A Random Matrix Bayesian framework for out-of-sample quadratic optimization

Marc Potters
joint work with Joël Bun and Jean-Philippe Bouchaud
6 November 2013

CFM CAPITAL FUND MANAGEMENT

Outline

Markowitz optimization

- N correlated random variables (zero mean, unit variance)
- Find the linear combination (weight vector $\mathbf{w}$) with minimum variance under a linear constraint
- In matrix notation:

$$\text{variance:} \quad R^2 = \mathbf{w}^T \mathbf{C} \mathbf{w}$$

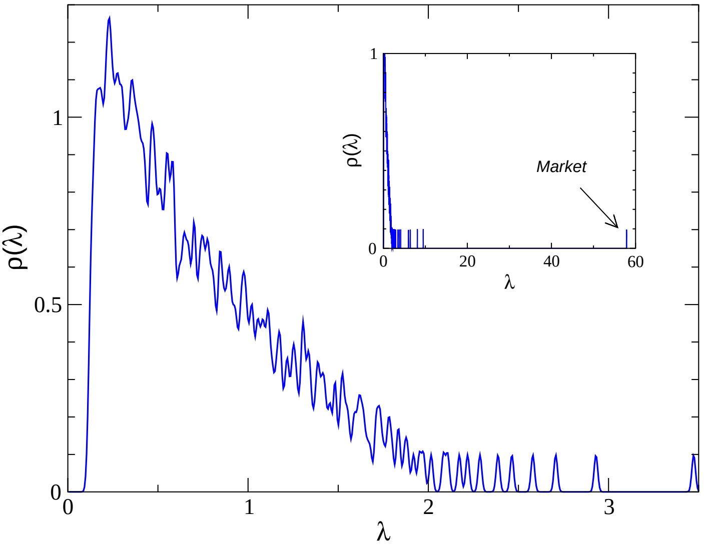
$$\text{gain constraint:} \quad G = \mathbf{w}^T \mathbf{g}$$

$$\text{optimal weights:} \quad \mathbf{w}_C = G \frac{\mathbf{C}^{-1}\mathbf{g}}{\mathbf{g}^T \mathbf{C}^{-1} \mathbf{g}}$$

- $\mathbf{w}$ overweighs small eigenvalues

Empirical Eigenvalues

$N = 406 \qquad T = 1300$



Marčenko and Pastur (1967)

- Given a true $N \times N$ covariance matrix $\mathbf{C}$, what is the resolvant and eigenvalues density of a sample covariance matrix $\mathbf{E}$ (with $q = N/T$)?

$$zG_{\mathbf{E}}(z) = ZG_{\mathbf{C}}(Z)$$

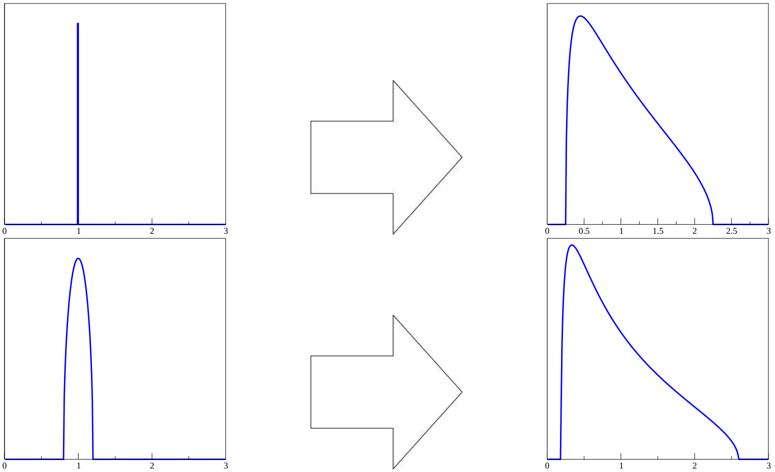
with

$$Z = \frac{z}{1 + q(zG_{\mathbf{E}}(z) - 1)}.$$

- When there are no correlations ($\mathbf{C} = \mathbb{I}$), they gave an explicit result:

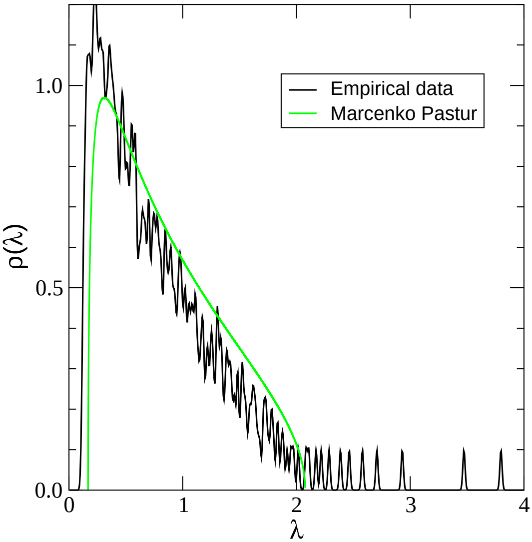
$$\rho(\lambda) = \frac{\sqrt{(\lambda_+ - \lambda)(\lambda - \lambda_-)}}{2\pi q \lambda} \text{ with } \lambda_\pm = (1 \pm \sqrt{q})^2$$

