Numerical Solutions to the General Marcenko Pastur Equation

Miriam Huntley

May 20, 2013

1 Motivation: Data Analysis

A frequent goal in real world data analysis is estimation of the covariance matrix of some multivariate variable. For example say we have some true covariance matrix Σ that is dictating the values of our p dimensional variable $\mathbf{x}$. Then we would like to know, from a given n samples of $\mathbf{x}$, how to estimate Σ.

We will call our data matrix X, a p by n matrix, with p as the dimension of our multivariate variable and n the number of independent samples. We assume that X is drawn from a mean-zero p-dimensional multivariate distribution $X \sim N_p(0, \Sigma)$. Alternatively, we can write $X = Y\Sigma^{1/2}$, where Y is a pxn matrix with random iid and mean-zero entries, and $\Sigma^{1/2}$ is the square root of Σ.

To estimate the covariance matrix, one might just look at the empirical covariance matrix, $E = \frac{1}{n}X^T X$. When $p >> n$, this is a good approximation to Σ. However, for p/n finite, the spurious correlations due to finite sampling will greatly distort the resulting estimated matrix.

For this reason, many areas of data analysis perform Principle Components Analysis (PCA) to separate signal from noise in the empirical covariance matrix. To do this, one first looks at the eigenvalues of E, which have some spread. To estimate Σ, PCA takes the eigenvectors associated with the largest k eigenvalues of E to form a rank k matrix. In principle, to choose how many of the eigenvectors one should keep, the procedure is to only take those eigenvalues that fall well outside the eigenvalue distribution of the null model. The null model that is being ruled out in this case is that the true covariance matrix is just an identity matrix; the expected eigenvalue distribution of E is then given by the White Marcenko Pastur equation, discussed below. Thus in PCA, one chooses the eigenvalues that are well outside this null distribution and calls those the 'signal' eigenvectors.

Despite this practice being widespread, it seems clear that there is room for improvement. The reliance on using an identity matrix as a null model for the true covariance matrix is quite crude in areas of statistics where more detailed and informative null models are available. In fact, there are null models where the distribution of the empirical eigenvalues has no right edge, which would imply that performing some sort of simple cutoff using PCA would be inappropriate. Therefore, in this project we examine the eigenvalue distributions of E for general covariance matrices.

2 General Marcenko Pastur Equation

2.1 Historical Note

In 1967 Marcenko and Pastur wrote a paper describing the general solution to the spectrum of empirical covariance matrices. They also solved this problem for a specific covariance matrix of interest, the white Wishart matrix (I_p). This simple closed form solution is often called the Marcenko Pastur equation, which can lead to a bit of confusion. For the purpose of this paper, the white Marcenko Pastur (or WMP) will refer to the equation that describes the distribution of white Wishart matrices, while the Marcenko Pastur (MP) equation will refer to the self consistent equation that is valid for general covariance matrices.

2.2 The Stieltjes Transform

Before proceeding to the MP equation, we need to define the Stieltjes transform (also known as the Cauchy transform). For some spectral distribution $f(\lambda)$, we define its Stieltjes transform $m_f(z) = \int \frac{f(\lambda)d\lambda}{\lambda - z}$ defined $\forall z \in C^+$. The need for such a transform is not immediately obvious, but it turns out that many results, including the MP equation and in general free probability manipulations, rely on it. The important thing to remember about the Stieltjes transform, besides for how to correctly spell it, is that it carries the same information as the density distribution which it represents. Like a fourier transform, we can transform from one to the other. The 'reverse Stieltjes transform' is given by the relation $f(\lambda) = \lim_{\epsilon \to 0} \frac{1}{\pi} Im[m_f(x + i\epsilon)]$

We can also define the Stieltjes transform of a matrix: for a pxp matrix A we can compute its eigenvalues,$\{\lambda_i\}$. From this we can create a continuous eigenvalue distribution which just places a dirac delta function at the locations of each of the eigenvalues, $f_A(x) = \frac{1}{p}\sum_{i=1}^{p} \delta(x - \lambda_i)$. From this we could write $m_{f_A} = \frac{1}{p}\sum_{i=1}^{p} \frac{1}{\lambda_i - z}$, or alternatively $m_{f_A}(z) = \frac{1}{p}Tr[(A - zI)^{-1}]$. (For convenience we will do away with the f in the subscript and write m_{f_A} as m_A below.)

