International Journal of Applied Mathematical Research, 2 (1) (2013) 91-98
©Science Publishing Corporation
www.sciencepubco.com/index.php/IJAMR

A Monte Carlo approach to estimating the spectrum of large positive semidefinite matrices

Roberto Ragona

ENEA, Dept. of Advanced Technologies for Energy and Industry, Via Anguillarese, 301 - 00123 Rome (Italy)
Corresponding author E-mail: roberto.ragona@enea.it

Abstract

The method presented in this paper aims at furnishing a series of numerical values to approximate the matrix eigenvalues. The method is based on a statistical model that involves the first three central moments of the eigenvalue distribution, and it relies on the solution of a nonlinear system of equations that implements the moment-matching method and on a subsequent procedure of Monte Carlo simulations. The method is only applicable to real positive semidefinite matrices (PSD), and it is especially useful when other techniques lead to computational problems, e.g., when the matrices become too large to be processed or the required storage space is the cause of heavy limits to the computational process.

Keywords: *Eigenvalue estimation, Matrix spectrum, Monte Carlo simulations, Moment-matching, Powered gamma distribution.*

1 Introduction

Matrix eigenvalue problems are present in a large number of scientific and engineering fields, including statistical data analysis, building design, vibration analysis, and electric networks. Most approaches use algebraic methods to solve eigenvalue problems, which may involve matrices of a general form or a particular form (e.g., symmetric matrices), and several software packages that can solve such problems are available; this type of software is readily useable to solve eigenvalue problems and works properly if the available computer memory is sufficient to store the matrix in question. Typically, this type of software returns all of the eigenvalues and eigenvectors.

In some applications, the matrix becomes too large to be stored or processed (i.e., when the intermediate results need a large amount of extra storage space); in this case, special approaches must be considered, and these approaches are executed using dedicated procedures, for example, for large sparse matrices [1].

In many cases, the largest or smallest eigenvalue and/or the range between them provide valuable information. In other cases, for example, in least-squares problems, the need to evaluate the numerical rank of a matrix arises; this problem is not a straightforward computational problem and it can be affected by numerical questions.

For this type of problem, techniques exist that are useable deterministically [1]; however, the present technique uses a different approach. This technique is based on a Monte Carlo procedure, which in general can provide results in applications where the system being studied is not stochastic, but nevertheless a stochastic algorithm is the most efficient, or the most accurate, or the only feasible method for studying the system.

The method discussed in this paper is especially useful when the given problem involves large matrices. To demonstrate this method, we will use it on medium-size matrices (on the order of 2000 × 2000), even though the eigenvalues and eigenvectors of these matrices could be found directly on many desktop computers.

The method aims at furnishing a series of values to approximate the matrix eigenvalues, and it can easily be implemented on large matrices; it is based on a statistical model that involves the first three central moments of the eigenvalue distribution, as applied in a Monte Carlo context.

Eigenvalues are supposed to be organised in descending order (i.e., with respect to their rank), implementing an order statistics [2], and furthermore, their approximating values are assumed to conform to this assumption. The reasons for this assumption will be presented later.

Finally, it is worthwhile to mention that this method is only applicable to real positive semidefinite matrices, although such matrices have many applications, e.g., considering the covariance matrices and one of their key applications, Principal Component Analysis (PCA).

2 Some facts of matrix theory

The fact that the eigenvalues of an $(n \times n)$ matrix **A** are the roots of the n-th degree polynomial equation, $\det(\mathbf{A} - \lambda I) = 0$, is well known.
We now establish a connection among the central moments of the eigenvalue distribution and the traces of **A**, $\mathbf{A}^2$, …, $\mathbf{A}^k$, …, which are the consecutive powers of **A**. This connection can be demonstrated in a variety of ways, for example, by recalling the following theorem [3], which is presented without a demonstration:

Theorem 2.1: *If* $\lambda_1, \lambda_2, \ldots, \lambda_n$ *are the eigenvalues (with the proper multiplicities) of an* $(n \times n)$ *matrix* **A** *and if* $g(x)$ *is a scalar polynomial, then* $g(\lambda_1), g(\lambda_2), \ldots, g(\lambda_n)$ *are the eigenvalues of* $g(\mathbf{A})$.