Marčenko-Pastur at work



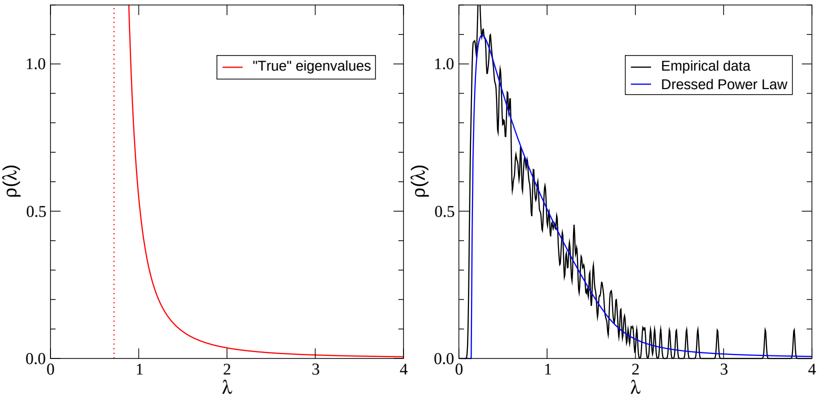
Empirical measure widens the eigenvalue distribution

Empirical Eigenvalues II



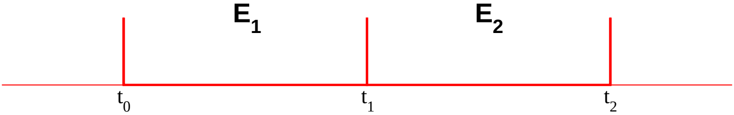
- Marčenko-Pastur distribution fits the data well: most eigenvalues are noise
- But the fit is not perfect: there is 'signal'

Empirical Eigenvalues III



- ‘True’ eigenvalues are mostly clustered around 1 with some very large outliers (power-law tail)

In and out of sample risk



- Estimate $\mathbf{E_1}$ using data from t_0 to t_1 and then test out-of-sample risk with $\mathbf{E_2}$ measured from t_1 to t_2

"In-sample" risk: $$R^2_{\text{in}} = \mathbf{w}_E^T \mathbf{E_1} \mathbf{w}_E = \frac{G^2}{\mathbf{g}^T \mathbf{E}_1^{-1} \mathbf{g}}$$

True minimal risk: $$R^2_{\text{true}} = \mathbf{w}_C^T \mathbf{C} \mathbf{w}_C = \frac{G^2}{\mathbf{g}^T \mathbf{C}^{-1} \mathbf{g}}$$

"Out-of-sample" risk: $$R^2_{\text{out}} = \mathbf{w}_E^T \mathbf{E_2} \mathbf{w}_E = \frac{G^2 \mathbf{g}^T \mathbf{E}_1^{-1} \mathbf{C} \mathbf{E}_1^{-1} \mathbf{g}}{(\mathbf{g}^T \mathbf{E}_1^{-1} \mathbf{g})^2}$$

- Using optimality: $R^2_{\text{in}} \leq R^2_{\text{true}} \leq R^2_{\text{out}}$

Is $\mathbf{E}$ biased?

- The empirical matrix $\mathbf{E_1}$ is an unbiased estimator of $\mathbf{C}$

$$\mathbb{E}[\mathbf{E_1}] = \mathbf{C} \text{ or equivalently } \mathbb{E}[E_{1ij}] = \langle r_i r_j \rangle$$

- $\mathbf{E_1}$ gives an unbiased estimate of the risk of an independent portfolio w.

$$\mathbb{E}\left[\sum_{ij} w_i E_{1ij} w_j\right] = \sum_{ij} w_i C_{ij} w_j = \left\langle \left(\sum_i w_i r_i\right)^2 \right\rangle$$

- But an optimized portfolio such as w_E is not independent of $\mathbf{E_1}$, so using $\mathbf{E_1}$ will generated a biased estimate:

$$\mathbb{E}\left[\sum_{ij} w_{Ei} E_{1ij} w_{Ej}\right] \leq \mathbb{E}\left\langle \left(\sum_i w_{Ei} r_j\right)^2 \right\rangle$$

In and out of sample risk II

- We can make these inequalities more precise
- For N large and T large with $q = N/T$

$$\mathrm{Tr}(\mathbf{E}^{-1}) = -G_{\mathbf{E}}(0) = -\frac{G_{\mathbf{C}}(0)}{1-q} = \frac{\mathrm{Tr}(\mathbf{C}^{-1})}{1-q}$$

$$\mathrm{Tr}(\mathbf{E}^{-1}\mathbf{C}\mathbf{E}^{-1}) = \frac{\mathrm{Tr}(\mathbf{C}^{-1})}{(1-q)^2}$$

- Allowing to compute the different risks:

$$R^2_{\mathrm{in}} = R^2_{\mathrm{true}}(1-q) \quad \text{and} \quad R^2_{\mathrm{out}} = \frac{R^2_{\mathrm{true}}}{1-q}$$

- This result is independant of the 'true' **C**.
- As $N \to T$ ($q \to 1$) R^2_{in} goes to zero and R^2_{out} diverges!

