_{t-1}^2}{\sigma_t^2}. \quad (2)$$

Canonical decomposition

- Consider a portfolio $P_t = x^T S_t$ with $x \in \mathbf{R}^n$, using (1) we know that

$$x^T S_t = x^T A S_{t-1} + x^T Z_t,$$

so its predicability can be measured as:

$$\lambda_x = \frac{x^T A \Gamma A^T x}{x^T \Gamma x}$$

where $\Gamma = \mathbf{E}[S S^T]$.

- The portfolio with maximum (respectively minimum) predictability will be the eigenvector corresponding to the largest (respectively smallest) eigenvalue of the matrix:

$$\Gamma^{-1} A \Gamma A^T. \tag{3}$$

- We then only need to estimate A . . .

Canonical decompositions

- The **Box-Tiao procedure** finds linear combinations of the assets ranked in order of predictability by computing the eigenvectors of the matrix:

$$(S^T S)^{-1} \left(\hat{S}_t^T \hat{S}_t \right) \quad (4)$$

where \hat{S}_t is the least squares estimate computed above.

- The **Johansen procedure**: following Bewley, Orden, Yang & Fisher (1994), we rewrite equation (1) as:

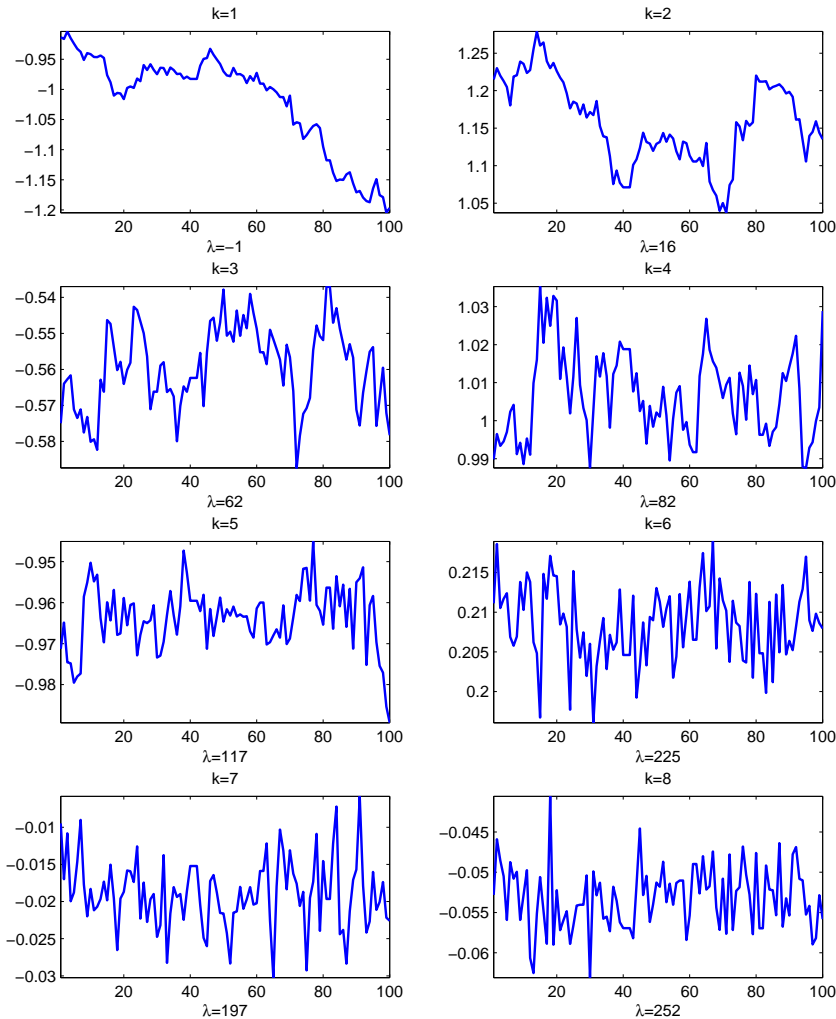
$$\Delta S_t = Q S_{t-1} + Z_t$$

where $Q = A - \mathbf{I}$. The basis of cointegrating portfolios is then found by solving the following generalized eigenvalue problem:

$$\lambda S_{t-1}^T S_{t-1} - S_{t-1}^T \Delta S_t (\Delta S_t^T \Delta S_t)^{-1} \Delta S_t^T S_{t-1} \quad (5)$$

in the variable $\lambda \in \mathbf{R}$.

Mean-reversion: canonical decompositions



Box & Tiao (1977) canonical decomposition on 100 days of U.S. swap rate data (in percent), ranked in decreasing order of predictability. The mean reversion coefficient λ is listed below each plot.

Mean-reversion: related works

- Fama & French (1988), Poterba & Summers (1988) model and test for market predictability in excess returns.
- Cointegration techniques, (see Engle & Granger (1987), and Alexander (1999) for a survey of applications in finance) are usually used to extract mean reverting portfolios.
- Several authors focused on the optimal investment problem when expected returns are mean reverting, with Kim & Omberg (1996) and Campbell & Viceira (1999) or Wachter (2002) among others, obtaining closed-form solutions in some particular cases.
- Liu & Longstaff (2004) study the optimal investment problem in the presence of a “textbook” finite horizon arbitrage opportunity, modeled as a Brownian bridge. Jurek & Yang (2006) study this same problem when the arbitrage horizon is indeterminate. Gatev, Goetzmann & Rouwenhorst (2006) also studied the performance of pairs trading, which are classic examples of structurally mean-reverting portfolios.
- The LTCM meltdown in 1998 focused a lot of attention on the impact of leverage limits and liquidity, see Grossman & Vila (1992) or Xiong (2001) for a discussion.

Sparse methods

- ℓ_1 regularized regression (LASSO): Tibshirani (1996).
- Feature selection: ℓ_1 penalized support vector machines.
- Compressed sensing: Candès & Tao (2005), Donoho & Tanner (2005).
- Basis pursuit: Chen, Donoho & Saunders (2001), . . .
- Sparse PCA and covariance selection: d'Aspremont, El Ghaoui, Jordan & Lanckriet (2007) and d'Aspremont, Banerjee & El Ghaoui (2006).

Outline

- Small Mean Reverting Portfolios
 - Canonical decomposition
 - **Sparse generalized eigenvalue problems**
 - Estimation and trading
 - Numerical results
- Covariance Selection
 - Introduction
 - Robust Maximum Likelihood Estimation
 - Algorithms
 - Numerical Results

Sparse generalized eigenvalue problems

Both canonical decompositions involve solving a **generalized eigenvalue problem** of the form:

$$\det(\lambda A - B) = 0 \quad (6)$$

in the variable $\lambda \in \mathbf{R}$, where $A, B \in \mathbf{S}^n$. This is usually solved using a QZ decomposition. The largest solution of this problem can be written in variational form as:

$$\lambda^{\max} = \max_{x \in \mathbf{R}^n} \frac{x^T A x}{x^T B x}.$$

Here however, we seek to maximize (or minimize) that ratio while constraining the cardinality of the (portfolio) coefficient vector x and solve instead:

$$\begin{aligned} & \text{maximize} && x^T A x / x^T B x \\ & \text{subject to} && \mathbf{Card}(x) \leq k \\ & && \|x\| = 1, \end{aligned} \quad (7)$$

where $k > 0$ is a given constant and $\mathbf{Card}(x)$ is the number of nonzero coefficients in x .

Sparse generalized eigenvalue problems

- Solving generalized eigenvalue problems is **easy**: takes $O(n^3)$ operations.
- Solving sparse generalized eigenvalue problems is **hard**: equivalent to subset selection which is NP-Hard.

Here, we seek good approximate solutions to:

$$\begin{array}{ll} \text{maximize} & x^T A x / x^T B x \\ \text{subject to} & \mathbf{Card}(x) \leq k \\ & \|x\| = 1, \end{array}$$

using two algorithms:

- **Greedy search**: Incrementally scan all variables.
- **Semidefinite relaxation**: form a tractable convex relaxation.

Greedy Search

- Define:

$$I_k = \{i \in [1, n] : x_i \neq 0\},$$

- We build approximate solutions **recursively** in k . When $k = 1$, we can simply find I_1 as:

$$I_1 = \operatorname{argmax}_{i \in [1, n]} A_{ii}/B_{ii}.$$

- Given I_k , we add one variable with index i_{k+1} to produce the largest increase in predictability:

$$\max_{\{x \in \mathbf{R}^n : \operatorname{supp}(x) = I_k \cup \{i\}\}} \frac{x^T A x}{x^T B x}.$$

- The complexity of computing solutions for all k is in $O(n^4)$.