2.3 Relation between $X^T X$ and XX^T

For the data matrix X we can of course define two empirical covariance matrices, $E_p = \frac{1}{n}X^T X$ and $E_n = \frac{1}{n}XX^T$ (where for convenience we have divided by n in both cases). As defined above, the "interesting" covariance matrix that we care about is that which describes the relation between the p dimensions, which would be embodied in the $X^T X$ matrix. $X^T X$ and XX^T share the same nonzero eigenvalues; it is just that if $n > p$ then XX^T will have $n - p$ zero eigenvalues (and vice versa is $n < p$). The useful outcome of this fact is that since the eigenvalue distributions of the two covariance matrices are very cleanly defined, we can also write down the formula for the relation between the Stieltjes transforms of the two distributions:

$$m_{E_n}(z) = \frac{1}{z}(\frac{p}{n} - 1) + \frac{p}{n} m_{E_p}(z) \tag{1}$$

The results below will be stated in terms of $m_{E_n}(z)$, despite the fact that as stated above, we really are interested in m_{E_p}. However, transforming between the two using the above equation is easy to do.

2.4 The MP Equation

The MP equation has been reformulated a few times since its appearance in the original paper. Here we choose a particular instantiation of the equation that was laid out by Silverstein in 1995 [2]. In his paper, he proves the MP equation under the following set of assumptions.

Let X be the data matrix, which we assume can be written $X = Y\Sigma_p^{1/2}$. Here Σ_p is the 'true covariance matrix', a pxp positive definite matrix. Y is an nxp matrix with iid entries. Let $E[Y_{i,j}] = 0$ and $E[|Y_{i,j}|^2] = 1$. Let the spectral distribution of Σ_p be $f_{\Sigma_p(}\lambda)$, and assume that f_{Σ_p} converges weakly to f_{Σ_∞} as $p \to \infty$. Let $m_{E_n}(z)$ be the Stieltjes transform of the empirical covariance matrix $E_n = \frac{1}{n}XX^T$. Then the Marcenko Pastur result says that $m_{E_n}(z)$ approaches $m_\infty(z)$ asymptotically, where m_∞ is given by the equation:

$$-\frac{1}{m_\infty(z)} = z - \gamma \int \frac{\lambda f_{\Sigma_\infty(}\lambda) d\lambda}{1 + \lambda m_\infty(z)} \tag{2}$$

For general forms of $f_{\Sigma_\infty(}\lambda)$, it is not possible to find an analytical solution for m. However, we note that in the well known WMP case, we can readily solve Eqn 2 by plugging in $f_{\Sigma_\infty(}\lambda) = \delta(\lambda - 1)$ to give us $\frac{1}{m_\infty} = z - \gamma\frac{1}{1+m_\infty}$, which can then be solved for m and then f_{E_p}.

There are a few other cases where an analytical solution to Eqn 2 is readily obtainable. However, for the general covariance matrices that we are interested in, we must turn to numerical solutions.

3 Numerical Solutions to the MP Equation

3.1 How to Solve MP

The road to numerical solutions of the MP equation is straightforward and relatively obstacle-free. In order to solve Eqn 2 above, we first break up the equation into its real and imaginary parts. Then we can discretize in z: that is, for each z of interest we solve for the real and imaginary parts of m that solve the equations. To solve, one could use their favorite optimization algorithm of choice. Here we decided to use the built in Matlab function fsolve, and this seems to work reasonably well. Once we obtain m_{E_n}, we can turn that into m_{E_p} using Eqn 1 and then obtain f_{E_p} using the reverse Stieltjes transform.

3.2 Examples

As a first test, we input the identity matrix as the true covariance matrix to our solver. This corresponds to a spectral distribution $f_\Sigma(x) = \delta(x-1)$, the classical WMP. As shown in Fig 1, the solver gives us the correct distribution, confirmed with Monte Carlo simulations as well.