Bayesian approach

- Ad-hoc cleaning methods exist but is there an optimal one?
- The Markowitz problem needs the expectation value of **C**

$$\langle r_i r_j \rangle = \langle C_{ij} \rangle$$

- The empirical covariance matrix is the maximum-likelihood estimator of **C** but not its expectation value.
- To define the expectation value, we need a Bayesian framework.
- What is the probablity of a 'true **C**' given what we have observed?

$$P(\mathbf{C}|\{r_i^t\}) = \frac{P(\{r_i^t\}|\mathbf{C})P_0(\mathbf{C})}{P(\{r_i^t\})}$$

- The optimal cleaning will depend on the choice of $P_0(\mathbf{C})$.
- Did we gain anything?

Posterior distribution as a matrix model

- The measurement process.

$$P(\{r_i^t\}|\mathbf{C}) = \frac{(\det \mathbf{C})^{-\frac{T}{2}}}{(2\pi)^{\frac{NT}{2}}} \exp\left(-\frac{1}{2}\sum_{ijt} C_{ij}^{-1} r_i^t r_j^t\right)$$

$$\propto \exp\left(-\frac{T}{2}\mathrm{Tr}\left\{\mathbf{E}\mathbf{C}^{-1} + \log \mathbf{C}\right\}\right)$$

- We will assume a rotationally invariant prior of the form

$$P_0(\mathbf{C}) \propto \exp\left\{-\frac{N}{a}\mathrm{Tr}\, V_0(\mathbf{C})\right\}$$

where $\langle \mathbf{C}\rangle_0 = \mathbb{I}$ and a governs the width the distribution.

- Posterior distribution $\Longleftrightarrow$ Matrix model:

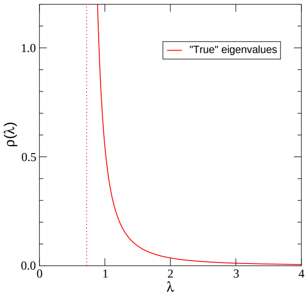
$$\widehat{\mathbf{C}} = \frac{\int \mathcal{D}\mathbf{C}\mathbf{C} \exp\{-N\mathrm{Tr}\, V_{\mathbf{E}}(\mathbf{C})\}}{\int \mathcal{D}\mathbf{C} \exp\{-N\mathrm{Tr}\, V_{\mathbf{E}}(\mathbf{C})\}}.$$

where $V_{\mathbf{E}}(\mathbf{C})$ is our Bayes potential function ($q = N/T$):

$$V_{\mathbf{E}}(\mathbf{C}) = \frac{1}{2q}\log \mathbf{C} + \frac{1}{2q}\mathbf{E}\mathbf{C}^{-1} + \frac{1}{a}V_0(\mathbf{C})$$

The prior is in the data!

- Assume rotational invariance
- Use the self-averaging properties of RMT ⇒ 1 sample gives the whole distribution.
- Need to inverse the M-P formula:
  - parametrically
  - non-parametrically (?)
- We will come back to this later (numerical method).



Some simple priors for analytical computation

- A Wigner matrix centered at the Identity

$$P_0(\mathbf{C}) \propto \exp\left(-\frac{N}{4a}\sum_{i,j}(\mathbf{C}_{i,j}-\delta_{i,j})^2\right)$$

whose potential function is

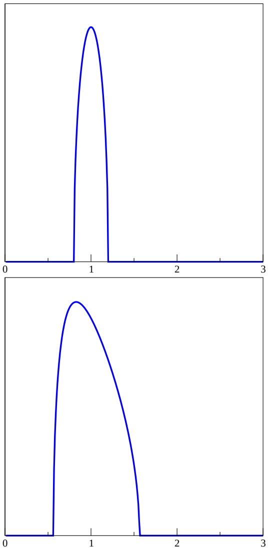
$$V_0(\mathbf{C}) = \frac{1}{4}\left(\mathbf{C}^2 - 2\mathbf{C} + \mathbb{I}\right)$$

- A Wishart matrix

$$P_0(\mathbf{C}) \propto \det(\mathbf{C})^{\frac{N(a^{-1}-1)-1}{2}} \exp\left(-\frac{N}{2a}\mathrm{Tr}\mathbf{C}\right)$$

whose potential function is

$$V_0(\mathbf{C}) = \frac{1}{2}\left(\mathbf{C} + (1-a)\log\mathbf{C}\right)$$



Inverse-Wishart prior

- **C** is an Inverse-Wishart matrix if $\mathbf{C} = \mathbf{C}_W^{-1}$ where $\mathbf{C}_W$ is a Wishart matrix
- The simplest prior for computations. It has the same form as the 'measurement process':

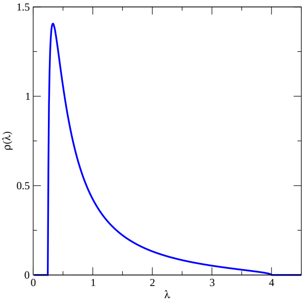
$$P_0(\mathbf{C}) \propto \exp\left(-\frac{N}{a}\mathrm{Tr}\left\{\mathbf{C}^{-1} + (a+1)\log\mathbf{C}\right\}\right)$$

