Denoising and Detoning

Machine learning for asset managers

Agenda

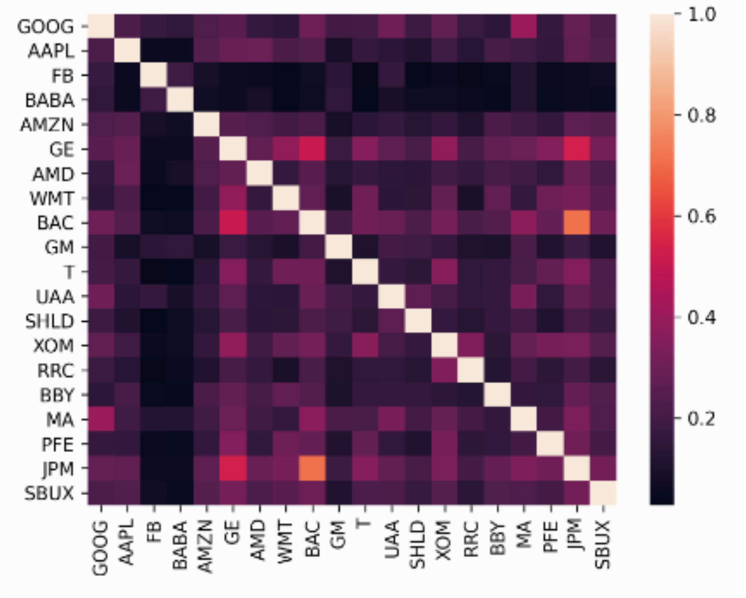
- Background
- Random Matrix Theory:
  - Marcenko-Pastur PDF and theorem
- Matrix Denoising
- Matrix Detoning
- Illustration on portfolio optimization
- Alternative Approaches
- References
- Appendix

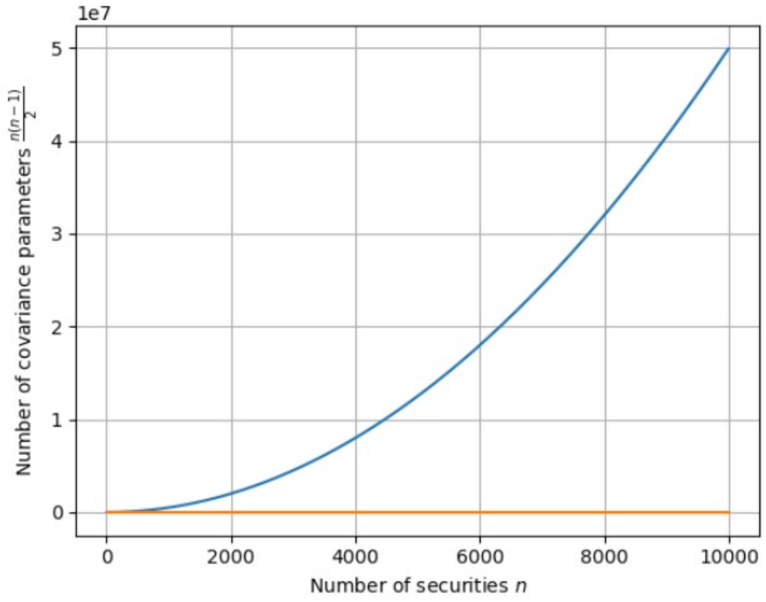
Background

- Covariances are ubiquitous in finance
  - Run regressions
  - Estimate risks
  - Optimize portfolios
  - Monte Carlo scenario analysis
  - Clustering
  - Dimension reduction etc

