

9 Supplementary Material: A Primer on Nonlinear Shrinkage

Nonlinear shrinkage estimation of the unconditional covariance matrix is a burgeoning field of probability and statistics which may not be very accessible to applied researchers in economics and finance. This supplementary material provides a self-contained introduction. It is intended to be descriptive, qualitative, and as non-technical as the nature of the subject matter will allow. It is not intended as a substitute for the rigorous treatment provided in [Ledoit and Wolf \(2017a\)](#).

The exposition here is couched in terms of the covariance matrix, but in the DCC context the described estimator should be applied to the *devolatilized* residuals. The resulting estimate should then be renormalized as per Section 3.5 in order to generate a proper correlation matrix.

Σ denotes the population covariance matrix. Y_T denotes a stationary data set of dimension $T \times N$ with covariance matrix Σ . We assume mean zero for simplicity.

9.1 Importance of the Eigenvalues for Portfolio Selection

Following [Markowitz \(1952\)](#), if μ denotes a vector of expected returns, then the weights of the tangency portfolio are

$$w^{\text{TANGENCY}} = \text{scalar} \times \Sigma^{-1} \mu . \quad (9.1)$$

Inverting a matrix is not a particularly intuitive operation, and when an experienced practitioner like [Michaud \(1989\)](#) warns that it leads to “error maximization”, it is hard to see what is going wrong or how to fix it.

Fortunately, the covariance matrix is not just *any* matrix, it is a *symmetric* matrix. The covariance of the return on Intel shares with Nike shares is the same as Nike with Intel by definition. Symmetric matrices enjoy a very special property: they always admit a *spectral decomposition*. This decomposition is given by

$$\Sigma =: V \begin{bmatrix} \tau_1 & & & \\ & \tau_2 & & 0 \\ & & \ddots & \\ & 0 & & \ddots \\ & & & & \tau_N \end{bmatrix} V' , \quad (9.2)$$

where $\tau := (\tau_1, \dots, \tau_N)$ are the population eigenvalues and V is a rotation matrix, meaning that $V' = V^{-1}$. The i th column of V is the population eigenvector v_i .

The best way to interpret this decomposition is to look at the dimension $N = 2$. Figure 1 gives a graphical illustration.

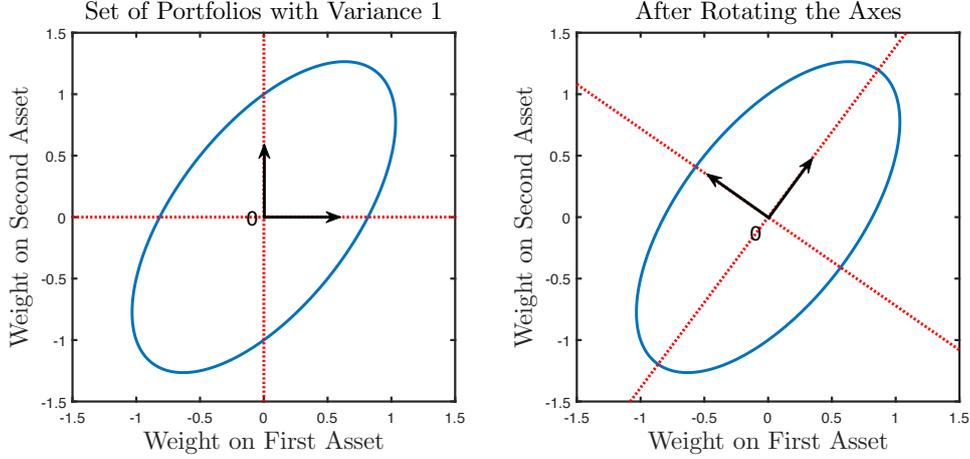


Figure 1: Decomposition into eigenvalues and eigenvectors in dimension $N = 2$.

On the left panel, the ellipsis represents all the portfolios that have the same variance (which we can take as normalized to one). The fact that it bulges out into the northeast and southwest quadrants is due to nonzero covariance between the two asset returns. What the eigenvectors do is define a change of basis, a rotation of the axes, so that (when viewed from this new angle) all quadrants look the same. This is shown on the right panel. One eigenvector corresponds to the axis across which the ellipsis is narrowest, and another corresponds to the axis across which the ellipsis is widest. The ellipsis is not a perfect circle because the two eigenvalues are not equal to each other.

Economically, what this rotation does is to repackage the original menu of assets into N funds whose returns are all mutually uncorrelated. The weights of each fund are given by the corresponding eigenvector; and the corresponding eigenvalue is the variance of the return on the fund. Since they represent fund return variances, all eigenvalues must be non-negative, meaning that the covariance matrix is *positive semi-definite*.

The N funds span the same space of investment opportunities as the original assets, therefore we can rewrite Equation (9.1) as

$$w^{\text{TANGENCY}} = \text{scalar} \times \sum_{i=1}^N \frac{v_i' \mu}{\tau_i} v_i . \quad (9.3)$$

Equation (9.3) demonstrates that the tangency portfolio is best viewed not as a combination of the N original assets, but as a combination of the N uncorrelated eigenvector funds. The capital

assigned to each fund is proportional to its expected return and inversely proportional to its variance, which makes economic sense.