- The eigenvalue density has a reasonable form:

$$\rho(\lambda) = \frac{\sqrt{2(a+1)\lambda - \lambda^2 - 1}}{a\pi\lambda^2}$$

- With this prior, linear shrinkage is optimal

$$\boxed{\hat{\mathbf{C}} = (1-\alpha)\mathbf{E} + \alpha\mathbb{I}} \quad \text{with} \quad \alpha = \frac{2q}{2q+a}$$



Inverse-Wishart prior (II)

- The Inverse-Wishart distribution is the conjugate prior of the Multivariate Gaussian distribution known by statisticians (e.g. [Haff, 1980]).
- The linear Shrinkage was popularized by [Ledoit and Wolf, 2004] where they found a nice way to estimate the Shrinkage parameter α from the data.
- As far as we know, nobody ever considered this prior as a '*true*' distribution of eigenvalues.
- Does the eigenvalues spectrum make sense?
  - Does the α parameter of Ledoit-Wolf correspond to a reasonable *a* for the prior?

Matrix saddle-point

- Our aim: evaluate $\langle \mathbf{C} \rangle_{\mathbb{P}(\mathbf{C}|\mathbf{E})}$
- Explicit solution for the Inverse-Wishart prior, but not for other priors

⇒ First method: use a matrix Saddle-point to have a suitable point at which one can start a perturbation theory in the number of loops.

- The saddle-point $\mathbf{C}_0$ is such that

$$V'_{\mathbf{E}}(\mathbf{C}_0) = \frac{1}{2q}\mathbf{C}_0^{-1} - \frac{1}{2q}\mathbf{E}\mathbf{C}_0^{-2} + \frac{1}{a}V'_0(\mathbf{C}_0) = 0.$$

- Applications of the saddle-point (let $\alpha = q/a$)
  - For the Wigner prior:

$$\boxed{\alpha\mathbf{C}_0 - \alpha\mathbb{I} + \mathbf{C}_0^{-1} + \mathbf{E}\mathbf{C}_0^{-2} = 0}$$

  - For the Wishart prior :

$$\boxed{(1-\alpha+q)\mathbf{C}_0^{-1} - \mathbf{E}\mathbf{C}_0^{-2} + \mathbb{I} = 0}$$

Matrix saddle-point (II)

$$\frac{1}{2q}\mathbf{C}_0^{-1} - \frac{1}{2q}\mathbf{E}\mathbf{C}_0^{-2} + \frac{1}{a}V_0'(\mathbf{C}_0) = 0.$$

- Our matrix saddle point $\mathbf{C}_0$ is not exact.
- The are still fluctuations coming form the measurement process (q) and from the prior distribution (a).
- It is exact in the limit $q \to 0$ and $a \to 0$ with fixed $\alpha = q/a$.
- $\mathbf{C}_0$ also contains higher order terms in q, we denote $\mathbf{C}_{00} = \lim_{q\to 0} \mathbf{C}_0$
- $\mathbf{C}_0$ and $\mathbf{E}$ commute.
- At this order, the Baysian estimator is a (non-linear) shrinkage function applied to the eigenvalues of $\mathbf{E}$.
- Eigenvectors of $\mathbf{E}$ are left unchanged.

Perturbation theory on C

- Let $\mathbf{C}_0$ the solution of the saddle-point equation. By a simple change of variable

$$\mathbf{C} = \mathbf{C}_0^{1/2}(\mathbb{I} + X)\mathbf{C}_0^{1/2}$$

- Our Bayes potential function becomes

$$\begin{aligned} V_{\mathbf{E}}(\mathbf{C}) &= \frac{1}{2q} \log\left(\mathbf{C}_0^{1/2}(\mathbb{I}+X)\mathbf{C}_0^{1/2}\right) + \frac{1}{2q}\mathbf{E}\left(\mathbf{C}_0^{1/2}(\mathbb{I}+X)\mathbf{C}_0^{1/2}\right)^{-1} \\ &+ \frac{1}{a} V_0(\mathbf{C}_0^{1/2}(\mathbb{I}+X)\mathbf{C}_0^{1/2}) \end{aligned}$$

- Ignoring constants and cyclical permutations ($\alpha = q/a$)

$$V_{\mathbf{E}}(\mathbf{C}) = \frac{1}{2q}\left[\log(\mathbb{I}+X) + \mathbf{E}\mathbf{C}_0^{-1}(\mathbb{I}+X) + 2\alpha V_0(\mathbf{C}_0^{1/2}(\mathbb{I}+X)\mathbf{C}_0^{1/2})\right]$$

Perturbation theory on **C** (II)

- Let $V_X(X) = V_0(\mathbf{C}_0^{1/2}(\mathbb{I}+X)\mathbf{C}_0^{1/2})$, we can now proceed to a Taylor series expansion

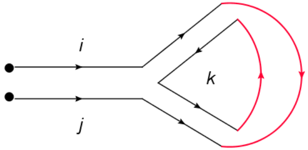
$$V_{\mathbf{E}}(\mathbf{X}) = \frac{1}{2q} \left(\sum_{k=2}^{\infty} (-1)^k \left[\mathbf{E}\mathbf{C}_0^{-1} - \frac{1}{k} \right] \mathbf{X}^k + 2\alpha \left[\frac{1}{2} \sum_{i,j,k,l} X_{i,j} X_{k,l} \frac{\partial^2 V_X}{\partial X_{i,j} \partial X_{k,l}} \Big|_{X=0} + \mathcal{O}(\mathbf{X}^3) \right] \right)$$

- As the constant and linear terms vanish, the first contribution (quadratic term) leads directly to the propagator D in order to use Wick's theorem
- In the large N limit, it is known from ['t Hooft, 1974] that the only diagrams which survive are planar

$\Rightarrow$ If we truncate the loop expansion to a certain level k, we can compute our estimator to order q^k.