In particular, if **A** has the eigenvalues $\{\lambda_1, \lambda_2, \ldots, \lambda_n\} \stackrel{\text{def}}{=} \Lambda$, then $\mathbf{A}^k$, which represents the matrix **A** raised to the power k, has the eigenvalues $\{\lambda_1^k, \lambda_2^k, \ldots, \lambda_n^k\} \stackrel{\text{def}}{=} \Lambda_k$.
It follows that

$$\left.\begin{array}{lll} \text{trace}(\mathbf{A}) = \Sigma\, \lambda_i & \longrightarrow & \Lambda_m = \Sigma\, \lambda_i / n = \text{trace}(\mathbf{A})/n; \\ \text{trace}(\mathbf{A}^2) = \Sigma\, \lambda_i^2 & \longrightarrow & \sigma^2_\Lambda = 1/n \bullet \Sigma\, (\lambda_i - \Lambda_m)^2 = \text{trace}(\mathbf{A}^2)/n - [\text{trace}(\mathbf{A})/n]^2 \\ \text{trace}(\mathbf{A}^3) = \Sigma\, \lambda_i^3 & \longrightarrow & Sk_\Lambda = \Sigma\, (\lambda_i - \Lambda_m)^3 / (n \bullet \sigma^3_\Lambda) = \{[\text{trace}(\mathbf{A}^3) - 3\Lambda_m\, \text{trace}(\mathbf{A}^2)]/n + 2\Lambda_m^3\}/\sigma^3_\Lambda \end{array}\right\} \quad (1)$$

where Λ_m, σ^2_Λ, and Sk_Λ represent the sample mean value, the sample variance and the sample skewness of the eigenvalue distribution, respectively.
Although higher moments are obtainable, we will limit ourselves to the first three moments, due to the application of our method to large matrices.
In fact, it can easily be demonstrated that the first three moments are calculable with in-place procedures, that is, without the need to store any intermediate results in memory areas as large as the base matrix **A**. Indeed, the computing procedure requires only a small amount of extra storage space.
The following pseudo-code justifies this assertion:

```
Vr=zeros(1,1:n);          % A zeroed n-element row vector to contain the intermediate results
tr2=0;                    % The initial setting for trace(A^2)
tr3=0;                    % The initial setting for trace(A^3)
for i=1:n                 % A(i,1:n) represents the i-th row of A, A(1:n, j) represents the j-th column of A
        for j=1:n
            Vr(j)= A(i,1:n) • A(1:n,j);        % dot product
        end
        tr2=tr2+Vr(i);
        tr3=tr3+Vr • A(1:n,i);                 % dot product
end
```

Upon exiting the outerloop, the variables tr2 and tr3 will contain $\text{trace}(\mathbf{A}^2)$ and $\text{trace}(\mathbf{A}^3)$, respectively.
In practical applications, it is of interest to estimate the largest eigenvalue that could arise. An iterative procedure called the *power method* is available, and it is one of the most widely used processes for the estimation of the largest eigenvalue of a real matrix **A**. As a result, the power method is also suitable for matrices of higher orders. For details, see for example [1, §4.1]. In general, the power method's convergence is not assured to be fast.

3 Statistical assumptions

Fitting a distribution to a set of data is an important task in data analysis, for various reasons.
In our case, the fitted distribution forms the basis of generating simulation scenarios, as dictated by the Monte Carlo method that is adopted here.
In section 2, we mentioned the fact that a probability distribution can affect the eigenvalues, and we showed how to calculate the distribution's sample central moments. In effect, this assumption is crucial in our paper: the eigenvalues are stochastic in nature and a probability distribution can represent them conveniently.
With this assumption in mind, we now have to choose a proper law of eigenvalue distribution and find a method to identify this distribution's free parameters based on the data we possess, namely, the first three sample moments assigned by the relations reported in (1).

