Eigenvalue Spectra of Random Matrices for Neural Networks

Kanaka Rajan and L. F. Abbott

Center for Neurobiology and Behavior, Columbia University,, College of Physicians and Surgeons, New York, New York 10032, USA

(Received 18 July 2006; published 2 November 2006)

The dynamics of neural networks is influenced strongly by the spectrum of eigenvalues of the matrix describing their synaptic connectivity. In large networks, elements of the synaptic connectivity matrix can be chosen randomly from appropriate distributions, making results from random matrix theory highly relevant. Unfortunately, classic results on the eigenvalue spectra of random matrices do not apply to synaptic connectivity matrices because of the constraint that individual neurons are either excitatory or inhibitory. Therefore, we compute eigenvalue spectra of large random matrices with excitatory and inhibitory columns drawn from distributions with different means and equal or different variances.

DOI: 10.1103/PhysRevLett.97.188104 PACS numbers: 87.18.Sn, 02.10.Yn, 05.90.+m, 87.19.La

Knowledge of the statistical properties of eigenvalues of large random matrices has proven valuable in a wide range of applications [1,2]. In neuroscience, networks of neurons are often studied using models in which interconnections are represented by a synaptic matrix with elements drawn randomly [3,4]. The distribution of eigenvalues of this matrix is useful for studying spontaneous activity and evoked responses in such models [3–7]. For example, the existence of spontaneous activity depends on whether the real parts of any of the eigenvalues are large enough to destabilize the silent state in a linear analysis, and the spectrum of eigenvalues with large real parts provides strong clues about the nature of the spontaneous activity in the full, nonlinear models. A classic result in random matrix theory is Girko's circle law [8], which states that, for large N, the eigenvalues of an $N \times N$ asymmetric random matrix lie uniformly within the unit circle in the complex plane, if the elements are chosen from a distribution with zero mean and variance $1/N$. When partial symmetry is included, the circle changes to an ellipse [9].

Unfortunately, these results do not apply to the synaptic matrices used in realistic neural network models. Neurons are either excitatory or inhibitory, which means that the synapses they make on their targets are all of one type or the other. Thus, the elements of the synaptic matrix must be drawn from two distributions with different means and perhaps different variances. Furthermore, each column of the matrix must have all excitatory (positive) or all inhibitory (negative) elements. Our goal is to determine the eigenvalue spectra of such matrices.

We construct a random synaptic matrix by choosing the elements of fN ''excitatory'' columns from an excitatory distribution and the elements of the remaining $(1-f)N$ ''inhibitory'' columns from an inhibitory distribution. Initially, we consider distributions for the excitatory and inhibitory columns with different means ($\mu_E/\sqrt{N} > 0$ for excitatory and $\mu_I/\sqrt{N} < 0$ for inhibitory) but the same variance, $1/N$. We primarily study a ''balanced'' situation [10,11] in which the average of the combined excitatory and inhibitory distributions is 0, i.e., $f\mu_E + (1-f)\mu_I = 0$.

The result of numerically calculating the eigenvalues of such a matrix is shown in Fig. 1(a). Most of the eigenvalues lie within the unit circle, but there are a number of outliers. The location of these outlying eigenvalues varies from matrix to matrix, and their number does not appear to go to zero as N increases. This makes it difficult to study them analytically. Interestingly, the circle shown in Fig. 1(a) that contains most of the eigenvalues has unit radius. If each element of the matrix were chosen to be either excitatory or inhibitory and then assigned a value from the appropriate distribution, the eigenvalues of the synaptic strength matrix would obey a circle law, but the radius of the circle would be $\sqrt{1 + f\mu_E^2 + (1-f)\mu_I^2}$, rather than 1. Thus, the columnwise assignment of distributions has a dramatic effect on the distribution of eigenvalues.

It is possible to remove the outliers in Fig. 1(a) [see Fig. 1(b)] by imposing a constraint that we now derive. When the variances of the excitatory and inhibitory distributions are equal, we can write our $N \times N$ synaptic matrix as $J + M$, where the elements of J obey $\langle J_{ij} \rangle = 0$ and $\langle J_{ij}^2 \rangle = 1/N$, with the angle brackets representing an average over the distribution from which these elements are

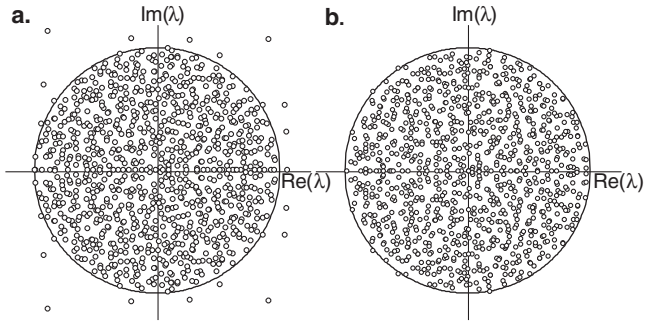

FIG. 1. Numerical results for the distribution of eigenvalues in the complex plane for $N = 1000$. (a) If excitatory and inhibitory elements are drawn from distributions with different means but the same variance, a few eigenvalues lie outside the unit circle. (b) When Eq. (2) is imposed, the eigenvalues lie inside the unit circle.

drawn. Thus, the eigenvalues of J lie uniformly inside the unit circle in the complex plane [8].