Systematic approach by diagrammatic expansion

- For the first order correction term in q, there is only one planar diagram given by $\left\langle \mathbf{X}\mathrm{Tr}\mathcal{M}^{(3)}\mathbf{X}^3 \right\rangle$



Explicit expression for this contribution (in the diagonal basis)

$$\widehat{\mathbf{C}}_{i,i} = (\mathbf{C}_0)_{i,i} + (\mathbf{C}_0^{1/2})_{i,i} \left\langle \mathbf{X}\mathrm{Tr}\mathcal{M}^3\mathbf{X}^3 \right\rangle_{i,i} (\mathbf{C}_0^{1/2})_{i,i} + \mathcal{O}(q^2)$$

First order correction

Applications:

- For the Wigner prior:

$$\widehat{\mathbf{C}}_{i,i} = (\mathbf{C}_0)_{i,i} + q \left[\frac{(\mathbf{C}_0)_{i,i}}{3\alpha(\mathbf{C}_0)_{i,i}^2 - 2\alpha(\mathbf{C}_0)_{i,i} + 1} \right]$$

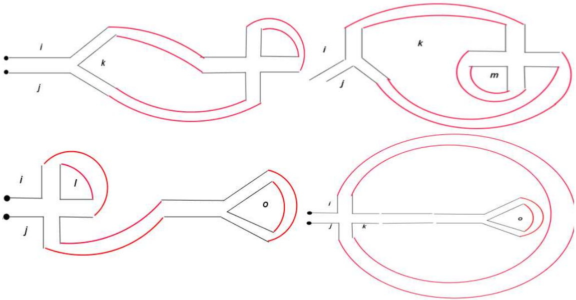
$$\times \left(1 - \frac{1}{N} \sum_k \left[\frac{\alpha(\mathbf{C}_0)_{i,i}(\mathbf{C}_0)_{j,j} - \alpha(\mathbf{C}_0)_{i,i}((\mathbf{C}_0)_{i,i} - 1) - 1}{\alpha\left((\mathbf{C}_0)_{i,i}((\mathbf{C}_0)_{i,i} - 1) + (\mathbf{C}_0)_{k,k}((\mathbf{C}_0)_{k,k} - 1) + (\mathbf{C}_0)_{i,i}(\mathbf{C}_0)_{k,k}\right) + 1} \right] \right)$$

- For the Wishart prior:

$$\widehat{\mathbf{C}}_{i,i} = (\mathbf{C}_{00})_{i,i} + q \frac{(\mathbf{C}_{00})_{i,i}}{N} \frac{\alpha(\mathbf{C}_{00})_{i,i} + 1 - \alpha}{2\alpha(\mathbf{C}_{00})_{i,i} + 1 - \alpha} \sum_k \frac{1}{\alpha(\mathbf{C}_{00})_{i,i} + \alpha(\mathbf{C}_{00})_{k,k} + 1 - \alpha}$$

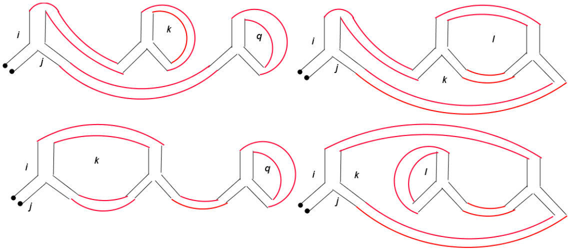
Second order correction

- First order correction via Feynman diagrammatic expansion leads to explicit expressions...
- But the second order correction leads to ten different planar diagrams and far more tedious computations!
- Contribution for $\left\langle X \mathrm{Tr} \mathcal{M}^{(3)} X^3 \mathcal{M}^{(4)} X^4 \right\rangle$

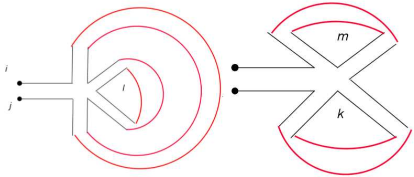


Second order correction (cont.)

- Contribution for $\left\langle X \mathrm{Tr} \mathcal{M}^{(3)} X^3 \mathcal{M}^{(3)} X^3 \mathcal{M}^{(3)} X^3 \right\rangle$



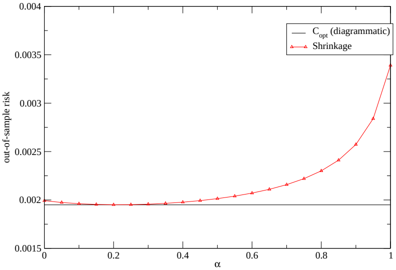
- Contribution for $\left\langle X \mathrm{Tr} \mathcal{M}^{(5)} X^5 \right\rangle$



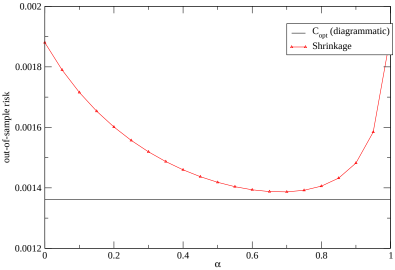
Out-of-sample risk for the one-loop solution

Test of the out-of-sample risk on simulated data for the one-loop approximation with an arbitrary q.