In general, a data sample is available and a fitting can be conducted on it in a variety of ways. Among all of these methods, we mention the *moment matching* method, which dates back at least to K. Pearson, in the late 1800's. This method proceeds by prescribing that a fixed number of sample (central) moments must match the moments of the theoretical distribution. The quality of the fitted distribution can be affected to some degree because even if the fixed number of moments is matched exactly, one cannot be assured that the resulting theoretical distribution will perfectly match the empirical distribution. In general, this method is recognised to be inferior to other approaches, especially the maximum likelihood method. However, in our case, this alternative approach is not applicable because some terms of the log likelihood function are out of reach.
Thus, despite the possible drawbacks, the moment matching method suits our context well.
In effect, in our method, we will pursue a match among the three sample moments that are generated by relations (1) and the first three moments of the theoretical distribution law, which we now introduce.
Various possible distributions to represent observed data have been defined, and among these distributions, we propose using the *powered gamma distribution* (or Kritskii-Menkel distribution) because it is extremely useful for representing observations on a continuous positive random variable [4, §11.11]. In fact, the possession of three adjustable parameters gives this family a large degree of flexibility that may make a satisfactory fit possible.
Let Gamma(α, β) be a standard gamma variate with shape parameter $\alpha > 0$ and scale parameter $\beta > 0$ [4, §11.3]. The powered gamma distribution, which has a power parameter $\nu > 0$, is the distribution of the variate

$$Z(\alpha, \beta, \nu) = [\text{Gamma}(\alpha, \beta)]^{\nu}, \quad Z \geq 0 \tag{2}$$

and this distribution is the distribution of a standard gamma variate raised to the power ν.
It can be demonstrated [4, §11.11] that the variate's first three central moments are given by

$$\left.\begin{aligned} Z_m &= E\{Z(\alpha, \beta, \nu)\} = \beta^{\nu} \bullet [\Gamma(\alpha + \nu) / \Gamma(\alpha)] = \beta^{\nu} \bullet G(\alpha,\nu) \\ \sigma_Z^2 &= \text{var}\{Z(\alpha, \beta, \nu)\} = \beta^{2\nu} \bullet [\Gamma(\alpha + 2\nu) / \Gamma(\alpha) - G^2(\alpha,\nu)] = \beta^{2\nu} \bullet [G(\alpha,2\nu) - G^2(\alpha,\nu)] \\ Sk_Z &= \text{skew}\{Z(\alpha, \beta, \nu)\} = [G(\alpha,3\nu) + 2\, G^3(\alpha,\nu) - 3\, G(\alpha,\nu) \bullet G(\alpha,2\nu)] / \sigma_Z^3, \end{aligned}\right\} \tag{3}$$

where $\Gamma(\bullet)$ is the gamma function and $G(\alpha, k\nu) = \Gamma(\alpha + k\nu) / \Gamma(\alpha)$. Furthermore, Z_m represents the mean value, σ_Z^2 represents the variance and Sk_Z represents the skewness of the powered gamma distribution with parameters α, β, and ν.

4 Computational details

Because we have defined the principal ingredients of the method, we are now ready to detail its computational process.

4.1 Matching the first three central moments

As explained above, in our case, the moment matching method has to find values for the free distributional parameters of (2) because of the goal of making the three sample moments (1) as close as possible to being equal to the corresponding theoretical moments (3). In other words, in our situation, we have to find a solution, even approximate and that is as accurate as possible, to the following highly nonlinear system of three equations in three unknown values α, β, and ν:

$$\left.\begin{aligned} Z_m(\alpha, \beta, \nu) &= \Lambda_m; \\ \sigma^2_Z(\alpha, \beta, \nu) &= \sigma^2_\Lambda; \\ Sk_Z(\alpha, \beta, \nu) &= Sk_\Lambda, \end{aligned}\right\} \tag{4}$$

where Λ_m, σ^2_Λ and Sk_Λ are calculated using both the traces of **A** and the powers of **A**.
The nonlinearity of the problem indicates that an iterative search for the solution may be ideal, and many methods and software resources exist that use iterative searches; thus, they are able to address the problem.
In our analyses, we used the Matlab routine *fsolve*, which demonstrated an adequate level of ability to find suitable solutions; in any case, the constraints of *positivity* on all three solving parameters must be taken into account.
How to utilise the solution (α_s, β_s, ν_s) to system (4) (once it has been found) will be explained in the next subsection.