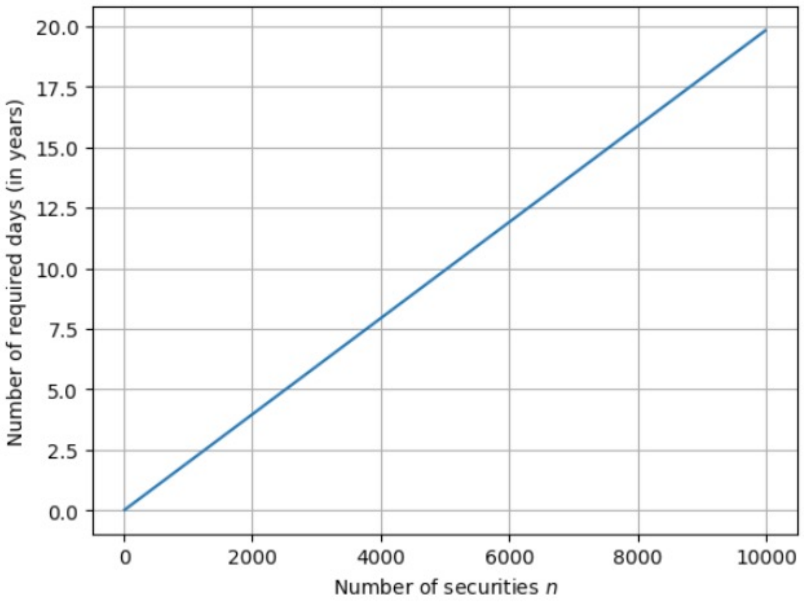
$$X \in \mathbb{R}^{T \times N}$$

$$\Sigma = \frac{1}{n-1}(X - \overline{X})^T (X - \overline{X})$$

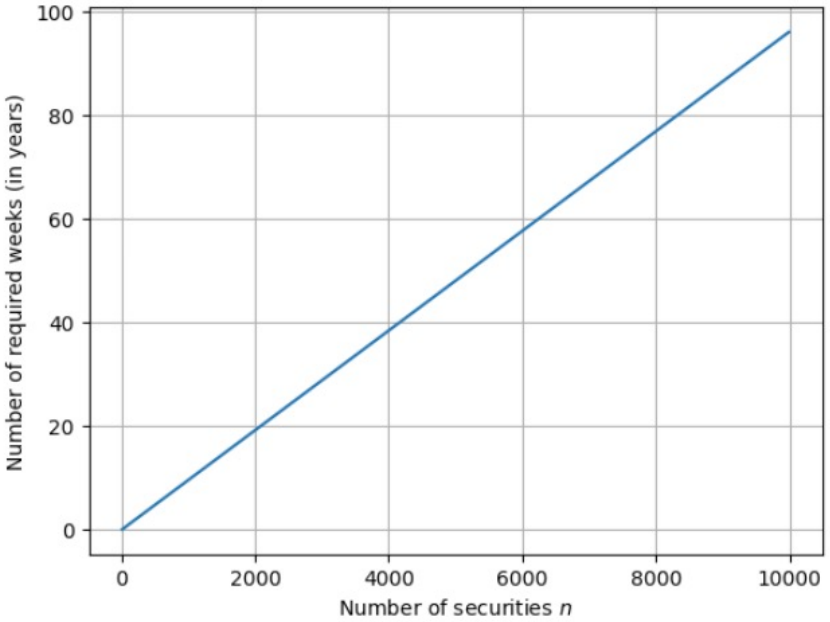




Years of data if we want daily covariance



Years of data if we want weekly covariance



Hard to estimate true covariances due to finite/thin data

Sample covariances

Eigen-value decomposition

$$\Sigma v = \lambda v$$

- Most real world financial problems don't have enough data to estimate high dimensional covariances
- This situation leas to ill-conditioned matrices:
  - Small changes in the input data can lead to large changes in the output
- Ill-conditioning can be quantified by the condition number:

$$\frac{\lambda_{max}}{\lambda_{min}}$$

- Solution: apply shrinkage which make the covariance matrix close to diagonal, therefore reducing its condition number

$$\Sigma \leftarrow (1-\alpha)\Sigma + \alpha \frac{\mathbf{Tr}(\Sigma)}{N} \mathbb{I}_N$$

- Problem: shrinkage accomplishes that without discriminating between noise and signal, as a result, it can further eliminate an already weak signal

The Marcenko-Pastur probability density function

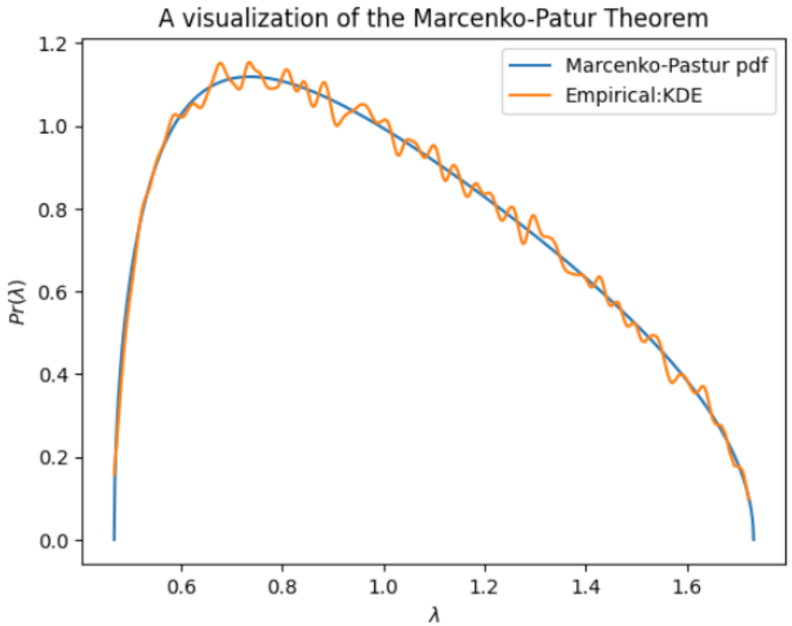
Let :

- X: a random observation matrix in $\mathbb{R}^{T \times N}$
- T: number of observations
- N: number of securities
- The underlying process generating the observations has zero mean and variance σ^2
- $q = \frac{T}{N}$
- $C = \frac{1}{T} X^T X$
- λ refers to the eigenvalues of C

The Marcenko-Pastur probability density function (PDF) is defined as:

$$f(\lambda) = \begin{cases} \frac{q}{2\pi\lambda\sigma^2}\sqrt{(\lambda_{\max} - \lambda)(\lambda - \lambda_{\min})} & \text{if } \lambda_{\min} \le \lambda \le \lambda_{\max} \\ 0 & \text{otherwise} \end{cases}$$ where

- $\lambda_{\max} = \sigma^2(1 + \sqrt{\frac{1}{q}})^2$
- $\lambda_{\min} = \sigma^2(1 - \sqrt{\frac{1}{q}})^2$



Theorem: As:

- $N \longrightarrow +\infty$
- $T \longrightarrow +\infty$
- $1 < q < +\infty$

Then the eigenvalues λ asymptotically converge to the Marcenko-Pastur PDF.

- when $\sigma^2 = 1$, C is the correlation matrix associated with X
- Eigenvalues $\lambda \in [\lambda_{\min}, \lambda_{\max}]$ are consistent with random/noise behavior
- Eigenvalues $\lambda \notin [\lambda_{\min}, \lambda_{\max}]$ are consistent with nonrandom/signal behavior
- Specifically, we associate $\lambda \in [0, \lambda_{\max}]$ with noise

How is Marcenko-Pastur is being fit ?