- Wigner with $N = 500$, $\sigma = 0.3$ and $q = 0.5$



- Wishart with $N = 500$, $q_0 = 0.5$ and $q = 0.5$



Eigenvalues saddle-point: the HCIZ integral problem

Alternative method: perform a saddle-point on eigenvalues to find the exact value of $\mathbf{C}_{opt}$. Suppose that $\mathbf{C} = O\Lambda O^T$, our problem is of the form

$$\mathbb{P}(\mathbf{C}|\mathbf{E}) \propto \int d\lambda_1 \ldots d\lambda_N \exp\left\{\log I(\mathbf{E},\Lambda) - N\left[\frac{1}{2q}\sum_{i=1}^{N}[\log(\lambda_i) + 2\alpha V_0(\lambda_i)] - \frac{1}{N}\sum_{i<j}^{N}\log|\lambda_i - \lambda_j|\right]\right\}$$

with $I(\mathbf{E},\Lambda)$ the well-known Harish-Chandra-Itzykson-Zuber integral

$$I(\mathbf{E},\Lambda) = \int \mathcal{D}O \exp\left\{-\frac{N}{2q}\text{Tr}\, O^T\mathbf{E}O\Lambda^{-1}\right\}$$

$\Rightarrow$ Main difficulty: the evaluation of the Orthogonal version of the HCIZ integral in the large N limit: $I \sim \exp -N^2 F(\mathbf{E},\Lambda)$.

- Some formulas are known for the large N limit of HCIZ but we haven't found a way to use them in our problem.

In order to make computation, we have to make a brutal hypothesis!

Special case: $\mathbf{E} = e\mathbb{I}$

- Denote by $\lambda_i, i \in \{1, ..., N\}$ (resp. $e_i, i \in \{1, ..., N\}$) the i-th eigenvalue of **C** (resp. of **E**), we suppose that

$$\boxed{\mathbf{E} = e \times \mathbb{I}}$$

that is to say $\lambda_i = F(e_i)$, where F is a function that depends of the prior.

- In this case

$$\mathbb{P}(\mathbf{C}|\mathbf{E}) \propto \int d\lambda_1 \dots d\lambda_N \exp\left\{-N\left[\frac{1}{2q}\sum_{i=1}^{N}[\log(\lambda_i) + \frac{e}{\lambda_i} + 2\alpha V_0(\lambda_i)] - \frac{1}{N}\sum_{i<j}^{N}\log|\lambda_i - \lambda_j|\right]\right\}$$

- Following the work of [BIPZ, 1978], this problem can be solved by using the Stieltjes transform. In the Orthogonal case, when $z^r V'(z)$ is a polynomial, we have

$$G(z) = V_E'(z) \pm \sqrt{V_E'(z)^2 - 2P(z)}.$$

with $z^r P(z)$ a polynomial with unknown coefficients.

One-cut assumption

- We consider $V_{\mathbf{E}}(\mathbf{C})$ a convex function such the density of the eigenvalues of $\mathbf{C}$ under the posterior distribution is given by an unique compact support $\Rightarrow$ one-cut assumption
- Under this one-cut assumption, the Stieltjes transform of $\mathbf{C}$ under the posterior distribution is now

$$G(z) = V_E'(z) \pm Q(z)\sqrt{z^2 - 2az + b^2}.$$

with $z^r Q(z)$ still a polynomial in z. To find a, b and the coefficients of Q, we have:

  - if $z^r V'(z)$ is a polynomial of order k, then $z^r Q(z)$ is a polynomial of order $k-1$;
  - $\mathbf{C}$ is a positive definite matrix: $G(z)$ is regular in 0;
  - $G(z)$ is the solution of the Riemann-Hilbert problem. In particular, for $|z| \to \infty$,

$$G(z) \sim \frac{1}{z} + o(1/z^2)$$

and G(z) is analytical outside its branch cut.

Application to the Wishart case

In the Wishart case

$$V_E(z) = \frac{1}{2q}\left[\log(z) + \frac{e}{z} + \alpha(z - (1 - q_0)\log(z))\right].$$

and

$$G(z) = V_E'(z) \pm Q(z)\sqrt{z^2 - 2az + b^2} = V_E'(z) \pm \frac{cz + d}{z^2}\sqrt{z^2 - 2az + b^2}.$$

We find, with $\gamma = e/(2q)$,

- when $z \to 0$

$$d = \frac{\gamma}{b}$$

$$a = \frac{b^2}{\gamma}\left[\frac{\beta}{2q} + cb\right]$$

- when $z \to \infty$

$$c = \frac{\alpha}{2q}$$

$$\alpha^2 b^4 + \alpha\beta b^3 - e(\alpha - 1 + q)b - e^2 = 0.$$

Link with our Bayes estimator

- We can retrieve our Bayes estimator from the Stieltjes transform: in the large z limit,

$$G(z) \sim \frac{1}{z} + \frac{\langle \mathbf{C} \rangle_{\mathbb{P}(\mathbf{C}|\mathbf{E})}}{z^2} + \mathcal{O}\left(\frac{1}{z^3}\right)$$