4.2 Monte Carlo simulations

Monte Carlo simulations are stochastic techniques in the sense that they are based on the use of random numbers and statistics to analyse problems.
Depending on the number of uncertainties, a Monte Carlo simulation could involve thousands or tens of thousands of recalculations before it is complete.
Our method uses this technique, because the three parameters α_s, β_s, and ν_s of the chosen powered gamma distribution have been identified as a solution to the nonlinear system (4).
The Monte Carlo simulation we propose is developed in consecutive steps. We summarise it thusly:

- a repeated extraction of gamma random samples of size n (where n is the order of **A**), with shape parameter α_s and scale parameter β_s, is performed. For example, this task can be carried out by calls to the Matlab routine *gamrnd;*
- every sample of size n is raised to the power ν_s, as prescribed by (2). In this way, we obtain a powered gamma random sample of size n;
- the above samples are then reorganised in descending order, which complies with the assumption of order statistics;
- finally, an index-by-index average of all of the ordered samples is performed, and these averages furnish the desired approximating values.

Therefore, the approximation for the i-th eigenvalue is obtained as an average of the i-th order statistics, that is, as an average of the i-th value of all of the ordered powered gamma random samples of size n.
In our work, the number of extractions of size n to consider was likewise fixed at n, the order of **A**.
The computational procedure is simple to perform, and furthermore, it requires only a small amount of extra storage space. If this space is available, then the simulation process can be implemented with an in-place procedure.
The rationale behind the key adoption of order statistics is that, in the absence of ordering, the Monte Carlo procedure would produce uniform averages because of randomness in the sample extraction. Therefore, ordering readdresses the results of averaging, and in effect, our procedure acts as a generator of realisations of random samples of ordered eigenvalues.

5 Results

In this section, we will present the results of our method as applied to arbitrary matrices, which were obtained by repeated applications of a uniformly distributed pseudorandom generator. The first three of these matrices have a common origin because they were derived from a random sample containing 2000 points $\mathbf{x}_i \in R^2$ located inside a square with sides of length 5 that was at the origin O.
Let us consider the following rule defined on the points $\mathbf{x}_i$ and producing results $A(i, k) \in R^+$:

$$A(i, k) = \exp(-\gamma \|\mathbf{x}_i - \mathbf{x}_k\|^2) \qquad (i, k = 1, 2, \ldots, 2000) \tag{5}$$

This rule generates matrices **A** (2000× 2000) that are defined by means of a kernel function (more precisely, a radial basis function, or RBF) [5], where γ is a positive parameter that controls the RBF radius. As a result, the matrices **A** are symmetric and positive semidefinite for any value of γ [5]. Furthermore, by varying this parameter, a variety of matrices may be obtained, and all of them have real positive or null eigenvalues, hence, they are assimilable to covariance matrices.
The initial three matrices we analyse pertain to the values $\gamma = 10$, $\gamma = 1$, and $\gamma = 0.1$, and they will be denoted by $\mathbf{A}_{10}$, $\mathbf{A}_1$, $\mathbf{A}_{0.1}$, respectively.
Fig.1 and fig. 2, which pertain to the matrices $\mathbf{A}_{10}$ and $\mathbf{A}_1$, illustrate the results of our method: we obtained good levels of agreement and overlapping between the ordered eigenvalues (the blue profile) and the corresponding approximating curve (the red profile). In addition, fig. 1 shows the entire overlapped population of 2000 ordered random powered gamma samples (the green band) of our simulation, but fig. 2 has a limited representation of only 500 eigenvalues; the remaining eigenvalues are graphically indistinguishable from zero.
Eigenvalues were invariably calculated by the double-precision Matlab routine *eig.*
Table 1 helps us to achieve a better understanding of this process.
In table 1, we report the following data: the solution (α_s, β_s, ν_s) to (4) that was furnished by the routine *fsolve*, the values Λ_m, σ^2_Λ and Sk_Λ, which were calculated by means of the traces, see (1), and their counterparts Z_m, σ^2_Z and Sk_Z, which were obtained by means of (3) when the solution (α_s, β_s, ν_s) was substituted there. Moreover, λ_{max} represents the largest eigenvalue and AC(1) represents the first value of the approximating curve, whose deviation from λ_{max} can constitute a sort of index of quality of the approximation relative to the region with the largest eigenvalues.