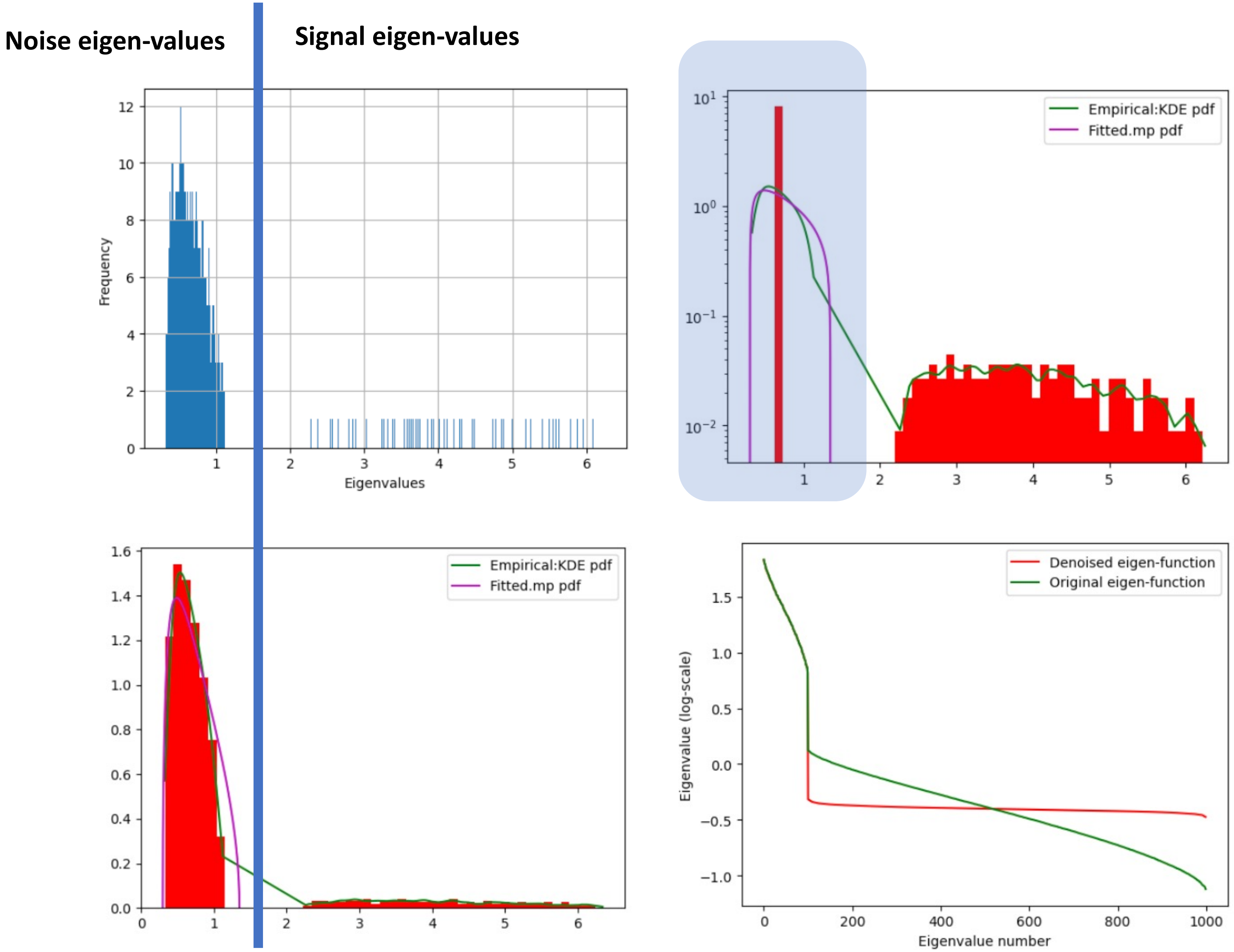
To fit the Marcenko-Pastur distribution to an arbitrary array of eigen vectors, we would need to identfiy the main parameters:

- σ^2
- q

The above can be achieved by:

- finding the optimal non-parametric empirical density
    - the search my require cross-validation to find the optimal Kernel and its parameters
- Fix a search grid for reasonable values of $q > 1$
- For each value of q:
    - find the optimal σ^2 that minimizes the sum of squared errors (SSE) between Marcenko-Pastur PDF and the Kernel PDF
        - Every candidate solution fully defines the MP PDF
        - Sample from the MP PDF
        - Evaluate the sampled data with the Kernel density
        - Compute the SSE between MP and Kernel PDF values
- Return q and σ^2 that minimize SSE
- Compute $\lambda_{\max} = \sigma^2(1 + \sqrt{\frac{1}{q}})^2$
- Adjust $\sigma^2 \approx \sigma^2(1 - \frac{\lambda_{\max}}{N})$

Constant Residual Eigenvalue Method- Laloux’ approach



Let

- $\Lambda = \{\lambda_n\}_{n=1,\cdots,N}$ be the set of all eigenvalues, ordered descending
- i be the position such that $\lambda_i > \lambda_{\max}$ and $\lambda_i < \lambda_{\min}$
- Then we set :
  $\lambda_j = \frac{1}{N-i} \sum_{k=i+1}^{N} \lambda_k$
  for $j = i+1, \cdots, N$, hence preserving the trace of the matrix.
- $\tilde{\Lambda}$ the new corrected eigenvalues that we refer to as
- Given the eigenvector decomposition $CW = W\Lambda$, then
  - $\tilde{C} = W\tilde{\Lambda}W^T$
  - $C = \tilde{C}\left[\left(\mathrm{diag}(\tilde{C})\right)^{\frac{1}{2}}\left(\mathrm{diag}(\tilde{C})\right)^{\frac{1}{2}T}\right]^{-1}$

Targeted Shrinkage

LeDoit and Wold shrinkage reduces the condition number of ill-conditioned covariance matrices but it applies to the full spectrum without discriminating between noise and signal.