Semidefinite relaxation

Start from our original problem:

$$\begin{aligned} & \text{maximize} && x^T A x / x^T B x \\ & \text{subject to} && \mathbf{Card}(x) \leq k \\ & && \|x\| = 1, \end{aligned}$$

with variable $x \in \mathbf{R}^n$, and rewrite it in terms of $X = x x^T \in \mathbf{S}_n$:

$$\begin{aligned} & \text{maximize} && \mathbf{Tr}(AX) / \mathbf{Tr}(BX) \\ & \text{subject to} && \mathbf{Card}(X) \leq k^2 \\ & && \mathbf{Tr}(X) = 1 \\ & && X \succeq 0, \mathbf{Rank}(X) = 1, \end{aligned}$$

in the variable $X \in \mathbf{S}_n$. This program is **equivalent** to the first one.

Semidefinite relaxation

$$\begin{aligned} & \text{maximize} && \mathbf{Tr}(AX)/\mathbf{Tr}(BX) \\ & \text{subject to} && \mathbf{Card}(X) \leq k^2 \\ & && \mathbf{Tr}(X) = 1 \\ & && X \succeq 0, \mathbf{Rank}(X) = 1, \end{aligned}$$

- Since $\mathbf{Card}(u) = q$ implies $\|u\|_1 \leq \sqrt{q}\|u\|_2$, we can replace the nonconvex constraint $\mathbf{Card}(X) \leq k^2$, by a weaker but **convex** constraint: $\mathbf{1}^T |X| \mathbf{1} \leq k$.
- We drop the rank constraint to get the following quasi-convex program:

$$\begin{aligned} & \text{maximize} && \mathbf{Tr}(AX)/\mathbf{Tr}(BX) \\ & \text{subject to} && \mathbf{1}^T |X| \mathbf{1} \leq k \\ & && \mathbf{Tr}(X) = 1 \\ & && X \succeq 0, \end{aligned}$$

in the variable $X \in \mathbf{S}_n$.

Semidefinite relaxation

Starting from the quasi-convex program:

$$\begin{aligned} & \text{maximize} && \mathbf{Tr}(AX)/\mathbf{Tr}(BX) \\ & \text{subject to} && \mathbf{1}^T |X| \mathbf{1} \leq k \\ & && \mathbf{Tr}(X) = 1 \\ & && X \succeq 0, \end{aligned}$$

we change variables:

$$Y = \frac{X}{\mathbf{Tr}(BX)}, \quad z = \frac{1}{\mathbf{Tr}(BX)}$$

and solve:

$$\begin{aligned} & \text{maximize} && \mathbf{Tr}(AY) \\ & \text{subject to} && \mathbf{1}^T |Y| \mathbf{1} - kz \leq 0 \\ & && \mathbf{Tr}(Y) - z = 0 \\ & && \mathbf{Tr}(BY) = 1 \\ & && Y \succeq 0, \end{aligned} \tag{8}$$

which is a semidefinite program in the variables $Y \in \mathbf{S}_n$ and $z \in \mathbf{R}_+$ and can be solved using standard SDP solvers such as SDPT3 by Toh, Todd & Tutuncu (1999).

Performance

Greedy algorithm:

- The optimal solutions of problem (7) might not have an increasing support set sequence $I_k \subset I_{k+1}$.
- However, the cost of this method is relatively low: with each iteration costing $O(k^2(n - k))$, the complexity of computing solutions for all k is in $O(n^4)$.
- This recursive procedure can also be repeated both forward and backward to improve the quality of the solution.
- Stability issues.

Semidefinite relaxation:

- Higher complexity.
- ℓ_1 penalization makes it potentially more stable.

Outline

- Small Mean Reverting Portfolios
 - Canonical decomposition
 - Sparse generalized eigenvalue problems
 - **Estimation and trading**
 - Numerical results
- Covariance Selection
 - Introduction
 - Robust Maximum Likelihood Estimation
 - Algorithms
 - Numerical Results

Estimation and trading

By integrating P_t over a time increment Δt we get:

$$P_t = \bar{P} + e^{-\lambda\Delta t}(P_{t-\Delta t} - \bar{P}) + \sigma \int_{t-\Delta t}^t e^{\lambda(s-t)} dZ_s,$$

so we can **estimate** λ and σ by simply regressing P_t on $P_{t-\Delta t}$ and a constant. We have the following estimators for the parameters of P_t :

$$\hat{\mu} = \frac{1}{N} \sum_{i=0}^N P_{t_i}$$

$$\hat{\lambda} = -\frac{1}{\Delta t} \log \left(\frac{\sum_{i=1}^N (P_{t_i} - \hat{\mu})(P_{t_{i-1}} - \hat{\mu})}{\sum_{i=1}^N (P_{t_i} - \hat{\mu})(P_{t_i} - \hat{\mu})} \right)$$

$$\hat{\sigma}^2 = \frac{2\lambda}{(1 - e^{-2\lambda\Delta t})(N - 2)} \sum_{i=1}^N \left((P_{t_i} - \hat{\mu}) - e^{-\lambda\Delta t}(P_{t_{i-1}} - \hat{\mu}) \right)^2$$

Estimation and trading

Trading O.U. processes: two classic strategies.

- **Threshold:** Invest when the spread $|\bar{P} - P_t|$ crosses a certain threshold, cf. Gatev et al. (2006).
- **Linear:** Under log-utility, the optimum strategy is **linear**:

$$N = \frac{\lambda(\bar{P} - P_t) - rP_t}{\sigma^2} W_t$$

where N is the number of units of portfolio the agent holds and W_t the investor's wealth at time t . See Jurek & Yang (2006).

A few remarks:

- None of these results account for transaction costs.
- Jurek & Yang (2006) also find the optimal strategy for CRRA utility defined over wealth at a finite horizon and Epstein-Zin utility defined over intermediate cash flows.
- Similar results hold with proportional fund flows, cf. Jurek & Yang (2006).

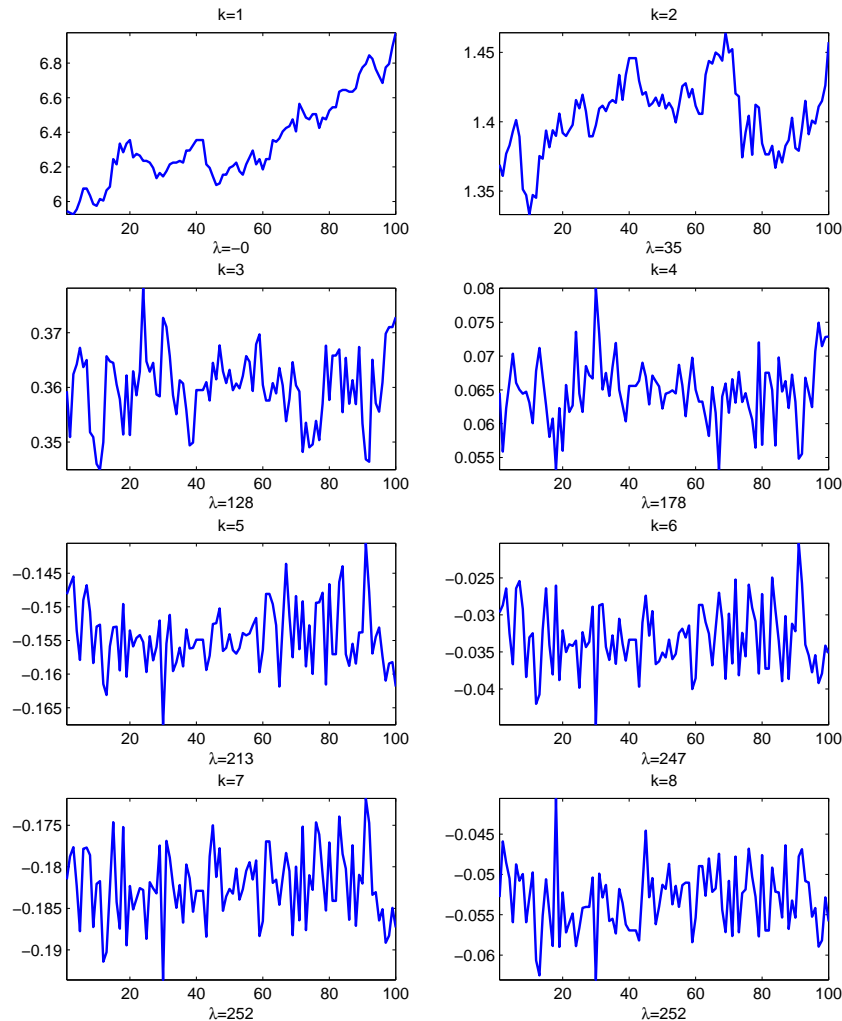
Outline

- Small Mean Reverting Portfolios
 - Canonical decomposition
 - Sparse generalized eigenvalue problems
 - Estimation and trading
 - **Numerical results**
- Covariance Selection
 - Introduction
 - Robust Maximum Likelihood Estimation
 - Algorithms
 - Numerical Results