It is easy to justify that the spectral decomposition is important for portfolio selection. Consider the hypothetical covariance matrix of monthly stock returns in Table 1.

	Apple	Boeing	Disney	IBM
Apple	0.2694	0.5714	0.2900	0.3080
Boeing	0.5714	1.3910	0.6674	0.6964
Disney	0.2900	0.6674	0.3275	0.3433
IBM	0.3080	0.6964	0.3433	0.4822

Table 1: Hypothetical covariance matrix between four US stocks.

To the naked eye, it looks fine. However, its eigenvalues are $(0, 0.0299, 0.1072, 2.3329)$. Even to the naked eye, the first eigenvalue looks wrong. The tangency portfolio does not exist when an eigenvalue is equal to zero. This is why extracting eigenvalues and eigenvectors is called the *spectral* decomposition: It enables us to penetrate right through the outer appearance of the matrix into its inner structure.

In practice, we do not know the true covariance matrix Σ , therefore we must use some estimator of it. It is known that the sample covariance matrix $S_T := Y_T'Y_T/T$ is a consistent estimator of Σ when the sample size T goes to infinity *while the dimension N remains fixed* (an often overlooked yet crucial assumption, to which we will return later). Mirroring Equation (9.2), define the spectral composition of S_T as

$$S_T =: U_T \begin{bmatrix} \lambda_{1,T} & & & & \\ & \lambda_{2,T} & & 0 & \\ & & \ddots & & \\ & 0 & & \ddots & \\ & & & & \lambda_{N,T} \end{bmatrix} U_T', \quad (9.4)$$

where $\lambda_T := (\lambda_{1,T}, \dots, \lambda_{N,T})$ are the sample eigenvalues and U_T is a rotation matrix ($U_T' = U_T^{-1}$) whose i th column is the sample eigenvector $u_{i,T}$. An implementable version of Equation (9.3) is

$$w^{\text{TANGENCY}} = \text{scalar} \times \sum_{i=1}^N \frac{u_{i,T}' \mu}{\lambda_{i,T}} u_{i,T}. \quad (9.5)$$

This formulation leads to a fundamental insight: In the denominator, we have $\lambda_{i,T}$, which is the *in-sample* variance of the i th eigenvector fund $u_{i,T}$, whereas for investment purposes we need its *out-of-sample* variance $u'_{i,T}\Sigma u_{i,T}$ instead. The whole point of the procedure advocated in this paper is to replace the former with (a consistent estimate of) the latter. Investment decisions are always evaluated out of sample.

One little-known mathematical fact about the eigenvalues is that they are the *most dispersed* diagonal elements that can be obtained through rotation; see [Ledoit and Wolf \(2004, Section 2.3\)](#). Given that the group of rotations has dimensionality of order N^2 , the potential for overfitting is tremendous when N is large. Overfitting causes excess dispersion: The smallest sample eigenvalues are too small, leading to over-investment, and the largest sample eigenvalues too large, leading to under-investment. The overall result is mal-investment. This insight goes a long way towards explaining the observation by [Michaud \(1989\)](#) about “error maximization”. However, to fix it requires a detour through multivariate statistics.

9.2 Importance of the Eigenvalues for Covariance Matrix Estimation

If we are going to look for estimators that improve upon the sample covariance matrix, the first task is to decide *where* to look. We need to specify a class of eligible estimators, and search within this class for one that beats the sample covariance matrix. In mathematics, a standard way to approach this kind of problem is to say that we want estimators that have certain appealing properties. One such property initially championed by [Stein \(1975\)](#) and subsequently adopted by many other authors is called *rotation equivariance*. A covariance matrix estimator $\widehat{\Sigma}_T(Y_T)$ is said to be rotation-equivariant if and only if for any N -dimensional rotation matrix W ,

$$\widehat{\Sigma}_T(Y_TW) = W' \widehat{\Sigma}_T(Y_T) W . \tag{9.6}$$

That is, the estimate based on the rotated data equals the rotation of the estimate based on the original data. Absent any *a priori* knowledge about the orientation of the true eigenvectors, it is natural to consider only covariance matrix estimators that are rotation-equivariant.

It can be proven that the class of rotation-equivariant estimators that are a function of the

sample covariance matrix is the class of estimators of the form

$$\widehat{\Sigma}_{\Psi_T} := U_T \begin{bmatrix} \psi_{1,T} & & & & \\ & \psi_{2,T} & & 0 & \\ & & \ddots & & \\ & & & 0 & \\ & & & & \ddots \\ & & & & & \psi_{N,T} \end{bmatrix} U_T', \quad (9.7)$$

where $\Psi_T := (\psi_{1,T}, \dots, \psi_{N,T})$ can be any vector in $[0, +\infty)^N$; for example, see [Perlman \(2007, Section 5.4\)](#). Thus, we preserve the sample eigenvectors, but are free to modify the sample eigenvalues in any way needed to improve upon the sample covariance matrix. Given [Section 9.1](#), the basic idea will be to set $\psi_{i,T}$ equal to $u_{i,T}' \Sigma u_{i,T}$, or if it is unavailable (the more likely scenario, given that it depends on the population covariance matrix, which is unobservable), a consistent estimator thereof.