Alternatively, since we can use MD PDF to distinguish between noise and signal, we can target the application of the shrinkage strictly to the noise eigenvectors.

Let

- W_{noise} and Λ_{noise} are the eigenvectors and eigenvalues associated with the **noise** part of the spectrum $\{n|\lambda_n \leq \lambda_{\max}\}$
- W_{signal} and Λ_{signal} are the eigenvectors and eigenvalues associated with the **signal** part of the spectrum $\{n|\lambda_n > \lambda_{\max}\}$

Consider the correlation matrix

$$C_1 = W_{\text{signal}}\Lambda_{\text{signal}}W_{\text{signal}}^T + \underbrace{\alpha W_{\text{noise}}\Lambda_{\text{noise}}W_{\text{noise}}^T + (1-\alpha)\text{diag}\left[W_{\text{noise}}\Lambda_{\text{noise}}W_{\text{noise}}^T\right]}_{\text{Shrinkage applies to the noise part}}$$

where

- α regulates the amount of shrinkage applied to the noise part.

Detoning

- Financial correlation matrices usually incorporate a market component
- The market component is characterized by the first eigenvector, with loadings
$$W_{n,1} \approx \sqrt{N}, n = 1, \cdots, N$$
- A market component affects every item of the covariance matrix
- In the context of clustering applications, it is useful to remove the market component, if it exists (a hypothesis that can be tested statistically)
  - The reason is, it is more difficult to cluster a correlation matrix with a strong market component, because the algorithm will struggle to find dissimilarities across clusters.
  - By removing the market component, we allow a greater portion of the correlation to be explained by components that affect specific subsets of the securities.
  - It is similar to removing a loud tone that prevents us from hearing other sounds.
- Detoning is the principal components analysis analogue to computing beta-adjusted (or market-adjusted) returns in regression analysis.
- We can remove the market component from the the denoised correlation matrix, C_1, to construct the detoned correlation matrix
$$\tilde{C}_2 = C_1 - W_M \Lambda_M W_M^T$$
$$C_2 = \tilde{C}_2 \left[\left(\mathrm{diag}(\tilde{C}_2)\right)^{\frac{1}{2}} \left(\mathrm{diag}(\tilde{C}_2)\right)^{\frac{1}{2}^T} \right]^{-1}$$
- The detoned matrix is singular (non invertible) , which might not be an issue for clustering but can be an empediment against portfolio optimization
  - we can optimize a portfolio using the covariance computed based on the selected (non zero) principal components
  - map the optimal allocation f^* back to the original basis
$$\omega^* = W_+ f^*$$
where W_+ contains only the eigenvectors that survived the detoning process.

Application to portfolio optimization (1)

$$\min_{w} w^T \Sigma w$$

subject to:

- $w^T \mu = \mu_p,$
- $w^T 1 = 1,$
- $w \geq 0$

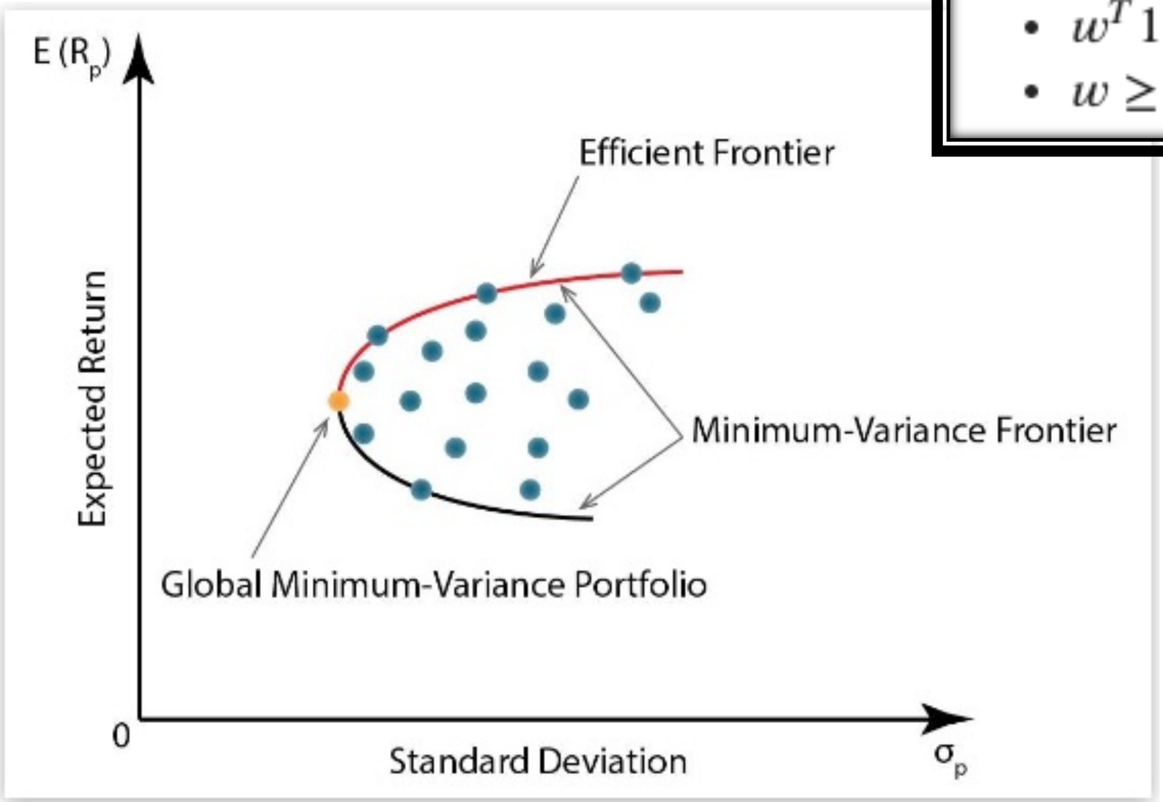
- Given
  - The covariance matrix
  - The expected returns vector
- We can identify:
  - The minimum-variance portfolio