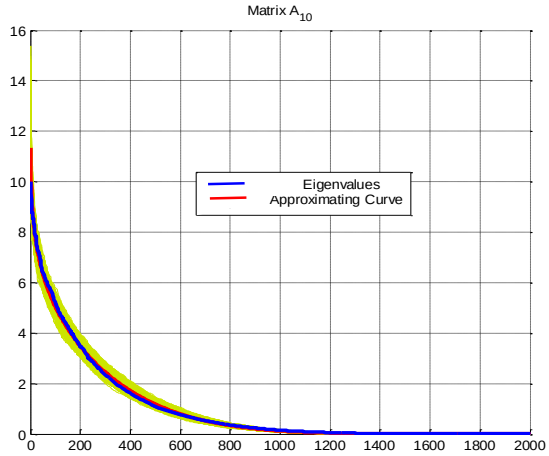

Fig. 1: The results for the matrix $\mathbf{A}_{10}$

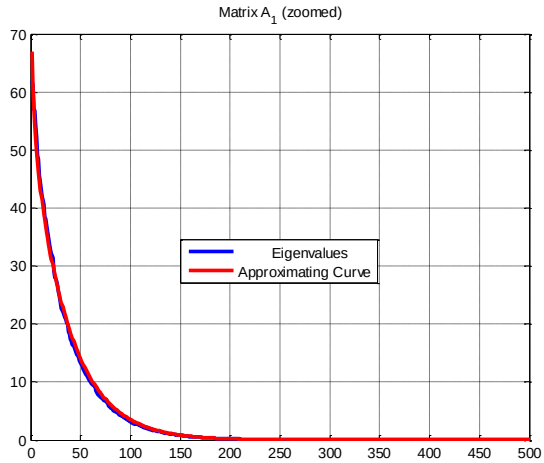

Fig. 2: The results for the matrix $\mathbf{A}_1$

Table 1: The characteristic values of the matrices $\mathbf{A}_{10}$, $\mathbf{A}_1$ and $\mathbf{A}_{0.1}$

	α_s	β_s	ν_s	Λ_m	Z_m	σ^2_Λ	σ^2_Z	Sk_Λ	Sk_Z	λ_{max}	AC(1)
$\mathbf{A}_{10}$	0.0572	443.229	0.3268	1	1.00	3.06	3.06	2.35	2.35	10.01	11.34
$\mathbf{A}_1$	0.0038	4.27e+8	0.2065	1	1.01	28.93	28.93	7.32	7.32	61.78	66.76
$\mathbf{A}_{0.1}$	9.35e-4	3.24e+7	0.3525	1	1.05	236.95	236.97	21.3	21.6	453.75	401.47

Fig. 3 presents the approximating curve relative to the matrix $\mathbf{A}_{0.1}$. To show the details more clearly, in the graph on the left, the representation is limited to the first 200 eigenvalues, and in the graph on the right, the representation is limited to the first 20 eigenvalues.

A general consideration can be portrayed: situations where the decay is fast and many eigenvalues tend to rapidly become smaller in their ordered sequence are harder to resolve because the solution to moment matching tends to assume extreme values to justify this large variation profile (for example, see the components α_s and β_s in the case corresponding to $\mathbf{A}_{0.1}$, which can be found on the third row of table 1).

In these situations, proper numerical procedures are to be employed. For example, extended precision may be required when calculating the values of the gamma function $\Gamma(\bullet)$ because it can rapidly tend to large values (in the limit to infinity when $\alpha \rightarrow 0$), which negatively affects the numerical procedure.