Numerical Results

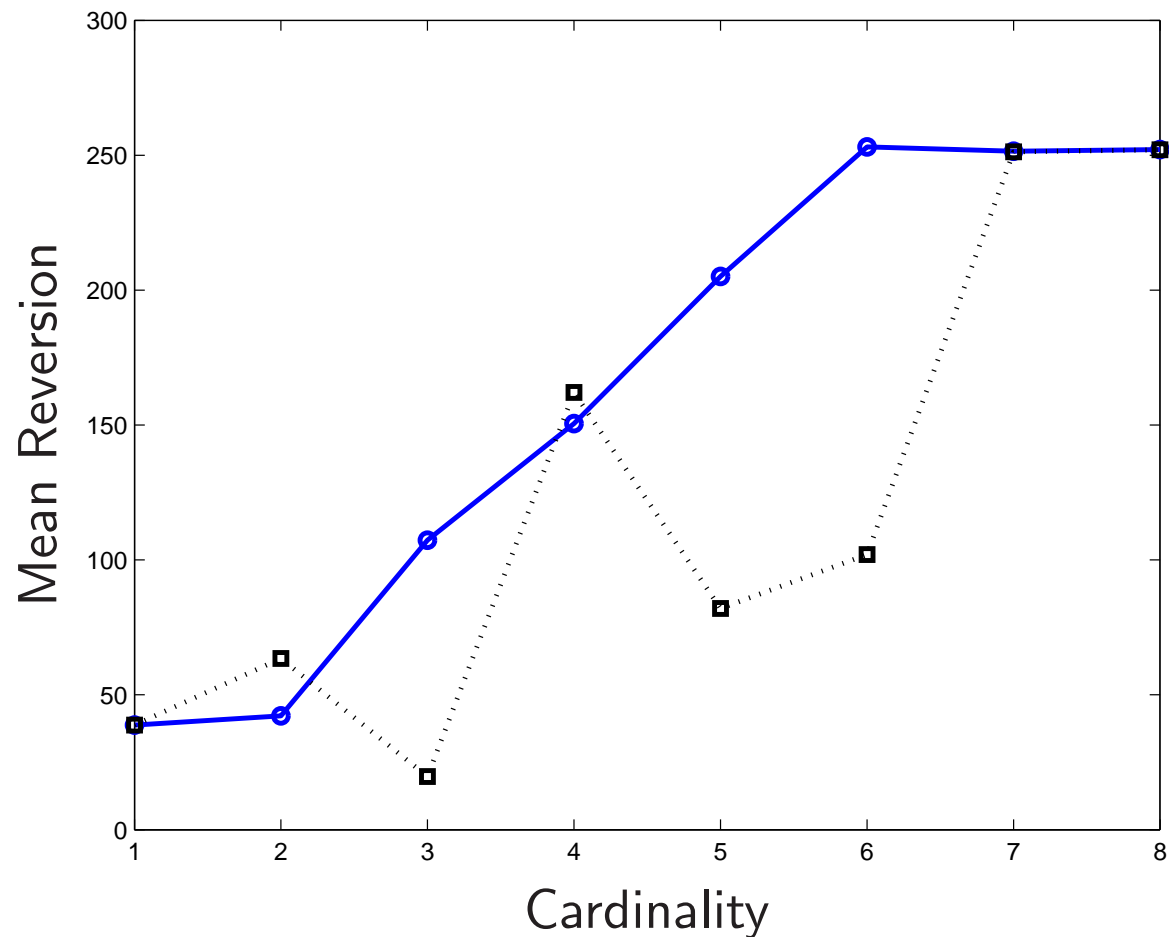
- U.S. swap rate data for maturities 1Y, 2Y, 3Y, 4Y, 5Y, 7Y, 10Y and 30Y from 1998 until 2005.
- Use greedy algorithm to compute optimally mean reverting portfolios of increasing cardinality for time windows of 200 days and repeat the procedure every 50 days.
- Update portfolios daily using linear rule.

Numerical Results



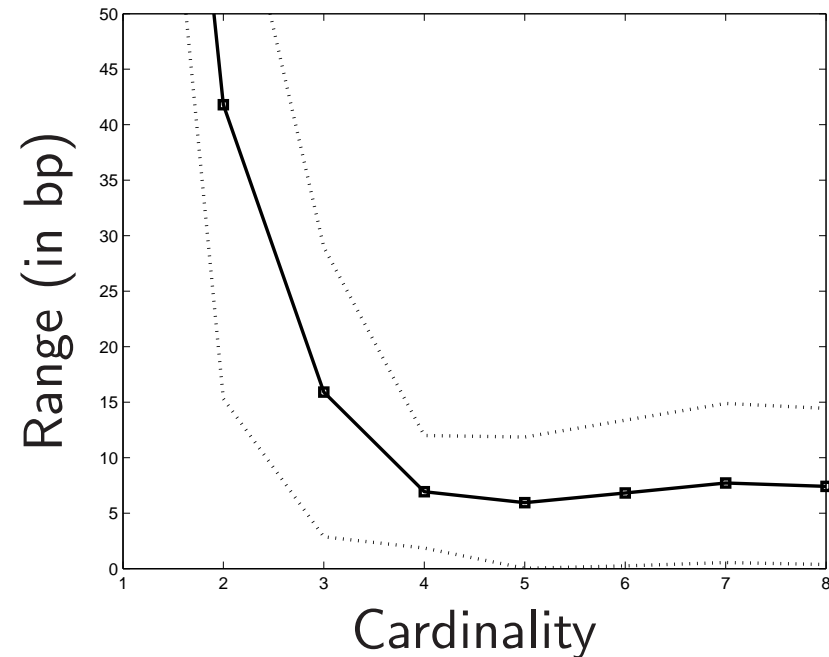
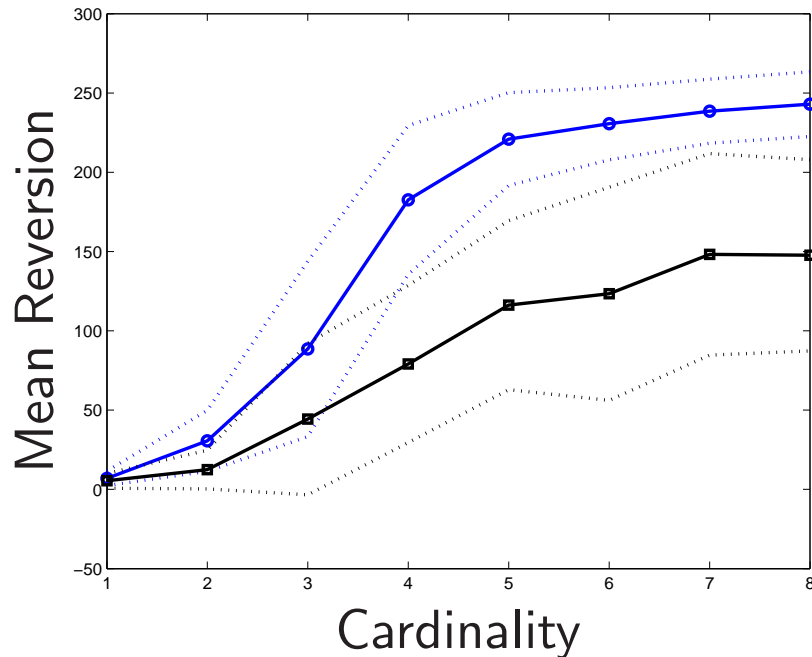
Sparse canonical decomposition on 100 days of U.S. swap rate data (in percent). The number of nonzero coefficients in each portfolio vector is listed as k on top of each subplot, the mean reversion coefficient λ is listed below each one.

Numerical Results



Mean reversion coefficient λ versus portfolio cardinality (number of nonzero coefficients) using the greedy search (solid line) and the semidefinite relaxation (dashed line) on U.S. swap rate data.