In summary, the key intuition is that we have to preserve the sample eigenvectors because we lack *a priori* information about the orientation of the true eigenvectors, and the goal is to modify the sample eigenvalues so we can beat the sample covariance matrix.

9.3 General Asymptotics

As mentioned before, the sample covariance matrix S_T is a consistent estimator of the population covariance matrix Σ when the sample size T goes to infinity while the dimension N remains fixed. This is strange: why is T allowed to move but not N ? When we have five years of daily data ($T = 1250$) on the components of the Russell 1000 stock index ($N = 1000$), it is easy to believe that T goes to infinity, as 1250 is a large number by any measure in statistics, but who is to say that N is finite? Shouldn't numbers that go to infinity be much bigger than those that are assumed to remain finite?

The answer is to simply relax the constraining assumption that N is fixed and instead allow the dimension to move along with the sample size: $N := N(T)$. This is called *general asymptotics*, large-dimensional asymptotics, or Kolmogorov asymptotics. Notation-wise, this kind of asymptotics requires appending the subscript T to the population covariance matrix, and also its eigenvalues and eigenvectors, a convention that we will uphold from here onwards.

Given that the number of eigenvalues N goes to infinity, it is no longer possible to make statements about individual eigenvalues. This is why it is necessary to introduce what is

known as *spectral distributions*. The population and sample spectral distributions are defined respectively as

$$\forall x \in \mathbb{R} \quad H_T(x) := \frac{1}{N} \sum_{i=1}^N \mathbf{1}_{\{\tau_{i,T} \leq x\}} \quad \text{and} \quad (9.8)$$

$$\forall x \in \mathbb{R} \quad F_T(x) := \frac{1}{N} \sum_{i=1}^N \mathbf{1}_{\{\lambda_{i,T} \leq x\}} , \quad (9.9)$$

where $\mathbf{1}$ denotes the indicator function. The spectral distribution can be interpreted as a cross-sectional cumulative distribution function (c.d.f.): it is a nondecreasing function having values between zero and one that returns the proportion of eigenvalues lower than its argument.

Two standard assumptions under general asymptotics are (i) that the population spectral distribution converges to a well-defined limit H called the *limiting spectral distribution* and (ii) that the ratio N/T converges to a finite limit c called the *concentration*:

$$H_T(x) \longrightarrow H(x) \quad \text{at all points of continuity of } H \quad (9.10)$$

$$\frac{N}{T} \longrightarrow c < +\infty . \quad (9.11)$$

Along with other technical assumptions that can vary from author to author, these two assumptions imply the fundamental result of general asymptotics, which is that the sample spectral distribution converges to a *nonrandom* limit F called the limiting spectral distribution:

$$F_T(x) \xrightarrow{\text{a.s.}} F(x) \quad \text{at all points of continuity of } F. \quad (9.12)$$

Remark 9.1. The fact that the matrix is random but its eigenvalues are not is a remarkable mathematical phenomenon first discovered by [Wigner \(1955\)](#) while investigating the properties of the wave functions of complicated quantum mechanical systems; see [Figure 2](#) for an illustration of this influential result. ■

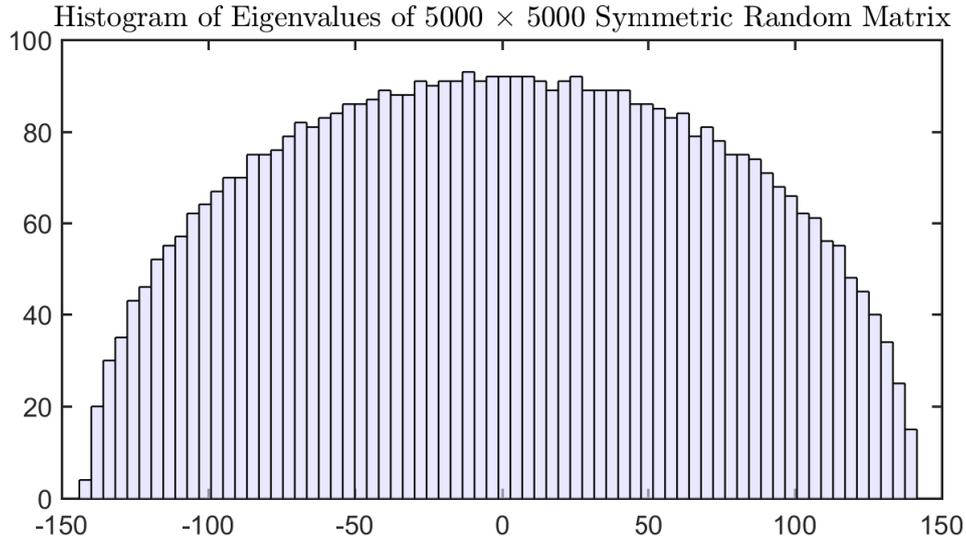


Figure 2: The eigenvalues of a large Wigner matrix follow Wigner’s semi-circular law. Wigner matrices are random symmetric matrix with i.i.d. standard normal entries. (They are different from covariance matrices, since they also have negative eigenvalues.) This picture does not represent an average across Monte Carlo simulations: it is just the result of one single draw.