$$\mathbf{w}^* = \frac{\Sigma^{-1}\mathbf{1}}{\mathbf{1}^T\Sigma^{-1}\mathbf{1}}$$

  - The maximum-Sharpe portfolio

$$\mathbf{SR} = \frac{\mu_p - r_f}{\sigma_p}$$

$$\mathbf{w}^* = \frac{\Sigma^{-1}(\mu - r_f\mathbf{1})}{\mathbf{1}^T\Sigma^{-1}(\mu - r_f\mathbf{1})}$$

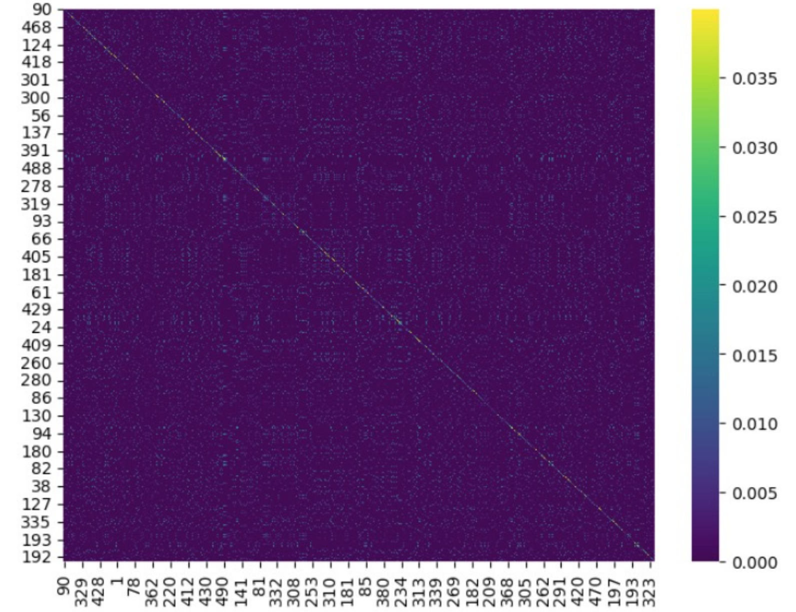


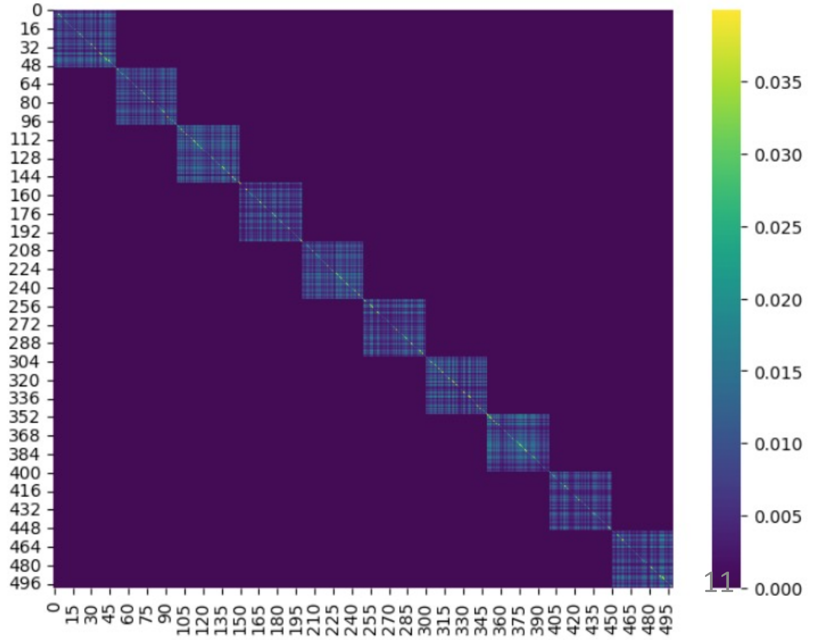
[Source]

Question: can RMT denoising of the covariance achieve better MVP and MSP ?