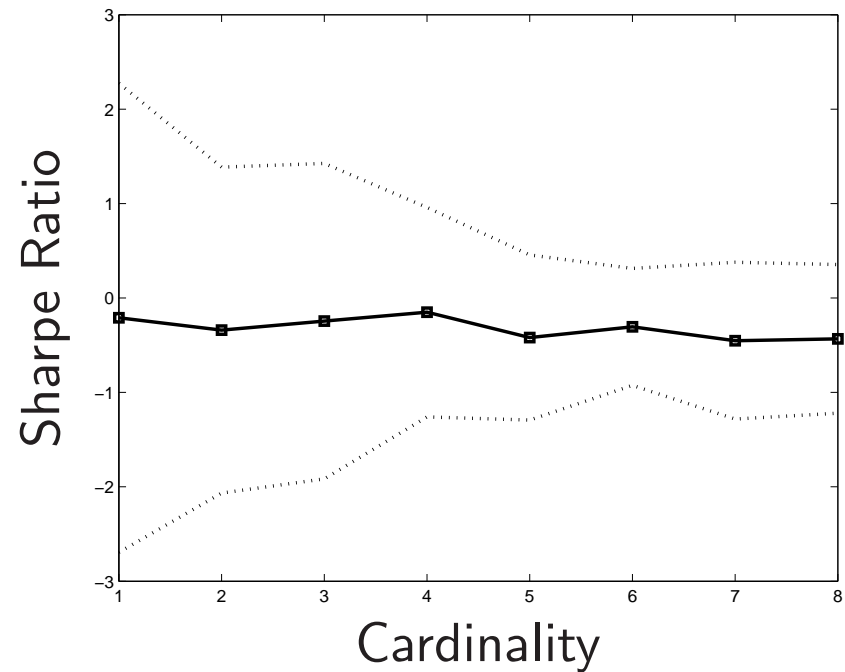
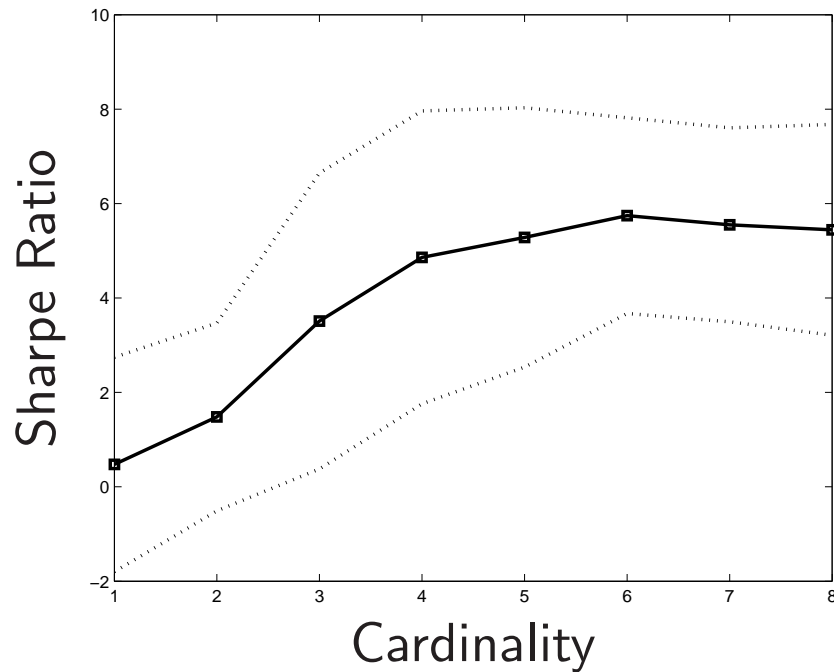
Numerical Results



Left: **mean reversion** coefficient λ versus portfolio cardinality (number of nonzero coefficients), in sample (blue circles) and out of sample (black squares) on U.S. swaps.

Right: out of sample portfolio **price range** (in basis points) versus cardinality (number of nonzero coefficients) on U.S. swap rate data. Dashed lines at plus and minus one standard deviation.

Numerical Results



Left: average out of sample **sharpe ratio** versus portfolio cardinality on U.S. swaps.

Right: idem, with **transaction costs** modeled as a Bid-Ask spread of 1bp. The dashed lines are at plus and minus one standard deviation.

Outline

- Small Mean Reverting Portfolios
 - Canonical decomposition
 - Sparse generalized eigenvalue problems
 - Estimation and trading
 - Numerical results
- **Covariance Selection**
 - Introduction
 - Robust Maximum Likelihood Estimation
 - Algorithms
 - Numerical Results

Covariance Selection

We estimate a **sample covariance matrix** Σ from empirical data. . .

- Objective: infer **dependence** relationships between variables.
- We want this information to be as **sparse** as possible.
- Basic solution: look at the magnitude of the covariance coefficients:

$$|\Sigma_{ij}| > \beta \quad \Leftrightarrow \quad \text{variables } i \text{ and } j \text{ are related,}$$

and simply threshold smaller coefficients to zero. (not always psd.)

- Use these dependence relationships to identify small groups of assets which might be cointegrated.

We can do better. . . Following Dempster (1972), look for zeros in the **inverse** covariance matrix.

Covariance Selection

Conditional independence:

- Suppose X, Y, Z have are jointly normal with covariance matrix Σ , with

$$\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$$

where $\Sigma_{11} \in \mathbf{R}^{2 \times 2}$ and $\Sigma_{22} \in \mathbf{R}$.

- Conditioned on Z , X, Y are still normally distributed with covariance matrix C satisfying:

$$C = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} = (\Sigma^{-1})_{11}^{-1}$$

- So X and Y are **conditionally independent** iff $(\Sigma^{-1})_{11}$ is diagonal, which is also:

$$\Sigma_{xy}^{-1} = 0$$

Covariance Selection

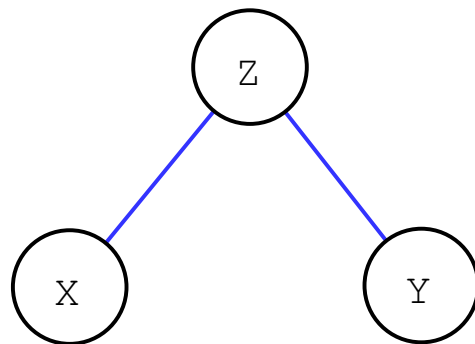
- Suppose we have iid noise $\epsilon_i \sim \mathcal{N}(0, 1)$ and the following linear model:

$$x = z + \epsilon_1$$

$$y = z + \epsilon_2$$

$$z = \epsilon_3$$

- Graphically, this is:

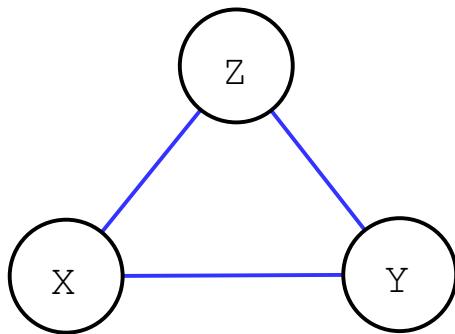


Covariance Selection

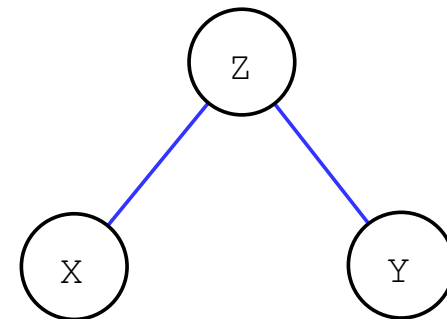
- The covariance matrix and inverse covariance are given by:

$$\Sigma = \begin{pmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 1 \end{pmatrix} \quad \Sigma^{-1} = \begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & -1 \\ -1 & -1 & 3 \end{pmatrix}$$

- The inverse covariance matrix has Σ_{12}^{-1} clearly showing that the variables x and y are independent conditioned on z .
- Graphically, this is again:

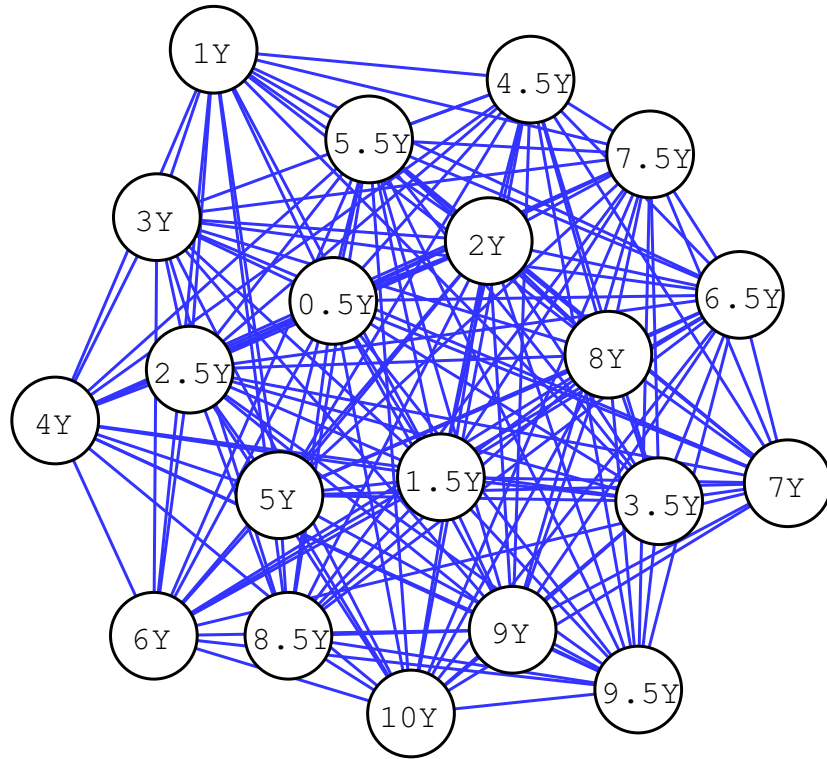


versus

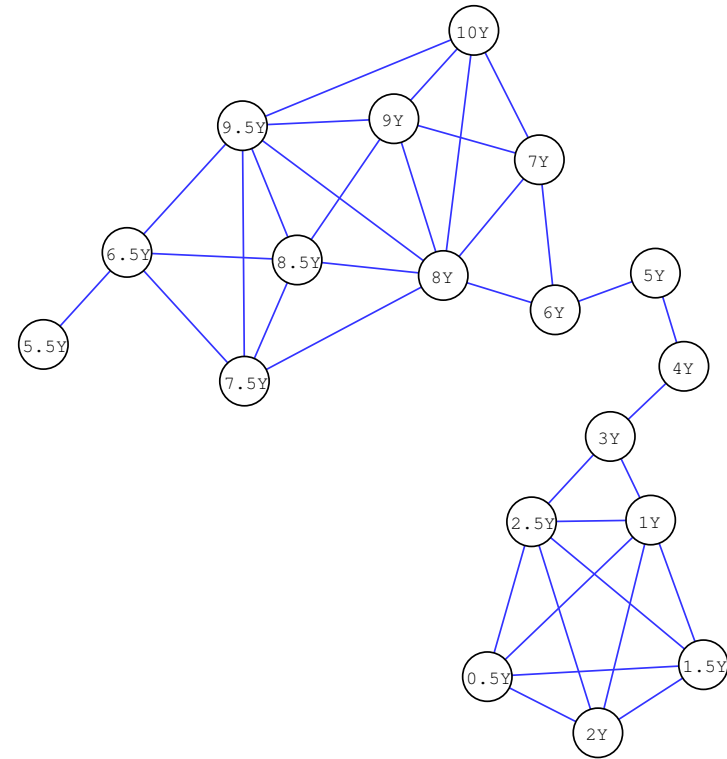


Covariance Selection

On a slightly larger scale. . .



Before



After

Applications & Related Work

- **Gene expression data.** The sample data is composed of gene expression vectors and we want to isolate links in the expression of various genes. See Dobra, Hans, Jones, Nevins, Yao & West (2004), Dobra & West (2004) for example.
- **Speech Recognition.** See Bilmes (1999), Bilmes (2000) or Chen & Gopinath (1999).
- **Finance.** Covariance estimation. Identifying correlation between **idiosyncratic** components of asset returns.
- Related work by Dahl, Roychowdhury & Vandenberghe (2005): interior point methods for large, sparse MLE.
- See also d'Aspremont et al. (2007) on sparse principal component analysis (PCA).

Outline

- Small Mean Reverting Portfolios
 - Canonical decomposition
 - Sparse generalized eigenvalue problems
 - Estimation and trading
 - Numerical results
- Covariance Selection
 - Introduction
 - **Robust Maximum Likelihood Estimation**
 - Algorithms
 - Numerical Results

Maximum Likelihood Estimation

- We can estimate Σ by solving the following maximum likelihood problem:

$$\max_{X \in \mathbf{S}^n} \log \det X - \mathbf{Tr}(SX)$$

- This problem is convex, has an explicit answer $\Sigma = S^{-1}$ if $S \succ 0$.
- Problem here: how do we make Σ^{-1} **sparse**?
- In other words, how do we efficiently choose I and J ?
- Solution: penalize the MLE.

AIC and BIC

Original solution in Akaike (1973), **penalize** the likelihood function:

$$\max_{X \in \mathbf{S}^n} \log \det X - \mathbf{Tr}(SX) - \rho \mathbf{Card}(X)$$

where $\mathbf{Card}(X)$ is the number of nonzero elements in X .

- Set $\rho = 2/(m + 1)$ for the Akaike Information Criterion (**AIC**).
- Set $\rho = \frac{\log(m+1)}{(m+1)}$ for the Bayesian Information Criterion (**BIC**).

Of course, this is a hard combinatorial problem. . .

Convex Relaxation

- We can form a **convex relaxation** of AIC or BIC penalized MLE

$$\max_{X \in \mathbf{S}^n} \log \det X - \mathbf{Tr}(SX) - \rho \mathbf{Card}(X)$$

replacing $\mathbf{Card}(X)$ by $\|X\|_1 = \sum_{ij} |X_{ij}|$ to solve

$$\max_{X \in \mathbf{S}^n} \log \det X - \mathbf{Tr}(SX) - \rho \|X\|_1$$

- Classic l_1 heuristic: $\|X\|_1$ is a **convex lower bound** on $\mathbf{Card}(X)$.
- See Fazel, Hindi & Boyd (2001) for related applications.

l_1 relaxation

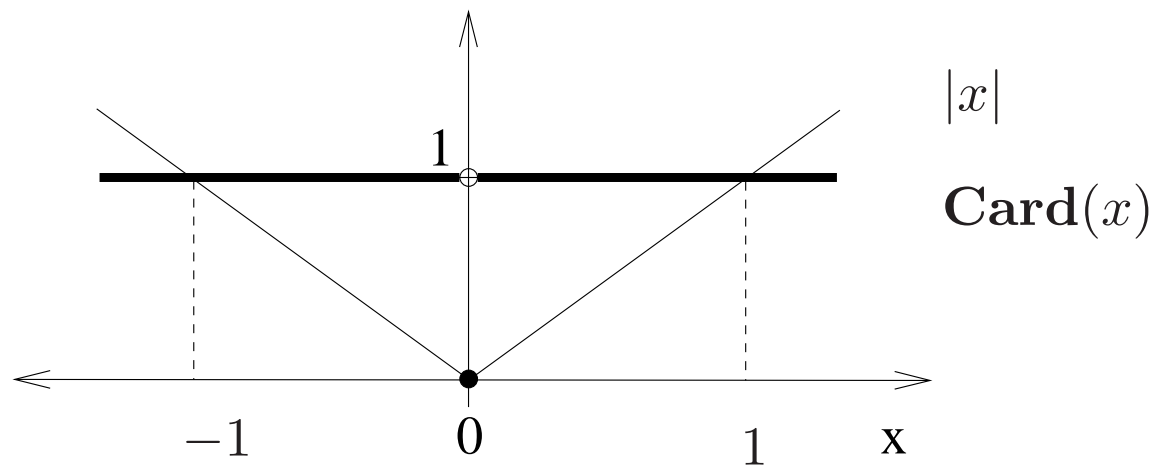
Assuming $|x| \leq 1$, this relaxation replaces:

$$\mathbf{Card}(x) = \sum_{i=1}^n 1_{\{x_i \neq 0\}}$$

with

$$\|x\|_1 = \sum_{i=1}^n |x_i|$$

Graphically, this is:



Robustness

- This penalized MLE problem can be rewritten:

$$\max_{X \in \mathbf{S}^n} \min_{|U_{ij}| \leq \rho} \log \det X - \mathbf{Tr}((S + U)X)$$

- This can be interpreted as a **robust MLE** problem with componentwise noise of magnitude ρ on the elements of S .
- The relaxed **sparsity** requirement is equivalent to a **robustification**.
- See d'Aspremont et al. (2007) for similar results on sparse PCA.