The limiting sample spectral distribution F is the key to knowing where the sample eigenvalues lie. There are a few things we can immediately say about this important object:

(a) F is uniquely determined by H and c ; see [Silverstein and Choi \(1995\)](#)

(b) $F = H \iff c = 0$

(c) $\int_{-\infty}^{+\infty} x dF(x) = \int_{-\infty}^{+\infty} x dH(x)$

(d) $\int_{-\infty}^{+\infty} x^2 dF(x) = \int_{-\infty}^{+\infty} x^2 dH(x) + c \left[\int_{-\infty}^{+\infty} x dH(x) \right]^2$

Statement (b) confirms that finite-dimensional asymptotics are included as a special case of general asymptotics. When N remains fixed and finite, N/T converges to zero as T goes to infinity. In this case, the eigenvalues of the sample covariance matrix are consistent estimators of their population counterparts. This remains true even if N goes to infinity along with T , as long as it grows sufficiently slowly (say in $\log(T)$ or \sqrt{T}). When $c = 0$ or, practically speaking, when N/T is minuscule, the sample covariance matrix works fine.

For five years of history on the Russell 1000, the ratio N/T is equal to 0.8, so it is definitely not minuscule. $c > 0$ is the relevant case for all large covariance matrices, because

when N is large it is very difficult to have a sample size such that the ratio is N/T minuscule. In this case, the sample eigenvalues never get close to their population counterparts, so we enter a qualitatively different regime where improvement over the sample covariance matrix is possible.

Statement (c) means that the cross-sectional average of the sample eigenvalues is in the right place even when $c > 0$: it never needs fixing. However, Statement (d) shows their cross-sectional dispersion around the average is systematically inflated, and excess dispersion increases in c ; see Yin (1986, Equation (4.14)). This confirms formally the intuition developed at the end of Section 9.1. To fix this problem, the filtering applied to sample eigenvalues will have to ‘shrink’ their cross-sectional dispersion towards the center; see Ledoit and Wolf (2004).

9.4 Single Mass Point

One way to get information about the limiting sample spectral distribution is to study what happens in the simplest case, when the population covariance matrix Σ_T is equal to the identity matrix. In this case, a closed-form solution exists: F is differentiable, and its derivative f , called the limiting spectral density, follows the so-called Marčenko-Pastur Law:

$$\forall x \in [a, b] \quad f(x) := \frac{\sqrt{(b-x)(x-a)}}{2\pi cx},$$

where the bounds of the support of f are $a := (1 - \sqrt{c})^2$ and $b := (1 + \sqrt{c})^2$ respectively; see Marčenko and Pastur (1967).

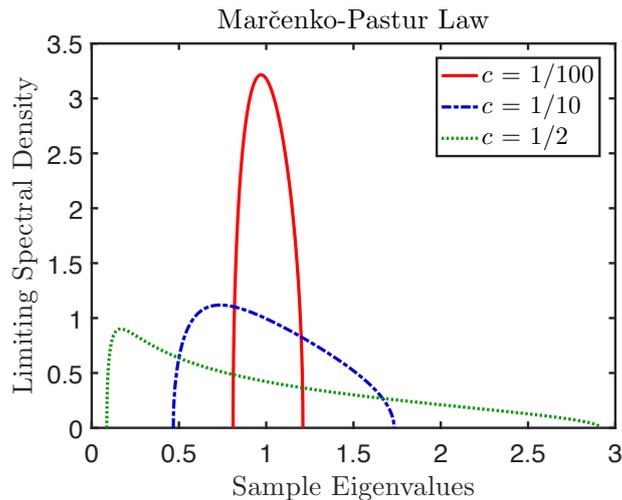


Figure 3: Limiting cross-sectional density of sample eigenvalues for various concentration levels.

Some lessons can be drawn, that reinforce and complement what we have learned from the first two moments:

1. Sample eigenvalues are smudged to the left and the right of the population eigenvalues.
2. The amount of excess spread increases in the concentration ratio c .
3. The smallest sample eigenvalues are too small, the largest ones too large.
4. This systematic bias must be filtered out by *shrinking* the distribution of sample eigenvalues towards the center.
5. The center (cross-sectional average) of the sample eigenvalues distribution is accurate: it matches its population counterpart.
6. The density is right-skewed: there are many small sample eigenvalues and few large ones.
7. It is only in the limit $c \rightarrow 0$ that sample eigenvalues start to conform with standard (fixed-dimension) asymptotics.

The concentration c needs to be *very* close to 0 before concerns about excess dispersion can be safely dismissed. For $c = 1/10$, meaning that we have 10 times more observations than variables, which many people would deem sufficient, some sample eigenvalues are still less than half their population counterparts. Even when $c = 1/100$, out-of-sample portfolio variances can be under/overestimated by 20%, inducing over/under-allocation of risk capital by the same percentage.