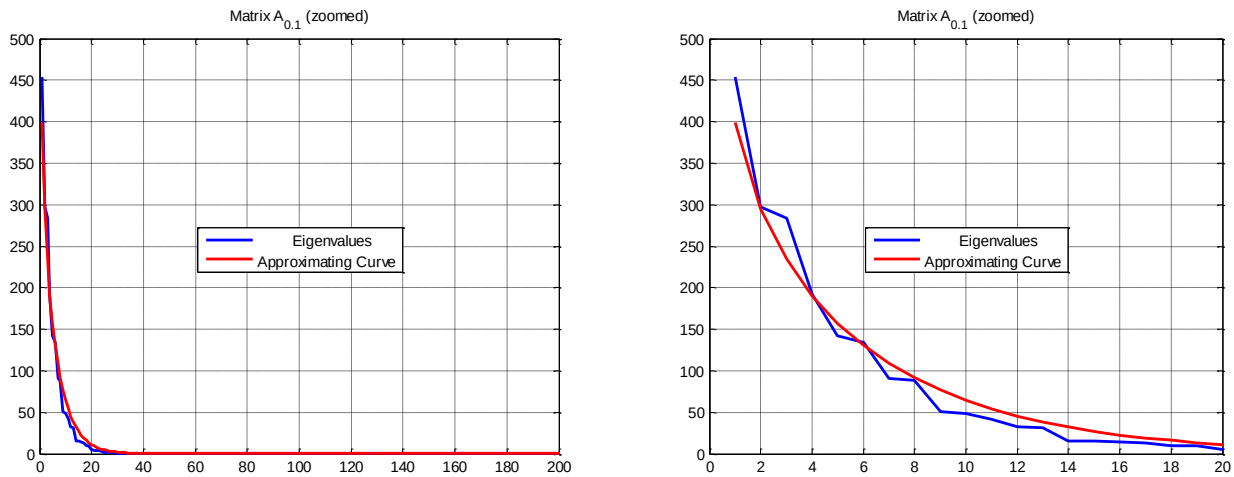

Fig. 3: The results for the matrix $\mathbf{A}_{0.1}$ (the graph on the left is limited to 200 eigenvalues and the graph on the right is limited to 20 eigenvalues)

The fourth analysed situation, which has a different origin, refers to a generic matrix **B** that is derived from a similarity transformation; this transformation involves a diagonal matrix **D** with uniform random entries in the range (0,100), an arbitrary random matrix **P** and its inverse $\mathbf{P}^{-1}$:

$$\mathbf{B} = \mathbf{P}^{-1} \bullet \mathbf{D} \bullet \mathbf{P}; \tag{6}$$

Although **B** is not symmetric, it possesses positive eigenvalues (the same diagonal elements of **D**); as usual, the order is fixed to (2000×2000).
In this case (see fig. 4), we note substantial differences in the first eigenvalues, which often resulted in our analyses being more difficult to perform.
Analogous effect is known in close contexts [6, §1] when n (the sample size) and p (the dimensions of the defining space) are both large, $n/p \to \theta$, θ finite (the so-called *large n large p* case).
Table 2 reports the numerical results that pertain to **B** (the first row), with the same meanings as table 1.
Apart from the values of skewness, namely Sk_Λ and Sk_Z, the other two moments Λ_m and σ^2_Λ are in good agreement with their counterparts Z_m and σ^2_Z. In table 2, we also note the following large deviation between the largest eigenvalue λ_{max} of **B** and the first value of the approximating curve AC(1): 99.94 versus 139.33 (+40.4%). This deviation is a result of the poor representation of the first eigenvalues.
In effect, the good practice of invariably estimating the largest eigenvalue by means of the power method (see section 2) is recommendable. At the cost of an acceptable amount of extra computing time, this estimation furnishes a value to compare with AC(1), and if the deviation is large, a local reconsideration or correction of the approximating curve can be attempted.

Table 2: The characteristic values for the matrices **B** and **C**

	α_s	β_s	ν_s	Λ_m	Z_m	σ^2_Λ	σ^2_Z	Sk_Λ	Sk_Z	λ_{max}	AC(1)
B	0.2643	6.40e+8	0.2246	50.36	50.36	836.26	836.26	-0.04	0.28	99.94	139.33
C	0.4430	7.29e+8	0.2196	60.11	60.11	609.50	609.50	-0.002	-0.002	94.97	132.34

In our case, at the cost of 20 additional iterations, the power method yields the value of 94.95. When compared to the above value of AC(1), the noticeable deviation indicates that we should reconsider the leftmost region of the approximating curve.
It is worthwhile to mention fig. 5, where the frequency distributions of the eigenvalues of **B** and of its approximating curve are presented. Whereas the former distribution shows in its domain a flat erratic profile (consequently, there is a low degree of departure from symmetry and a small skewness, see table 2), the latter distribution shows a unimodal profile, with a mode close to its mean value of $Z_m = 50.36$, with positive skewness and with a regular structure that is derived from both statistical modeling and Monte Carlo averaging. In spite of these differences, the first two central moments are practically undistinguishable; see table 2.