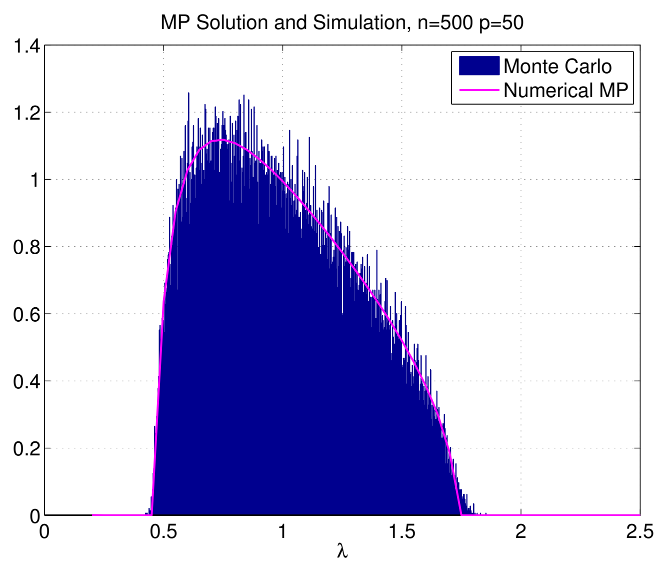

Figure 1: Numerical solution to the Marcenko Pastur equation (in pink) and Monte Carlo (in blue) of a white Wishart matrix. As expected, we see the WMP distribution.

Next, we look at the distribution of a slightly less trivial covariance matrix, as pictured below in Fig 2 . The solution to the MP equation and the simulations match up well.

We would also like to see solutions to the MP equation for true continuous distributions of eigenvalues, not just sums of delta functions. We provide two examples here. The first looks at $f_\Sigma(x) = \frac{1}{\mu}e^{-x/\mu}$. The second example gives as input the Laguerre matrix ensemble, with the WMP distribution:

$$\begin{aligned} f_\Sigma(x) &= \frac{1}{2\pi\gamma x}\sqrt{(b-x)(x-a)} \\ a &= (1-\gamma^{1/2})^2 \\ b &= (1+\gamma^{1/2})^2 \end{aligned}$$

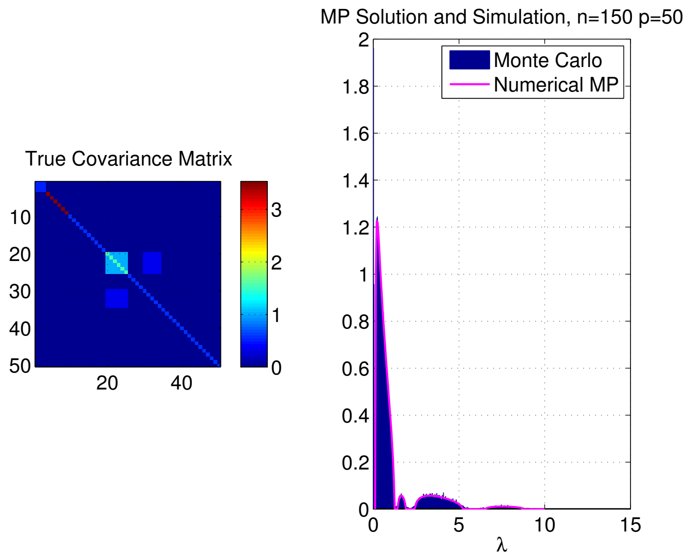

Figure 2: A slightly more complicated case. The input to the solver in this case is the eigenvalue density of the matrix on the left. The resulting numerical solution matches well with simulations.

In both examples, we see that the MP solution matches experiments well.

4 Further Directions

Thus far, we have used Eqn 2 by solving for m for a given $f_{\Sigma_p(}(\lambda)$. In principle one would in fact like to do the opposite. That is, for a given empirical distribution of eigenvalues, we would like to find the best true covariance spectral distribution $f_{\Sigma_p(}(\lambda)$. Unfortunately, searching over $f_{\Sigma_p(}(\lambda)$ space to find the distribution which best satisfies Eqn 2 seems difficult. However, it may be possible to perform such a search where the covariance matrix is parameterized by a small number of parameters. This seems to work in very simple cases (such as spiked covariance matrices), and it remains to be seen whether or not this could be generalized to be of practical use in covariance matrix estimation. For more on this topic see for example [3].

References

[1] Marcenko , V. A. and Pastur, L. A. (1967). Distribution of eigenvalues in certain sets of random matrices. Mat. Sb. (N.S.) 72 507?536.

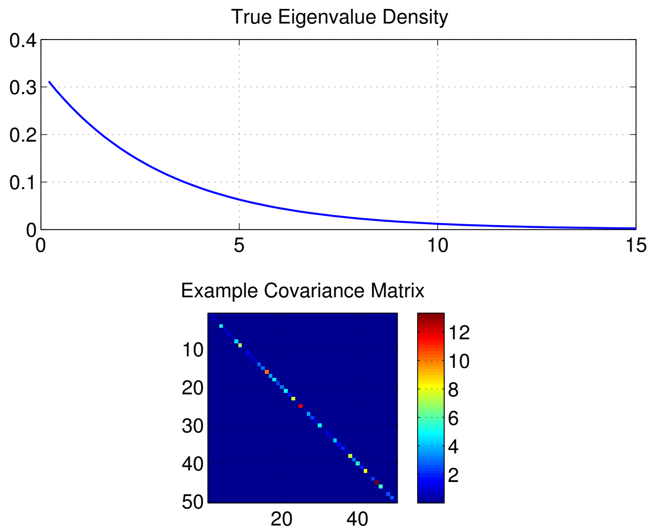

Figure 3: The exponential distribution of eigenvalues above, and an example covariance matrix below.