9.5 Two Mass Points

These seven observations carry over to the case where a fraction α of the population eigenvalues are equal to one, while the rest are equal to some $\tau > 1$. The behavior of the limiting spectral density is pretty much as one would expect intuitively: There are two clusters formed around each population eigenvalue. The clusters are either close to each other or distant, depending on how far τ is from 1. This finding adds one more qualitative observation to the list:

8. The locations of clusters of sample eigenvalues match the locations of the underlying population eigenvalues.

When one mass point in the distribution of population eigenvalues is heavier than the other, the weights of the clusters of population eigenvalues adjust in proportion, as one would intuitively expect. This finding yields yet one more common-sense observation:

9. The weights of clusters of sample eigenvalues match the multiplicities of the underlying population eigenvalues.

The only non-obvious mathematical phenomenon is a so-called “phase transition” that takes place when two clusters merge with each other to form a single one. It happens either because the underlying population eigenvalues are too close to each other, or the concentration c is too high. This phenomenon is illustrated in Figure 4 for the case $\alpha = 1/2$ and $\tau = 2$.

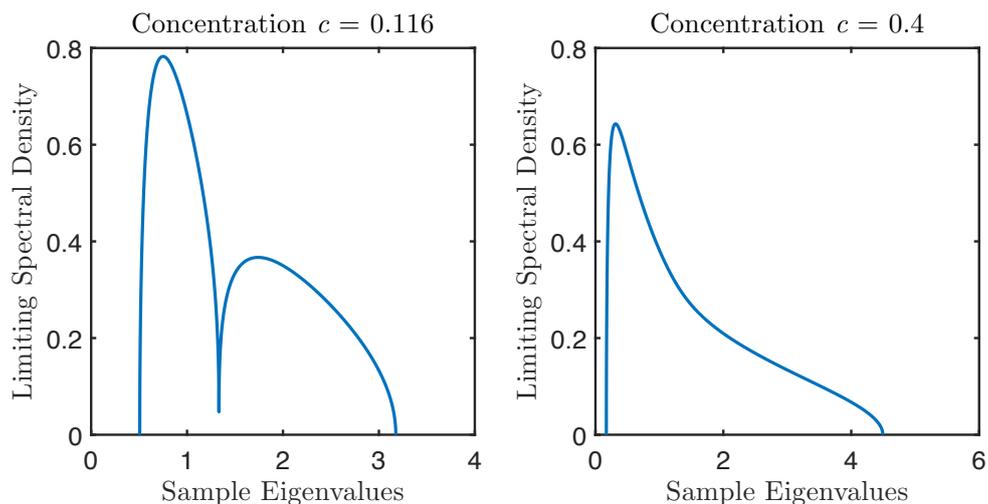


Figure 4: Phase transition: two clusters merge into one as the concentration ratio N/T increases. In the left panel, the two original clusters centered around 1 and 2, respectively, have already merged, but are still visually distinguishable from each other. In the right panel, the concentration ratio has increased so much that the two clusters can no longer be distinguished by visual inspection.

The tenth and final in our series of qualitative observations, which will all carry over to the general case, is:

10. Clusters that are too close to one another merge for sufficiently high concentration ratios.

On the right panel, the merger of the two clusters is so complete that we no longer even have a bimodal distribution. The naked eye cannot discern that half of the population eigenvalues are equal to one, and the other half to two. Only a purpose-built estimation process resting on advances in probability theory can. This estimation process is what Sections 9.6–9.8 detail.

9.6 Limiting Sample Spectral Distribution

In what follows, we describe only the case $0 < c < 1$. The case $c = 0$ is excluded because it is trivial: The sample eigenvalues converge to their population counterparts, and the sample covariance matrix is optimal. The case $c \geq 1$ is excluded because it would render the exposition less fluid, but it poses no great difficulty and can be dealt with just as effectively; see [Ledoit and Wolf \(2017a\)](#). One convenience that the assumption $0 < c < 1$ buys us is that the limiting sample spectral distribution admits a continuous derivative f .

In order to relate F to H quantitatively, a new mathematical object must be introduced: the [Stieltjes \(1894\)](#) transform. The Stieltjes transform of the distribution function F is defined on the half-plane of complex numbers with strictly positive imaginary part \mathbb{C}^+ by

$$\forall z \in \mathbb{C}^+ \quad m_F(z) := \int_{-\infty}^{+\infty} \frac{1}{\lambda - z} dF(\lambda). \quad (9.13)$$

Although it is difficult to visualize a complex function, there is an important result valid specifically for F that will help us gain an intuitive understanding of its Stieltjes transform. Indeed, F is smooth enough for the limit

$$\lim_{z \in \mathbb{C}^+ \rightarrow x} m_F(z) =: \check{m}_F(x) \quad (9.14)$$

to exist for all $x \in \mathbb{R}$. \check{m}_F extends the Stieltjes transform from the upper half of the complex plane onto the real line. Now \check{m}_F , being a complex-valued function of real argument, is much easier to comprehend. In particular, its imaginary part is simply the limiting spectral density f divided by π . Thus, if we have \check{m}_F , we get the density f and then also the distribution F (by integration of f). The quantiles of F tell us where the sample eigenvalues are located in the limit.