Outline

- Small Mean Reverting Portfolios
 - Canonical decomposition
 - Sparse generalized eigenvalue problems
 - Estimation and trading
 - Numerical results
- Covariance Selection
 - Introduction
 - Robust Maximum Likelihood Estimation
 - **Algorithms**
 - Numerical Results

Algorithms

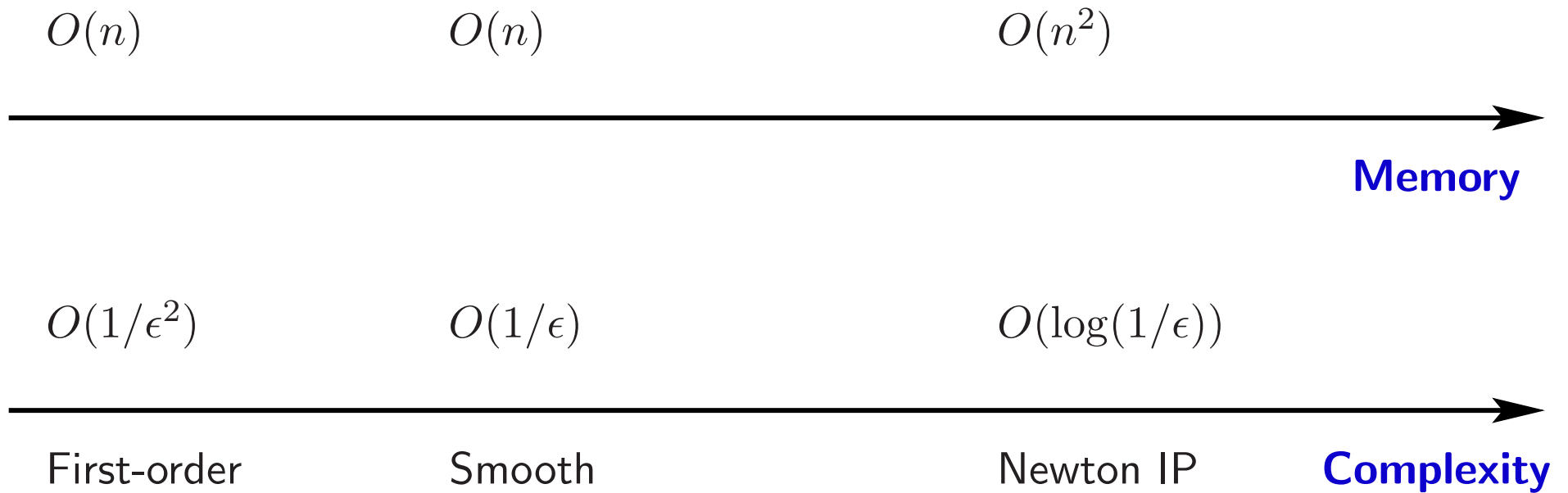
- We need to solve:

$$\max_{X \in \mathbf{S}^n} \log \det X - \mathbf{Tr}(SX) - \rho \|X\|_1$$

- For medium size problems, this can be done using interior point methods.
- In practice, we need to solve **very large, dense** instances. . .
- The $\|X\|_1$ penalty implicitly introduces $O(n^2)$ linear constraints and makes interior point methods too expensive.

Algorithms

Complexity options. . .



Algorithms

Here, we can exploit problem structure

- Our problem here has a particular **min-max** structure:

$$\max_{X \in \mathbf{S}^n} \min_{|U_{ij}| \leq \rho} \log \det X - \mathbf{Tr}((S + U)X)$$

- This min-max structure means that we use prox function algorithms by Nesterov (2005) (see also Nemirovski (2004)) to solve large, dense problem instances.
- We also detail a “greedy” block-coordinate descent method with good empirical performance.

Nesterov's method

Assuming that a problem can be written according to this min-max model, the algorithm works as follows. . .

- **Regularization.** Add strongly convex penalty inside the min-max representation to produce an ϵ -approximation of f with Lipschitz continuous gradient (generalized Moreau-Yosida regularization step, see Lemaréchal & Sagastizábal (1997) for example).
- **Optimal first order minimization.** Use optimal first order scheme for Lipschitz continuous functions detailed in Nesterov (1983) to solve the regularized problem.

Caveat: Only efficient if the subproblems involved in these steps can be solved explicitly or very efficiently. . .

Outline

- Small Mean Reverting Portfolios
 - Canonical decomposition
 - Sparse generalized eigenvalue problems
 - Estimation and trading
 - Numerical results
- Covariance Selection
 - Introduction
 - Robust Maximum Likelihood Estimation
 - Algorithms
 - **Numerical Results**

Numerical Examples

Generate random examples:

- Take a sparse, random p.s.d. matrix $A \in \mathbf{S}^n$
- We add a uniform noise with magnitude σ to its inverse

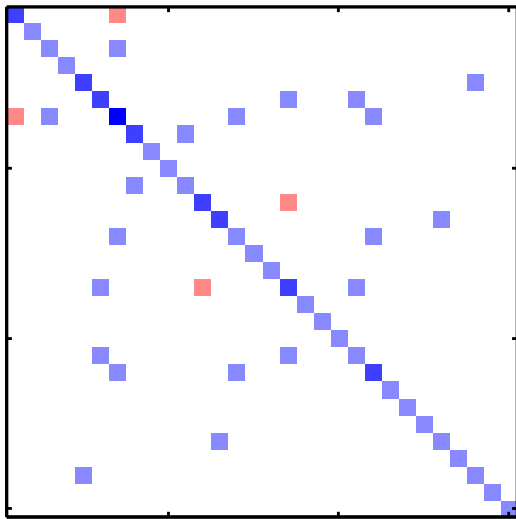
We then solve the penalized MLE problem (or the modified one):

$$\max_{X \in \mathbf{S}^n} \log \det X - \mathbf{Tr}(SX) - \rho \|X\|_1$$

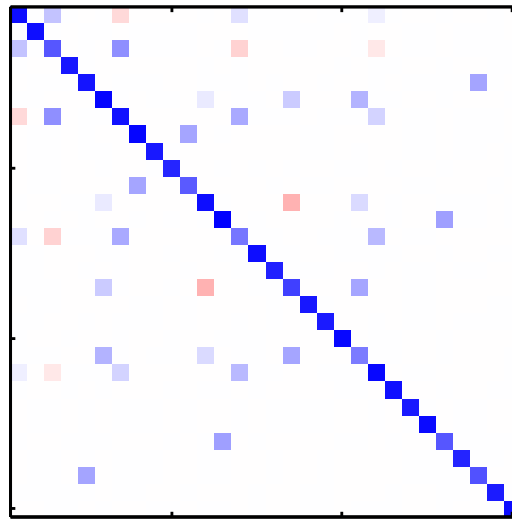
and compare the solution with the original matrix A .

Numerical Examples

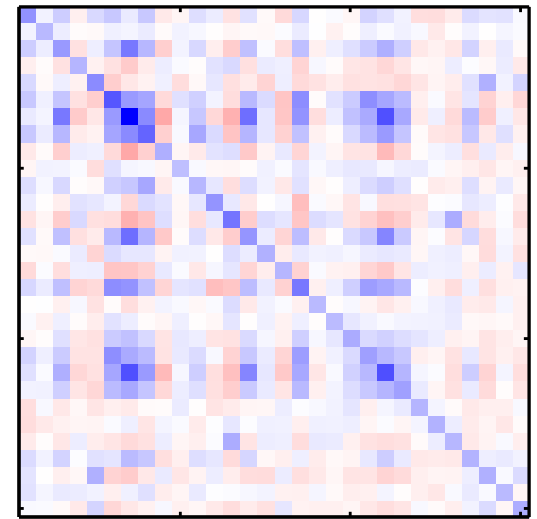
A basic example. . .