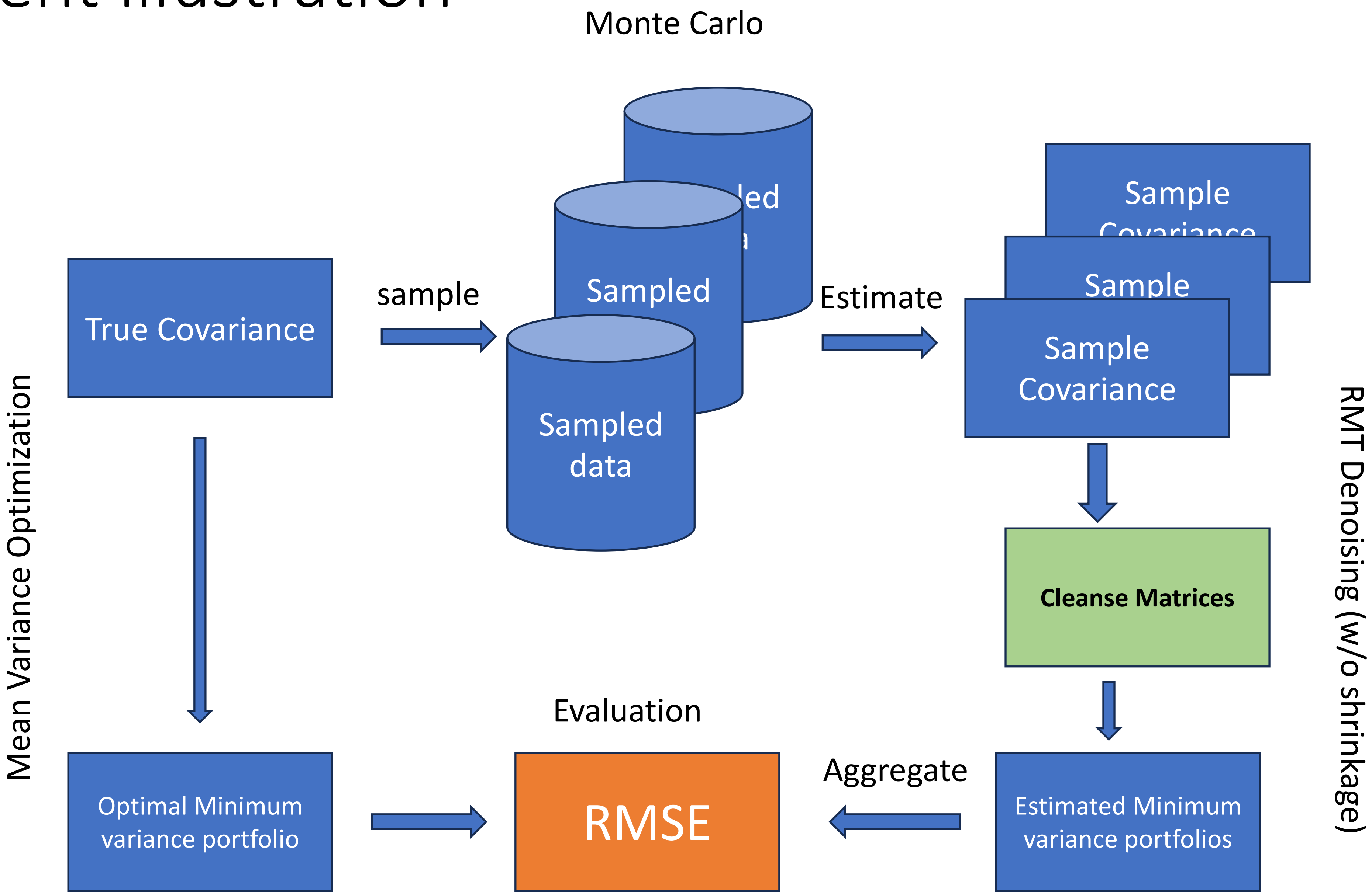
Application to portfolio optimization (2)

- Hypothesis: RMT denoising helps identifying more optimal portfolios
- Validation:
  - we need to design an experiment that allow us to compare portfolios constructed using denoised covariance matrices against *true/target* optimal portfolios.
  - To construct a true/target portfolio, we need to construct a true/target covariance matrix:
    - 10 blocks of 50 each
    - Off diagonal elements within each block have a correlation of 0.5
    - This covariance is a stylized representation of a true (nonempirical) detoned correlation matrix of the S&P 500 where each block is associated with an economic sector
    - Variances are drawn from a uniform distribution between 5% and 20%
    - Vector of means is drawn from a normal distribution with mean and standard deviation equal to the standard deviation from the covariance matrix
  - Use true/target covariance to compute true/target optimal portfolios
  - Sample observations from the true/target matrix
  - Compute corresponding sample covariances
  - Estimate optimal portfolios w/o denoising
  - Evaluate if denoising helps





Experiment illustration



Empirical Results

RMSE for combinations of denoising and shrinkage (minimum variance portfolio)

	Not denoised	Denoised
Not shrunk	4.95e-3	**1.99e-3**
Shrunk	3.45e-3	**1.7e-3**

RMSE for combinations of denoising and shrinkage (maximum Sharpe portfolio)