\check{m}_F can be deduced from H in only one known way, which is the following. For all $x \in \mathbb{R}$, $m := \check{m}_F(x)$ is the unique solution in \mathbb{C}^+ to the equation

$$\boxed{m = \int_{-\infty}^{+\infty} \frac{1}{\tau [1 - c - c x m] - x} dH(\tau)} \quad (9.15)$$

Equation (9.15) is the fundamental building block of all research in large-dimensional covariance matrix estimation. Although it may look daunting, and indeed there is generally no closed-form solution, it can be solved numerically in a matter of seconds for $N = 1000$ eigenvalues; see [Ledoit and Wolf \(2017b\)](#). This equation has been around for half a century in some form or other, and all researchers in the field are fully confident that “it does

exactly what it says on the tin”, even though it is difficult to provide intuitive insight; see Marčenko and Pastur (1967), Silverstein and Bai (1995), Silverstein (1995), as well as the authoritative monograph by Bai and Silverstein (2010).

F is smooth, it is more spread out than H , and the excess spread increases in c . In the limit, as $c \rightarrow 0$, we can recognize on the right-hand side the Stieltjes transform of H , so F becomes identical to H , which was to be expected from finite-dimensional asymptotics. All ten of the qualitative observations gathered from the study of simple cases in Sections 9.4–9.5 carry over to the solution of Equation (9.15). Indeed they pretty much encompass all the intuition that can be extracted from this equation. Figure 5 provides an illustration of the difference between F and H .

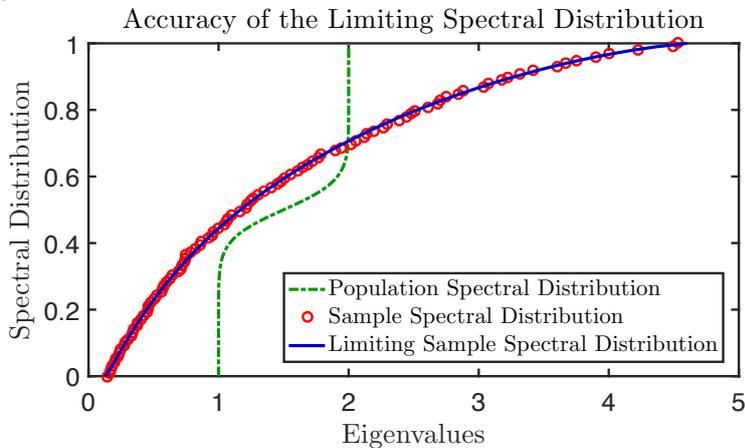


Figure 5: Three spectral distributions for $N = 100$ and $T = 200$. The population spectral distribution is the Beta(0.1,0.1) distribution shifted so the support is $[1, 2]$. Circles show the sample spectral distribution from one Monte Carlo simulation. One can observe that the sample eigenvalues are nowhere near their population counterparts, but their location is well predicted by the limiting spectral distribution F that comes out of Equation (9.15).

9.7 Discretization

Although Equation (9.15) theoretically solves the problem, it is not formulated in a directly usable way, as it relates the limiting spectral distributions F and H , whereas ideally we would want to relate the sample eigenvalues $(\lambda_{1,T}, \dots, \lambda_{N,T})$ to their population counterparts $(\tau_{1,T}, \dots, \tau_{N,T})$. A practical implementation is achieved through discretization.

Start from a family of N population eigenvalues $\boldsymbol{\tau} := (\tau_{1,T}, \dots, \tau_{N,T})$, which can be any

vector in $(0, +\infty)^N$. Construct the population spectral distribution H_T as per Equation (9.8) and inject it into Equation (9.15), replacing c with the ratio N/T . A trivial simplification shows that, for all $x \in \mathbb{R}$, $m := \check{m}_F(x)$ is the unique solution in \mathbb{C}^+ to the equation

$$m = \frac{1}{N} \sum_{i=1}^N \frac{1}{\tau_{i,T} \left[1 - \frac{N}{T} - \frac{N}{T} x m \right] - x}. \quad (9.16)$$

Having solved for the Stieltjes transform \check{m}_F numerically, we multiply its imaginary part by π to obtain the limiting sample spectral density f . Integration yields the limiting sample spectral distribution F . Finally, we can invert the function F to compute the N distribution quantiles.

This is exactly how the QuEST function is constructed. It maps an N -dimensional vector $\boldsymbol{\tau}$ of population eigenvalues into another N -dimensional vector $(q_{N,T}^1(\boldsymbol{\tau}), q_{N,T}^2(\boldsymbol{\tau}), \dots, q_{N,T}^N(\boldsymbol{\tau}))$, which represents a deterministic equivalent of the sample eigenvalues, by discretizing Equation (9.15). The output (quantiles of F) can be interpreted as the expectation of the sample eigenvalues, although this correspondence is rigorous only in the large-dimensional asymptotic limit.

9.8 Recovering Population Eigenvalues

As can be gathered from the above exposition, it is more straightforward to go from population to sample eigenvalues than the other way around. This is an intrinsic feature of Equation (9.15), which is our only tractable hook into the underlying mathematical truth.