Original inverse A



Solution for $\rho = \sigma$

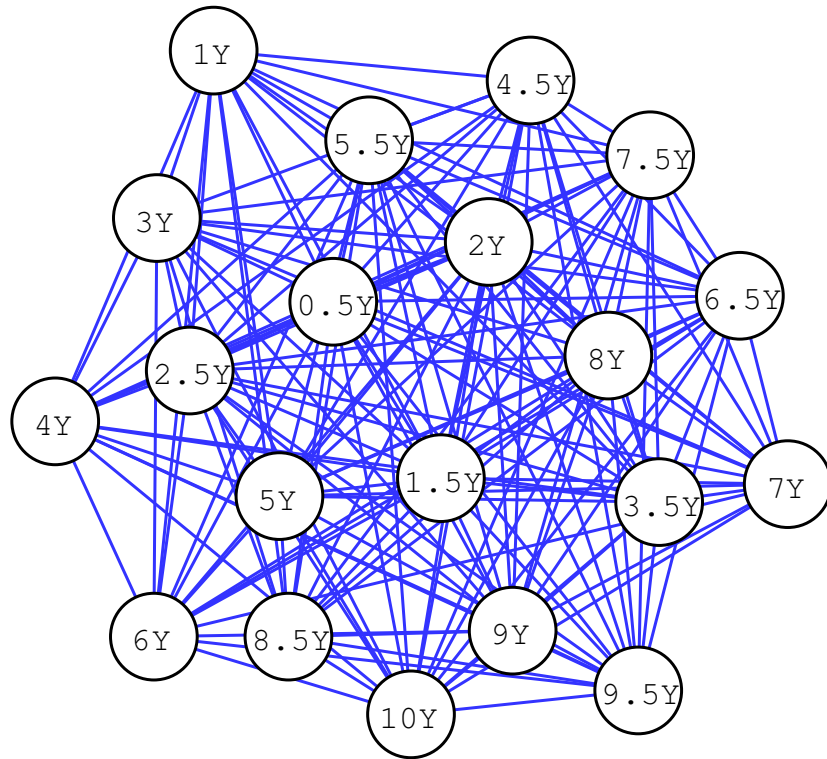


Noisy inverse Σ^{-1}

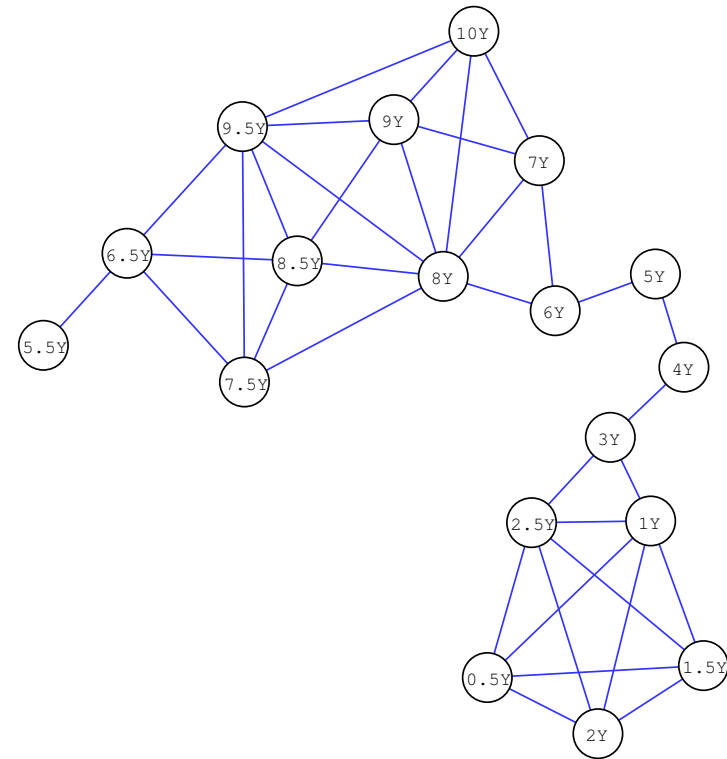
The original inverse covariance matrix A , the noisy inverse Σ^{-1} and the solution.

Covariance Selection

Forward rates covariance matrix for maturities ranging from 0.5 to 10 years.

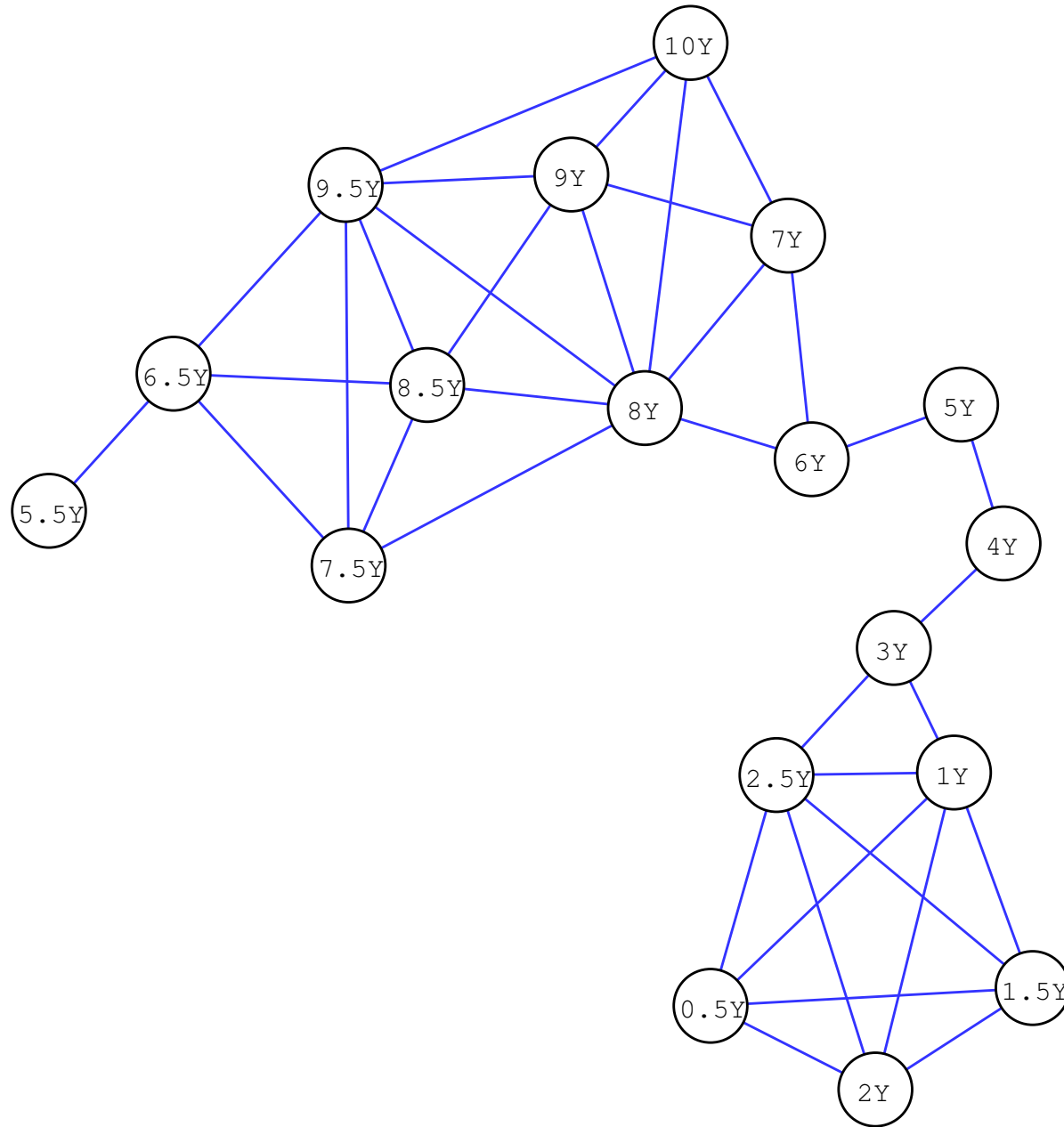


$$\rho = 0$$

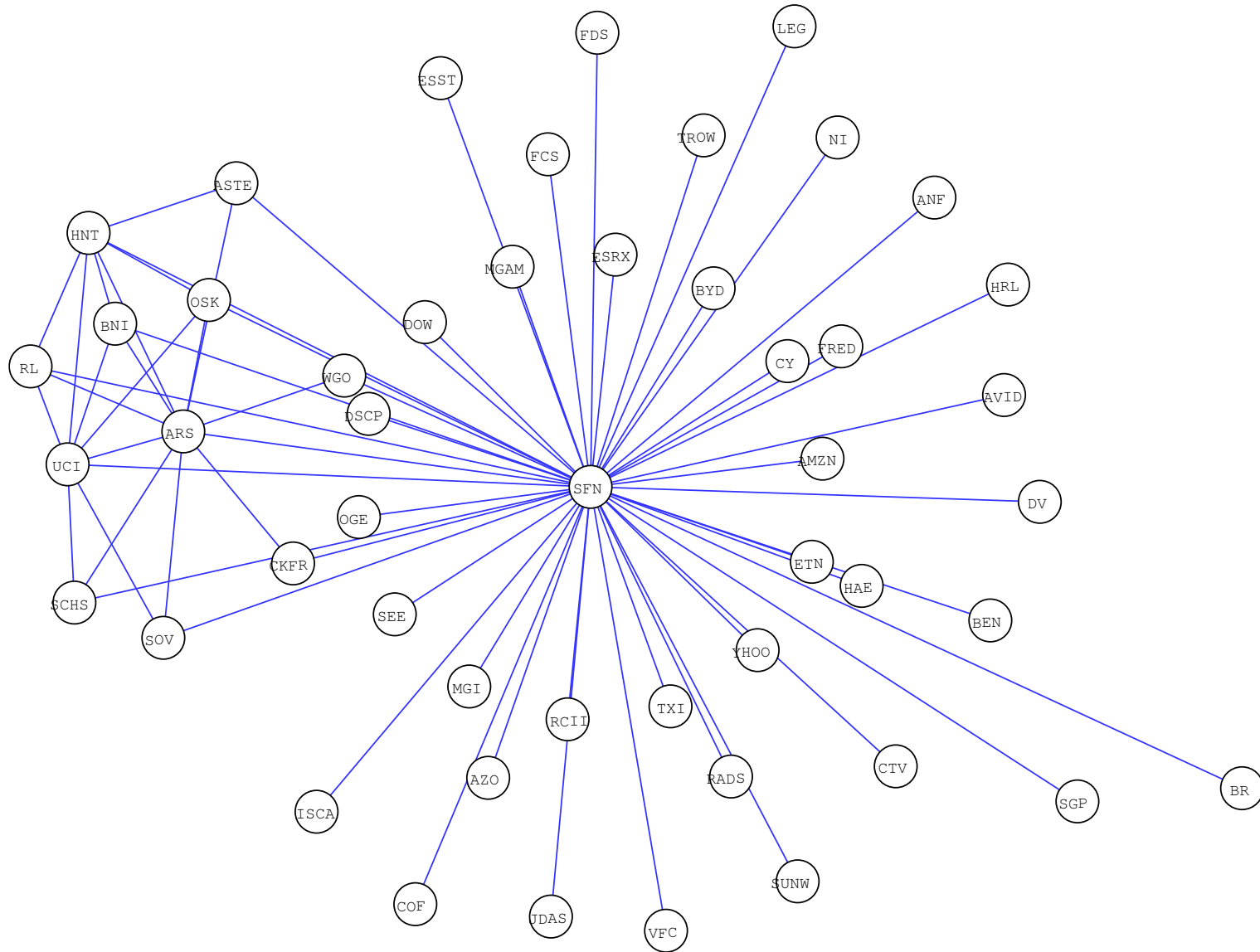


$$\rho = .01$$

Zoom. . .