[2] Silverstein, J. W. (1995). Strong convergence of the empirical distribution of eigenvalues of large-dimensional random matrices. J. Multivariate Anal. 55, 2, 331? 339.

[3] El Karoui, N., Spectrum estimation for large dimensional covariance matrices using random matrix theory, Ann. Statist. 36 (2008), 2757?2790

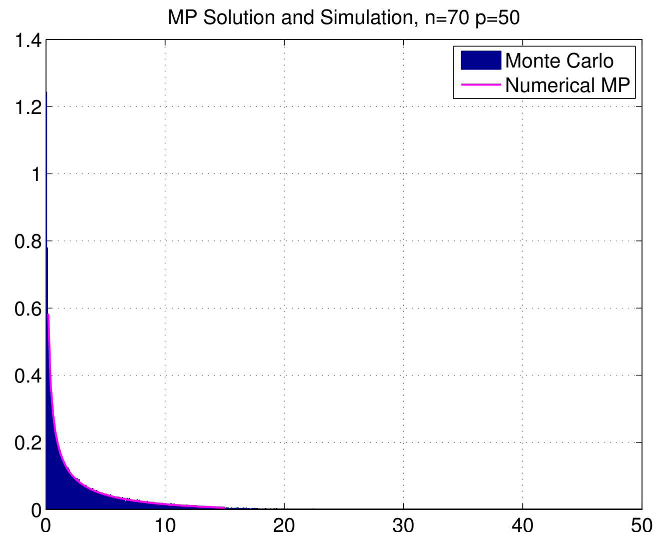

Figure 4: Numerical and experimental results for the exponential eigenvalue density.

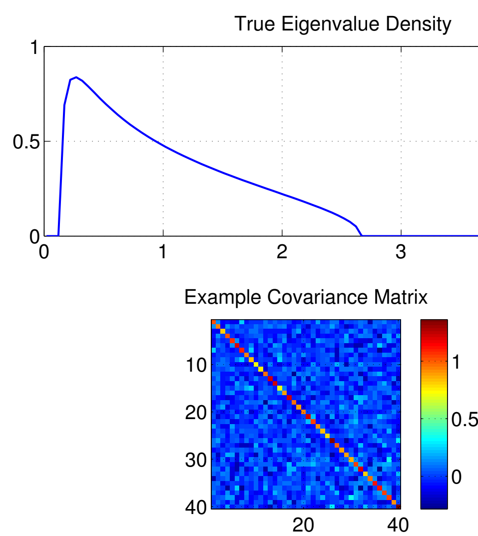

Figure 5: The WMP distribution of eigenvalues shown above, and an example covariance matrix below.

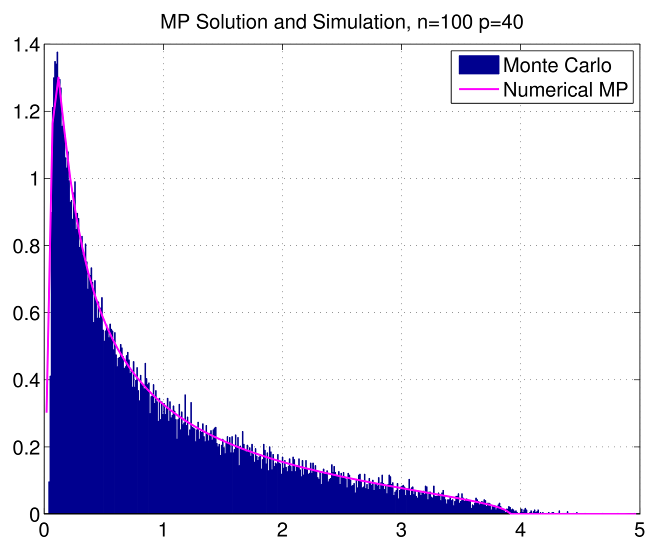

Figure 6: Numerical and experimental results for the WMP eigenvalue density.