However, given that Equation (9.15) has been made more practical through discretization inside the QuEST function as seen in Section 9.7, inverting it becomes a simple numerical problem. Find the vector $\boldsymbol{\tau} \in (0, +\infty)^N$ such that the function's output $(q_{N,T}^1(\boldsymbol{\tau}), q_{N,T}^2(\boldsymbol{\tau}), \dots, q_{N,T}^N(\boldsymbol{\tau}))$ matches most closely the observed sample eigenvalues $(\lambda_{1,T}, \dots, \lambda_{N,T})$. Any capable off-the-shelf nonlinear optimizer can solve this problem. Two recommended optimizers are Stanford Business Software's SNOPT and Matlab's `fmincon`; they both can easily handle dimensions up to $N = 1000$ within a reasonable amount of time, a few minutes at most. In a nutshell, the problem of recovering the population eigenvalues from the sample eigenvalues has been resolved by numerically inverting the multivariate function that discretizes Equation (9.15). Figure 6 illustrates the accuracy of this procedure.

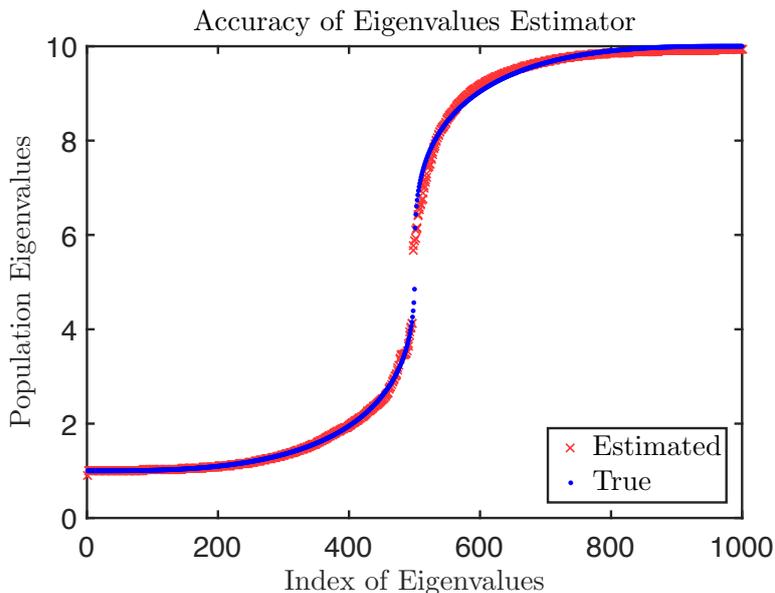


Figure 6: The matrix dimension is $N = 1000$ and the sample size is $T = 3000$. This graph is based on a single Monte Carlo simulation. Numerically inverting the discretized version of Equation (9.15) asymptotically recovers the population eigenvalues.

9.9 Nonlinear Shrinkage

At this juncture, it might be tempting to conclude that replacing the observed sample eigenvalues with the estimated population eigenvalues that come from inverting the QuEST function yields the optimal estimator of the covariance matrix. However this is not the case. The population eigenvalues (or consistent estimators thereof) are only optimal when recombined with the population eigenvectors. The latter are unobservable and, unlike the eigenvalues, there is no hope of recovering them through some advanced mathematics. The reason is that they live in a space of dimension $N(N - 1)/2$, which is infinitely too large given that we only collect $N \times T$ noisy data points, T being of the same order of magnitude as N .

In the terminology of Section 9.1, we do not want $\lambda_{i,T}$, the in-sample variance of the sample eigenvector $u_{i,T}$; but we also do not want $\tau_{i,T}$, the out-sample variance of the population eigenvector $v_{i,T}$. This is because we do not have $v_{i,T}$. What we want is a hybrid: $u'_{i,T} \Sigma u_{i,T}$, the out-sample variance of the sample eigenvector $u_{i,T}$. This quantity is estimated consistently

under general asymptotics by the following “nonlinear shrinkage” formula:

$$u'_{i,T} \Sigma u_{i,T} \approx \frac{\lambda_{i,T}}{\left| 1 - \frac{N}{T} - \frac{N}{T} \lambda_{i,T} \check{m}_F(\lambda_{i,T}) \right|^2}, \quad (9.17)$$

where $|\cdot|$ denotes the modulus of a complex number, and \check{m}_F is the function defined in (9.16); see Ledoit and Péché (2011, Theorem 3). Similar to Equation (9.15), it is hard to give intuition: this is just what comes out of the underlying mathematics. We are fortunate to have any explicit equation at all; it basically comes from a generalization of Equation (9.15). Extensive Monte Carlo simulations confirm the accuracy of this formula; see Ledoit and Wolf (2012, Section 6). One can see that when N/T is negligible, there is a negligible amount of shrinkage, as expected from finite-dimensional asymptotics.

Calling the right-hand side $\psi_{i,T}$ and injecting it into Equation (9.7) yields an estimator of the covariance matrix that improves upon the sample covariance matrix when the dimension N is not negligible with respect to the sample size T . A graphical illustration is given in Figure 7.