	Not denoised	Denoised
Not shrunk	9.48e-1	**5.27e-2**
Shrunk	2.77e-1	**5.17e-2**

Alternative approaches

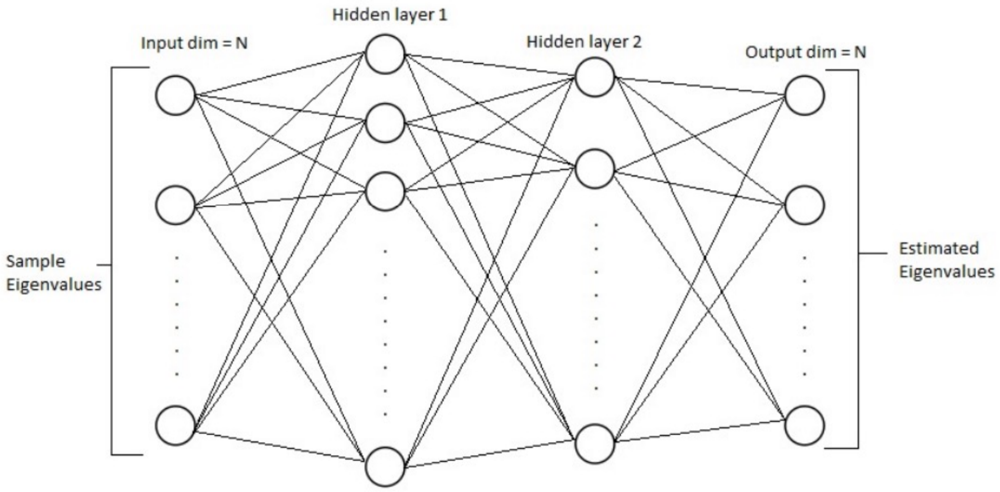
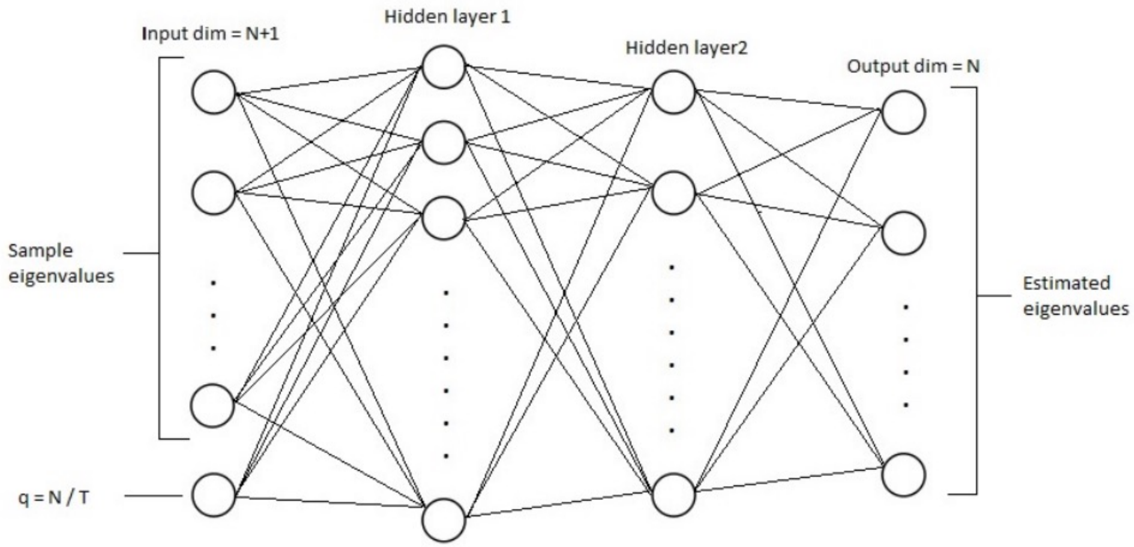


FIG. 2: Denoising autoencoder



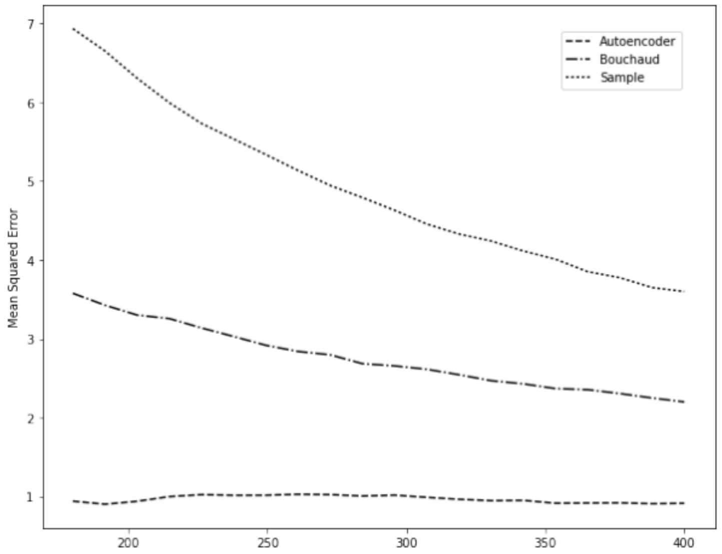


FIG. 7: MSE for T between 180 and 400 (q between 1 and 0.45)

Soufiane Hayou (2017), [Cleaning the correlation matrix with a denoising auto-encoder]

Rotationally Invariant Estimators (RIE)

Cleaning large correlation matrices: Tools from Random Matrix Theory

CrossMark

Joël Bun [a,b,*], Jean-Philippe Bouchaud [a], Marc Potters [a]

[a] *Capital Fund Management, 23–25, rue de l'Université, 75 007 Paris, France*
[b] *LPTMS, CNRS, Univ. Paris-Sud, Université Paris-Saclay, 91405 Orsay, France*

A R T I C L E   I N F O

Article history:
Accepted 24 October 2016
Available online 9 November 2016
editor: H. Orland

Keywords:
Random Matrix Theory
High dimensional statistics
Correlation matrix
Spectral decomposition
Rotational invariant estimator