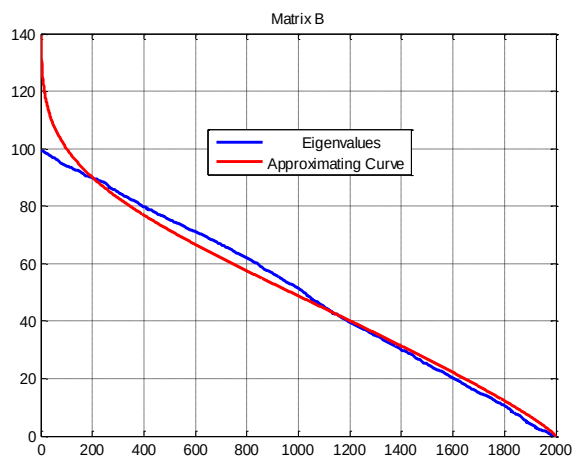

Fig. 4: The results for the matrix **B**

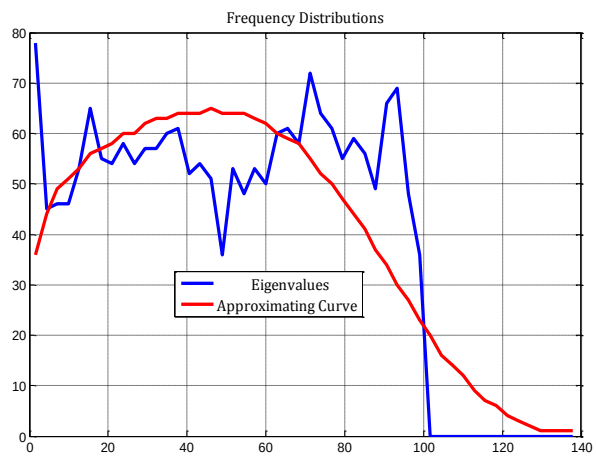

Fig. 5: Frequency Distributions

A final analysis that utilises another PSD matrix, denoted **C** and generated by the same rule (6), is presented in fig. 6 and fig. 7. This analysis shows analogous results; the random eigenvalues were in the range (0, 100), they were clustered uniformly around three central values (30, 60, 90), and they were chosen as entries of the diagonal matrix **D** of (6). Interestingly, the first three moment results perfectly matched (see the second row of table 2), but the differences in figures 6 and 7 are noticeable.

It appears that power raising in our statistical model (2) is capable of explaining the nearly exponential decays of the ordered eigenvalues (see figs. 1 to 3). In contrast, linear or piecewise-linear profiles are represented on a smaller scale of approximation (see fig. 4 and fig. 6).

Table 3: The RMSE relative to the five analyses

	$\mathbf{A}_{10}$	$\mathbf{A}_{1}$	$\mathbf{A}_{0.1}$	**B**	**C**
RMSE	0.0741	0.2248	2.0005	4.2757	7.7470
λ_{max}	10.01	61.78	453.75	99.94	94.97

Finally, to allow us a more quantitative interpretation of the obtained approximations, the RMSE (Root Mean Square Error) is considered; it is calculated in our analyses by means of its well-known definition:

$$\text{RMSE} = \sqrt{1/2000 \bullet \sum(\lambda_k - \widehat{\lambda_k})^2}, \text{ for } k = 1, 2, \ldots, 2000.$$

The RMSE furnishes an average measure of deviation of the true series of eigenvalues $\{\lambda_k\}$ from the approximating series $\{\widehat{\lambda_k}\}$.