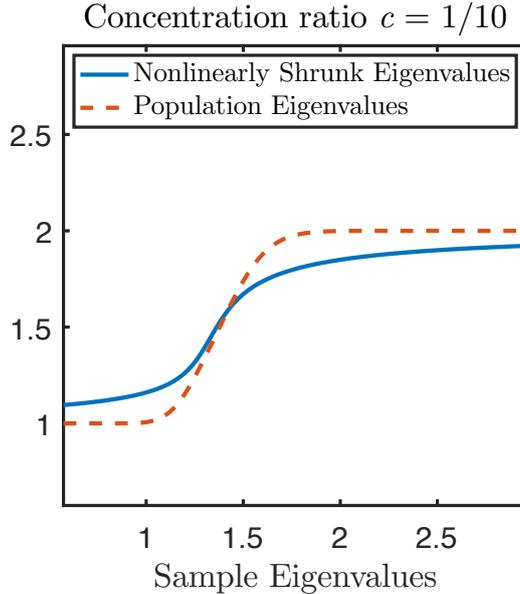


Figure 7: Nonlinearly shrunken eigenvalues as a function of sample eigenvalues. The population eigenvalues come from the Beta(0.1,0.1) distribution, shifted so the support is $[1,2]$. The optimal shrinkage transformation is highly nonlinear. The population eigenvalues are also plotted for reference as the dashed line. They are not the same as the shrunken eigenvalues, and are more spread out.

To summarize, the overall procedure consists of three consecutive steps:

- Step 1** Given the sample eigenvalues $(\lambda_{1,T}, \dots, \lambda_{N,T})$, invert the QuEST function defined in Section 9.7 to obtain consistent estimates of the population eigenvalues;
- Step 2** Plug the (estimated) population eigenvalues into Equation (9.16) to compute the complex-valued function \check{m}_F ;
- Step 3** Replace the sample eigenvalues with the nonlinear shrinkage formula on the right-hand side of Equation (9.17) while preserving the sample eigenvectors.

All of this is handled automatically by the QuEST software, available from the university faculty website of Michael Wolf. Ledoit and Wolf (2017a) prove that the resulting covariance matrix estimator is optimal for portfolio selection under general asymptotics within the rotation-equivariant class.

References

- Bai, Z. D. and Silverman, J. W. (2010). *Spectral Analysis of Large-Dimensional Random Matrices*. Springer, New York, second edition.
- Engle, R. F. (1982). Autoregressive conditional heteroskedasticity with estimates of the variance of United Kingdom inflation. *Econometrica*, 50:987–1007.
- Ledoit, O. and P ech e, S. (2011). Eigenvectors of some large sample covariance matrix ensembles. *Probability Theory and Related Fields*, 150(1–2):233–264.
- Ledoit, O. and Wolf, M. (2004). A well-conditioned estimator for large-dimensional covariance matrices. *Journal of Multivariate Analysis*, 88(2):365–411.
- Ledoit, O. and Wolf, M. (2012). Nonlinear shrinkage estimation of large-dimensional covariance matrices. *Annals of Statistics*, 40(2):1024–1060.
- Ledoit, O. and Wolf, M. (2017a). Nonlinear shrinkage of the covariance matrix for portfolio selection: Markowitz meets Goldilocks. *Review of Financial Studies*, 30(12):4349–4388.
- Ledoit, O. and Wolf, M. (2017b). Numerical implementation of the QuEST function. *Computational Statistics & Data Analysis*, 115:199–223.

- Marčenko, V. A. and Pastur, L. A. (1967). Distribution of eigenvalues for some sets of random matrices. *Sbornik: Mathematics*, 1(4):457–483.
- Markowitz, H. (1952). Portfolio selection. *Journal of Finance*, 7:77–91.
- Michaud, R. (1989). The Markowitz optimization enigma: Is optimized optimal? *Financial Analysts Journal*, 45:31–42.
- Perlman, M. D. (2007). *STAT 542: Multivariate Statistical Analysis*. University of Washington (On-Line Class Notes), Seattle, Washington.
- Silverstein, J. W. (1995). Strong convergence of the empirical distribution of eigenvalues of large-dimensional random matrices. *Journal of Multivariate Analysis*, 55:331–339.
- Silverstein, J. W. and Bai, Z. D. (1995). On the empirical distribution of eigenvalues of a class of large-dimensional random matrices. *Journal of Multivariate Analysis*, 54:175–192.
- Silverstein, J. W. and Choi, S. I. (1995). Analysis of the limiting spectral distribution of large-dimensional random matrices. *Journal of Multivariate Analysis*, 54:295–309.
- Stein, C. (1975). Estimation of a covariance matrix. Rietz lecture, 39th Annual Meeting IMS. Atlanta, Georgia.
- Stieltjes, T. J. (1894). Recherches sur les fractions continues. *Annales de la Faculté des Sciences de Toulouse 1^{re} Série*, 8(4):J1–J122.
- Wigner, E. P. (1955). Characteristic vectors of bordered matrices with infinite dimensions. *Annals of Mathematics*, 62(3):548–564.
- Yin, Y. (1986). Limiting spectral distribution for a class of random matrices. *Journal of Multivariate Analysis*, 20(1):50–68.