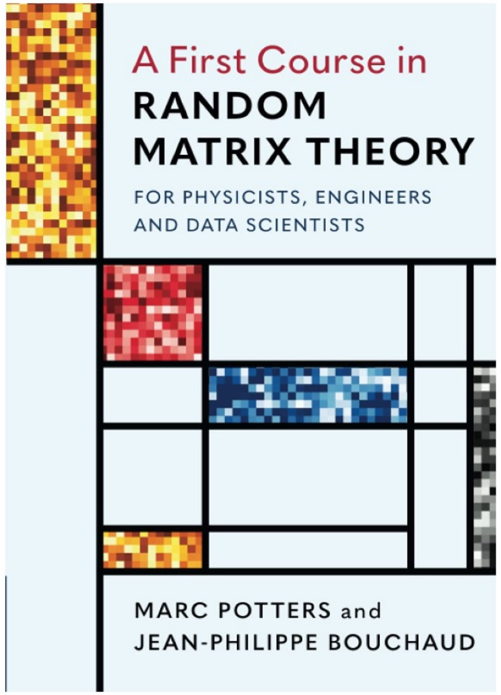
A B S T R A C T

This review covers recent results concerning the estimation of large covariance matrices using tools from Random Matrix Theory (RMT). We introduce several RMT methods and analytical techniques, such as the Replica formalism and Free Probability, with an emphasis on the Marčenko–Pastur equation that provides information on the resolvent of multiplicatively corrupted noisy matrices. Special care is devoted to the statistics of the eigenvectors of the empirical correlation matrix, which turn out to be crucial for many applications. We show in particular how these results can be used to build consistent "Rotationally Invariant" estimators (RIE) for large correlation matrices when there is no prior on the structure of the underlying process. The last part of this review is dedicated to some real-world applications within financial markets as a case in point. We establish empirically the efficacy of the RIE framework, which is found to be superior in this case to all previously proposed methods. The case of additively (rather than multiplicatively) corrupted noisy matrices is also dealt with in a special Appendix. Several open problems and interesting technical developments are discussed throughout the paper.

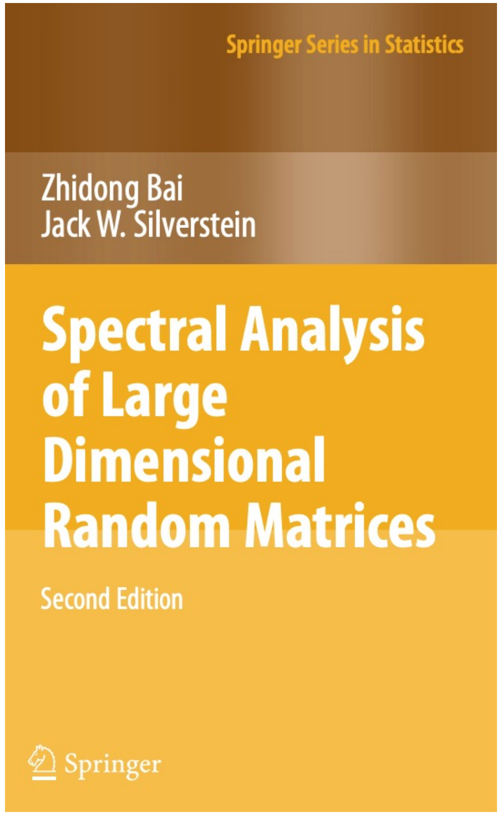
Source: Joel Bun, Jean-Philippe Bouchaud, Marc Potters, (2016) [Cleaning large correlation matrices: Tools from Random Matrix Theory]

References

- Ledoit, O., and M. Wolf (2004): "[A Well-Conditioned Estimator for Large- Dimensional Covariance Matrices]." Journal of Multivariate Analysis, Vol. 88, No. 2, pp. 365–411.
- Laloux, L., P. Cizeau, J. P. Bouchaud, and M. Potters (2000): "[Random Matrix Theory and Financial Correlations]." International Journal of Theoretical and Applied Finance, Vol. 3, No. 3, pp. 391–97.
- Lewandowski, D., D. Kurowicka, and H. Joe (2009): "[Generating Random Correlation Matrices Based on Vines and Extended Onion Method]." Journal of Multivariate Analysis, Vol. 100, pp. 1989–2001.
- *Tangential: RMT is relevant in understanding DL generalizability*
  - Charles Martin et al (2021), [Implicit Self-Regularization in Deep Neural Networks: Evidence from Random Matrix Theory and Implications for Learning.]



[Source]



[source]

Appendix

Deriving minimum variance portfolios

$$\min_{w} w^T \Sigma w$$

$$\text{subject to : } \; w^T 1 = 1$$

The Lagrangian is:

$$\mathcal{L}(w, \lambda) = w^T \Sigma w - \lambda(\sum_{i=1}^{n} w_i - 1), \lambda > 0$$

Taking the gradient with respect of w:

$$\frac{\partial \mathcal{L}}{\partial w}(w, \lambda) = 2\Sigma w - \lambda$$

Setting $\frac{\partial \mathcal{L}}{\partial w} = 0$:

$$w = \frac{\lambda}{2} \Sigma^{-1} \mathbf{1}$$

Plugging w in the constraint:

$$\frac{\lambda}{2} \mathbf{1}^T \Sigma^{-1} \mathbf{1} = 1$$

therefore

$$\lambda = \frac{2}{\mathbf{1}^T \Sigma^{-1} \mathbf{1}}$$

The optimal minimum variance portfolio is then:

$$w^* = \frac{\Sigma^{-1} \mathbf{1}}{\mathbf{1}^T \Sigma^{-1} \mathbf{1}}$$

Deriving maximum Sharpe portfolio

- Gregory Gundersen does an amazing job deriving the maximum Sharpe portfolio in this blog post.