- Application: for the Wishart prior:

$$\langle \mathbf{C} \rangle_{\mathbb{P}(\mathbf{C}|\mathbf{E})} = \frac{1}{2q}\left[e\left(\frac{a}{b} - 1\right) - \frac{\alpha}{2}(b^2 - a^2)\right]$$

$\Rightarrow$ **Generalize the previous approach as it is exact at all orders**:

- Perturbation theory $b = b_0 + qb_1 + q^2 b_2 + \mathcal{O}(q^3)$
- At first order, we find

$$\langle \mathbf{C} \rangle_{\mathbb{P}(\mathbf{C}|\mathbf{E})} = b_0 + q\frac{b_0(\alpha b_0 + \beta_0)}{(2\alpha b_0 + \beta_0)^2} + \mathcal{O}(q^2)$$

with $\beta_0 = 1 - \alpha$ and b_0 the solution of our Saddle-point equation for the Wishart prior with $q \to 0$ such that α finite

$$\alpha b_0 + \beta_0 b_0 - e = 0$$

$\rightarrow$ Same result than the Feynman diagrammatic expansion presented before for $\mathbf{E} \propto \mathbb{I}$

A Monte-Carlo based method

- We propose a numerical method to evaluate $\langle \mathbf{C} \rangle_{\mathbb{P}(\mathbf{C}|\mathbf{E})}$ using only a given prior matrix $\mathbf{C}$ and the empirical covariance matrix $\mathbf{E}$.
- Due to our rotational invariance hypothesis, we want to find $\mathbf{E}$ such that it minimizes the quadratic distance with $\mathbf{C}$ without modifying the eigenvectors
- By eigendecomposition $\mathbf{E} = U \Lambda U^{-1}$
- Our optimization problem is

$$\min_{\Lambda_{k,k}} \sum_{i,j} \left(\mathbf{C}_{i,j} - U_{i,k} \Lambda_{k,k} U_{j,k} \right)^2 .$$

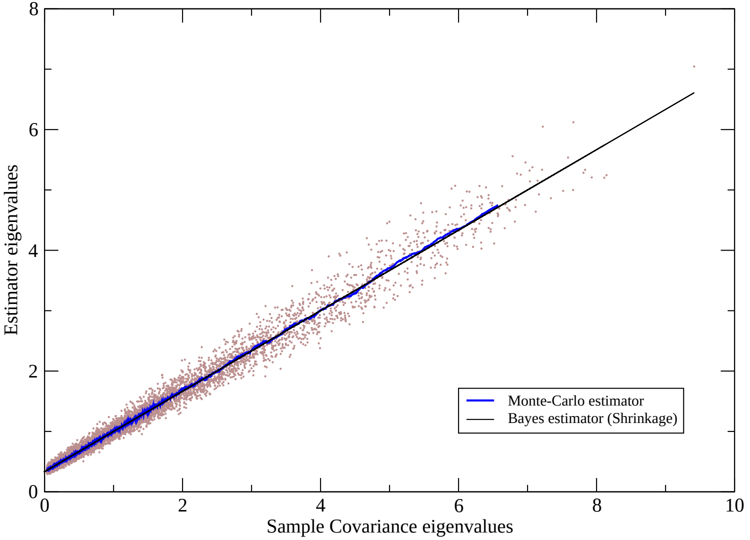
- The solution is

$$\boxed{\hat{\Lambda}_{k,k} = \sum_{i,j} U_{i,k} \mathbf{C}_{i,j} U_{j,k}}$$

- To get our Bayes estimator, we have

$$\langle \mathbf{C} \rangle_{\mathbb{P}(\mathbf{C}|\mathbf{E})} = \langle \hat{\Lambda} \rangle$$

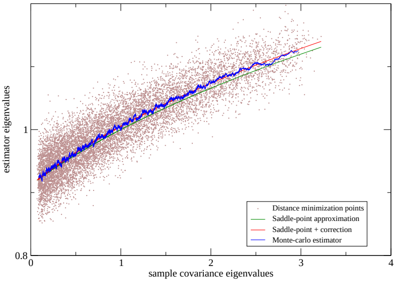
Test of Monte-Carlo method on Inverse-Wishart



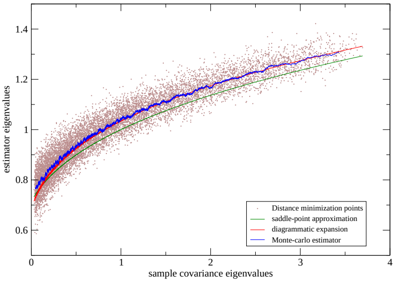
Optimality: Wigner Prior

Comparison of the optimality of our solution against the Monte-Carlo estimator (10000 points).