S&P 500



Conclusion

- Tractable solution to sparse canonical decompositions.
- The tradeoff between number of assets in the portfolio and mean reversion is quite favorable.

- A convex relaxation for sparse covariance selection.
- Robustness interpretation.
- Two algorithms for dense large-scale instances.
- Precision requirements? Thresholding? How do to fix ρ ? . . .

If you have financial applications in mind. . .

Network graphs generated using Cytoscape.

References

- Akaike, J. (1973), Information theory and an extension of the maximum likelihood principle, in B. N. Petrov & F. Csaki, eds, 'Second international symposium on information theory', Akademiai Kiado, Budapest, pp. 267–281.
- Alexander, C. (1999), 'Optimal hedging using cointegration', *Philosophical Transactions: Mathematical, Physical and Engineering Sciences* **357**(1758), 2039–2058.
- Bewley, R., Orden, D., Yang, M. & Fisher, L. (1994), 'Comparison of Box-Tiao and Johansen Canonical Estimators of Cointegrating Vectors in VEC (1) Models', *Journal of Econometrics* **64**, 3–27.
- Bilmes, J. A. (1999), 'Natural statistic models for automatic speech recognition', *Ph.D. thesis, UC Berkeley, Dept. of EECS, CS Division* .
- Bilmes, J. A. (2000), 'Factored sparse inverse covariance matrices', *IEEE International Conference on Acoustics, Speech, and Signal Processing* .
- Box, G. E. & Tiao, G. C. (1977), 'A canonical analysis of multiple time series', *Biometrika* **64**(2), 355.
- Campbell, J. & Viceira, L. (1999), 'Consumption and Portfolio Decisions When Expected Returns Are Time Varying', *The Quarterly Journal of Economics* **114**(2), 433–495.
- Candès, E. J. & Tao, T. (2005), 'Decoding by linear programming', *Information Theory, IEEE Transactions on* **51**(12), 4203–4215.
- Chen, S., Donoho, D. & Saunders, M. (2001), 'Atomic decomposition by basis pursuit.', *SIAM Review* **43**(1), 129–159.
- Chen, S. S. & Gopinath, R. A. (1999), 'Model selection in acoustic modeling', *EUROSPEECH* .
- Dahl, J., Roychowdhury, V. & Vandenberghe, L. (2005), 'Maximum likelihood estimation of gaussian graphical models: numerical implementation and topology selection', *Preprint* .
- d'Aspremont, A., Banerjee, O. & El Ghaoui, L. (2006), 'First-order methods for sparse covariance selection', *Arxiv math/0609812* .
- d'Aspremont, A., El Ghaoui, L., Jordan, M. & Lanckriet, G. R. G. (2007), 'A direct formulation for sparse PCA using semidefinite programming', *SIAM Review* **49**(3), 434–448.
- Dempster, A. (1972), 'Covariance selection', *Biometrics* **28**, 157–175.
- Dobra, A., Hans, C., Jones, B., Nevins, J. J. R., Yao, G. & West, M. (2004), 'Sparse graphical models for exploring gene expression data', *Journal of Multivariate Analysis* **90**(1), 196–212.
- Dobra, A. & West, M. (2004), 'Bayesian covariance selection', *working paper* .
- Donoho, D. L. & Tanner, J. (2005), 'Sparse nonnegative solutions of underdetermined linear equations by linear programming', *Proceedings of the National Academy of Sciences* **102**(27), 9446–9451.
- Engle, R. & Granger, C. (1987), 'Cointegration and error correction: representation, estimation and testing', *Econometrica* **55**(2), 251–276.
- Fama, E. & French, K. (1988), 'Permanent and Temporary Components of Stock Prices', *The Journal of Political Economy* **96**(2), 246–273.

- Fazel, M., Hindi, H. & Boyd, S. (2001), 'A rank minimization heuristic with application to minimum order system approximation', *Proceedings American Control Conference* **6**, 4734–4739.
- Gatev, E., Goetzmann, W. & Rouwenhorst, K. (2006), 'Pairs Trading: Performance of a Relative-Value Arbitrage Rule', *Review of Financial Studies* **19**(3), 797.
- Grossman, S. & Vila, J. (1992), 'Optimal Dynamic Trading with Leverage Constraints', *The Journal of Financial and Quantitative Analysis* **27**(2), 151–168.
- Hughes, T. R., Marton, M. J., Jones, A. R., Roberts, C. J., Stoughton, R., Armour, C. D., Bennett, H. A., Coffey, E., Dai, H., He, Y. D., Kidd, M. J., King, A. M., Meyer, M. R., Slade, D., Lum, P. Y., Stepaniants, S. B., Shoemaker, D. D., Gachotte, D., Chakraborty, K., Simon, J., Bard, M. & Friend, S. H. (2000), 'Functional discovery via a compendium of expression profiles', *Cell* **102**(1), 109–126.
- Johansen, S. (1991), 'Estimation and Hypothesis Testing of Cointegration Vectors in Gaussian Vector Autoregressive Models', *Econometrica* **59**(6), 1551–1580.
- Jurek, J. & Yang, H. (2006), Dynamic portfolio selection in arbitrage, Technical report, Working Paper, Harvard Business School.
- Kim, T. & Omberg, E. (1996), 'Dynamic Nonmyopic Portfolio Behavior', *The Review of Financial Studies* **9**(1), 141–161.
- Lemaréchal, C. & Sagastizábal, C. (1997), 'Practical aspects of the Moreau-Yosida regularization: theoretical preliminaries', *SIAM Journal on Optimization* **7**(2), 367–385.
- Liu, J. & Longstaff, F. (2004), 'Losing Money on Arbitrage: Optimal Dynamic Portfolio Choice in Markets with Arbitrage Opportunities', *Review of Financial Studies* **17**(3).
- Natsoulis, G., El Ghaoui, L., Lanckriet, G., Tolley, A., Leroy, F., Dunlea, S., Eynon, B., Pearson, C., Tugendreich, S. & Jarnagin, K. (2005), 'Classification of a large microarray data set: algorithm comparison and analysis of drug signatures', *Genome Research* **15**, 724–736.
- Nemirovski, A. (2004), 'Prox-method with rate of convergence $O(1/T)$ for variational inequalities with lipschitz continuous monotone operators and smooth convex-concave saddle point problems', *SIAM Journal on Optimization* **15**(1), 229–251.
- Nesterov, Y. (1983), 'A method of solving a convex programming problem with convergence rate $O(1/k^2)$ ', *Soviet Mathematics Doklady* **27**(2), 372–376.
- Nesterov, Y. (2005), 'Smooth minimization of non-smooth functions', *Mathematical Programming* **103**(1), 127–152.
- Poterba, J. M. & Summers, L. H. (1988), 'Mean reversion in stock prices: Evidence and implications', *Journal of Financial Economics* **22**(1), 27–59.
- Tibshirani, R. (1996), 'Regression shrinkage and selection via the LASSO', *Journal of the Royal statistical society, series B* **58**(1), 267–288.
- Toh, K. C., Todd, M. J. & Tutuncu, R. H. (1999), 'SDPT3 – a MATLAB software package for semidefinite programming', *Optimization Methods and Software* **11**, 545–581.
- Wachter, J. (2002), 'Portfolio and Consumption Decisions under Mean-Reverting Returns: An Exact Solution for Complete Markets', *The Journal of Financial and Quantitative Analysis* **37**(1), 63–91.

Xiong, W. (2001), 'Convergence trading with wealth effects: an amplification mechanism in financial markets', *Journal of Financial Economics* **62**(2), 247–292.