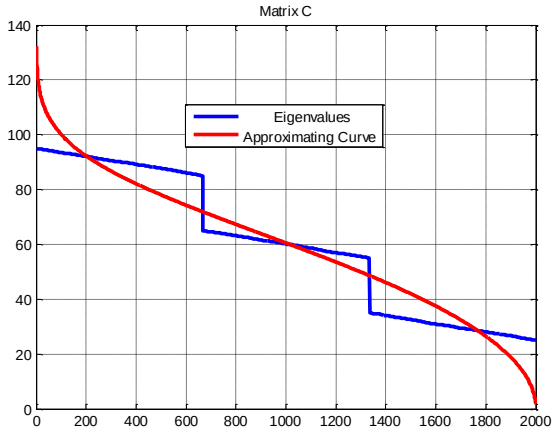

Fig. 6: The results for the matrix **C**

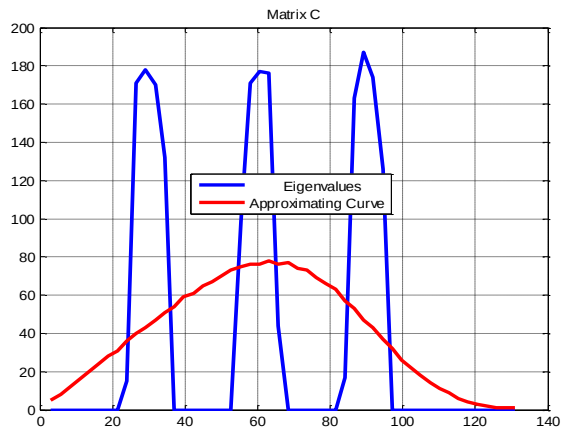

Fig. 7: Frequency Distributions

Table 3 shows the RMSE values; the second row reports the respective largest eigenvalues to have a term of comparison with this scale-dependent parameter.
The first three cases show evidence of small deviations, and these deviations are small either absolutely or relatively.

6 Conclusion

A method to trace the approximating curve of PSD matrix eigenvalues was proposed. This method relies on the solution of a nonlinear system of equations that implements the moment-matching method to fit a powered gamma distribution, and we assume that this distribution is able to conveniently represent the eigenvalue distribution. The subsequent step of Monte Carlo simulations is very simple to perform, and it implements the approximation of the i-th eigenvalue as an average of the i-th order statistics, which are relative to repeated extractions of powered gamma random samples.
The method is well suited to large matrices, when storage or numerical problems can arise in the course of attempts to calculate their spectra by algebraic procedures.
In our case, the assumed statistical distribution, namely, the powered gamma distribution, depends on three parameters, and it complies with the property that the first three sample central moments are calculable by means of in-place procedures. Other distributions exist that, in principle, could produce suitable fitting because of their flexibility, e.g., the four-parameter generalised λ distribution [7]. However, this alternative distribution excludes in-place procedures of computation because it requires more information, i.e., higher-order moments.
The proposed method showed its validity in situations that involve random matrices that possess nearly exponential and/or well-behaved ordered eigenvalue profiles; when this method is applied to inappropriate situations, it still furnishes informative results, but on a smaller scale of approximation.

References

[1] Y. Saad, Numerical Methods for Large Eigenvalue Problems, SIAM (2011); available online: http://www-users.cs.umn.edu/~saad/eig_book_2ndEd.pdf
[2] B.C. Arnold, N. Balakrishnan, H.N. Nagaraja, A First Course in Order Statistics, SIAM (2008)
[3] F.R. Gantmacher, The Theory of Matrices, Chelsea Publ. Company (1959), page 84, vol. I
[4] E. Floyd, Handbook of Applicable Mathematics, vol II: Probability, John Wiley & S. (1980)
[5] N. Cristianini, J. Shawe-Taylor, An Introduction to Support Vector Machines, Cambridge University Press (2000), pp. 42-44
[6] N. El Karoui, Spectrum Estimation for Large Dimensional Covariance Matrices Using Random Matrix Theory, Ann. Statist. (2008), Volume 36, Number 6, pp. 2757-2790
[7] J.S. Ramberg, E.J. Dudewicz, P.R. Tadikamalla, E.F. Mykytka, A probability distribution and its uses in fitting data, Technometrics (1979), vol. 21, No. 2, pp. 201-214.