- Wigner with $N = 100$, $\sigma = 0.2$ and $q = 0.5$



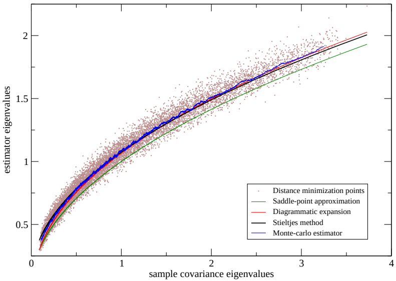
- Wigner with $N = 100$, $\sigma = 0.35$ and $q = 0.5$



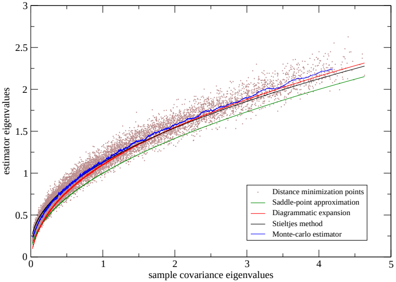
Optimality: Wishart Prior

Comparison of the optimality of our solution against the Monte-Carlo estimator (10000 points).

- Wishart with $N = 100$, $q_0 = 0.3$ and $q = 0.3$



- Wishart with $N = 100$, $q_0 = 0.5$ and $q = 0.5$



A fully numerical procedure

- Measure the sample covariance matrix on your data.
- Choose a parametric form for the 'true' distribution of eigenvalue for which you can compute the Resolvent $G(z)$.
- Fit the parameters to the SCM using Marčenko and Pastur

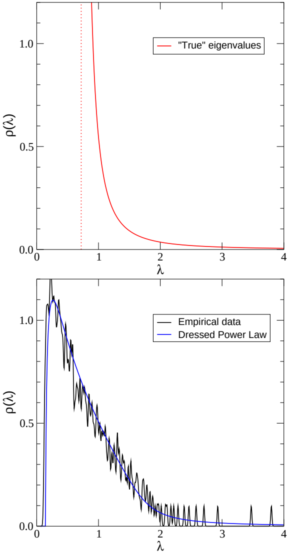
$$zG_{\mathbf{E}}(z) = ZG_{\mathbf{C}}(Z)$$

with

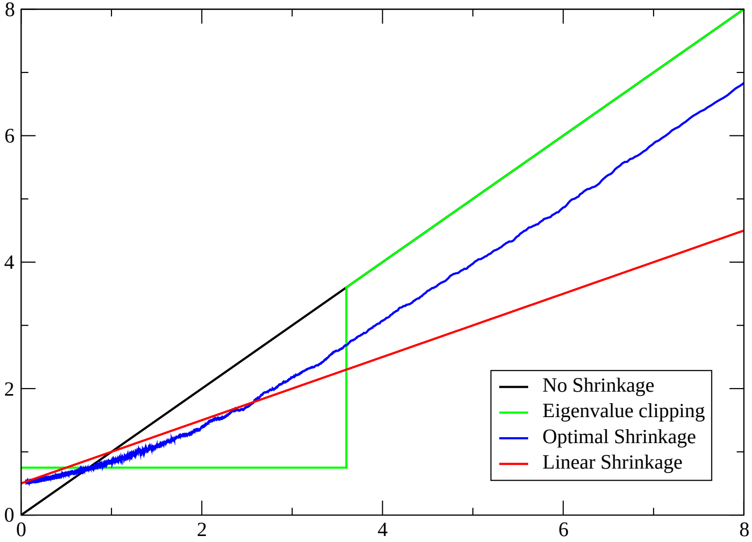
$$Z = \frac{z}{1 + q(zG_{\mathbf{E}}(z) - 1)}.$$

- Using the Monte Carlo procedure, compute the optimal shrinkage function.

$$\hat{\Lambda}_{k,k} = \sum_{i,j} U_{i,k} \mathbf{C}_{i,j} U_{j,k}$$



Numerics on our power-law prior



Summary & Conclusions

- The out-of-sample risk quadratic optimization problem can be rewritten in a Bayesian framework
- RMT allows us to characterize several prior on the true covariance matrix **C**.
- The computation of the Bayes estimator is reduced to the computation of an orthogonal version of a matrix model with an external field.
- One-loop perturbation theory gives satisfactory results for simple priors.
- We also present a simple numerical procedure that can be used for any prior.
- Open problems:
  - What kind of performance can we obtain on real data with those solutions?
  - Can we find a formulation of the large N limit of HCIZ that will allow us to solve the eigenvalue saddle point?
  - Extensions:
    - non-Gaussian data (e.g. Student Multivariate).
    - non-rotationnaly invariant prior (e.g. Market mode: permutation invariance)

References

Jean-Philippe Bouchaud, Marc Potters (2009)
Financial Applications of Random Matrix Theory: a short review
The Oxford Handbook of Random Matrix Theory.

Edouard Brézin, Claude Itzykson, Giorgio Parisi, and Jean-Bernard Zuber (1978)
Planar diagrams
Communications in Mathematical Physics

Gerard 't Hooft (1974)
A planar diagram theory for strong interactions
Nuclear Physics B

Vladimir Marcenko, Leonid Pastur (1967)
Distribution of eigenvalues for some sets of random matrices
Matematicheskii Sbornik.

L.R. Haff (1980)
Empirical Bayes estimation of the multivariate normal covariance matrix
The Annals of Statistics.

Olivier Ledoit, Michael Wolf (2004)
A Well-Conditioned Estimator for Large-Dimensional Covariance Matrices
Journal of Multivariate Analysis.