M is a matrix in which every row is identical and is equal to $1/\sqrt{N}$ times a vector m that has fN elements equal to μ_E and $(1-f)N$ elements equal to μ_I. The result of multiplying any vector v by M is

$$Mv = \frac{1}{\sqrt{N}}(m \cdot v)u \quad \text{where } u_i = 1 \quad \text{for } i = 1, \ldots, N. \tag{1}$$

All of the eigenvalues of M are zero and, in particular, the vector u is a right eigenvector of M with eigenvalue $m \cdot u/\sqrt{N} = \sqrt{N}[f\mu_E + (1-f)\mu_I] = 0$. Note that the elements of M are of the same order of magnitude $(1/\sqrt{N})$ as the elements of J, so the presence of the outliers in Fig. 1(a) is not surprising.

To force the outliers inside the unit circle, we impose the condition

$$Ju = 0, \tag{2}$$

which implies that the strengths of the synapses to each neuron (corresponding to one row of J) independently sum to zero [12]. This corresponds to a balance condition not only on the average synaptic strength but also on the specific sum for each neuron [10,11].

Equation (2) implies that u is a right eigenvector of J. If we denote the a^{th} left eigenvector of J by L_a, this means that $L_a \cdot u = 0$ for all the other eigenvectors of J. Thus, according to Eq. (1), $L_a M = 0$ for all these eigenvectors, so they are left eigenvectors of both J and $J + M$ with the same eigenvalues. Furthermore, because the vector u is a right eigenvector of J with eigenvalue 0, it is a right eigenvector of $J + M$ with eigenvalue $\sqrt{N}[f\mu_E + (1-f)\mu_I] = 0$. Thus, the eigenvalues of J and $J + M$ are identical, and *shifting the means of the excitatory and inhibitory distributions away from zero in a balanced manner has no effect on the eigenvalues, provided that Eq. (2) is satisfied*. This is illustrated in Fig. 1(b), which shows the eigenvalue distribution computed numerically for a matrix constructed as in Fig. 1(a), but with the balance condition imposed row by row by subtracting the same constant from each element in the row.

Although the balance constraint of Eq. (2) has a significant effect on the interaction between J and M, it does not change the distribution of eigenvalues [13] of J, for large N. This is because J acts as a random $(N-1) \times (N-1)$ matrix in the subspace spanned by the left eigenvectors of J orthogonal to u. It thus has $N-1$ eigenvectors distributed uniformly within a circle in the complex plane of radius $\sqrt{(N-1)/N}$, which goes to 1 as $N \to \infty$. The additional eigenvalue of J is the zero eigenvalue from Eq. (2).

In addition to having different means, the distributions for excitatory and inhibitory synapses are likely to have different variances. The inset in Fig. 2(a) shows that in this case, the eigenvalues lie inside a circle, but they are no longer distributed uniformly. We now consider this novel effect analytically.

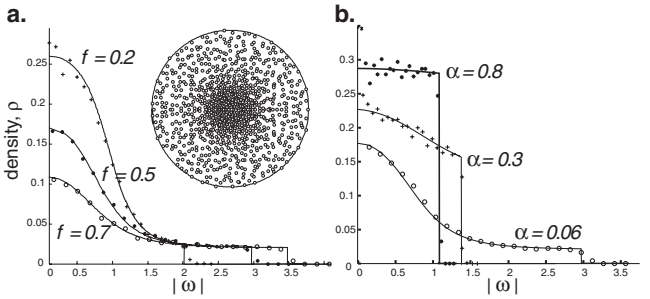

FIG. 2. Density ρ of eigenvalues as a function of radius in the complex plane $|\omega|$, for $N = 1000$. The solid lines are the result of the analytic calculation and symbols are numerical results. (a) Results for different fractions f of excitatory and inhibitory elements, and $\alpha = 0.06$. The inset shows eigenvalues in the complex plane computed numerically for $f = 0.5$ and $\alpha = 0.06$. (b) Results for different excitatory variances $1/(N\alpha)$ and $f = 0.5$, with the inhibitory variance equal to $1/N$.

According to the above results, the eigenvalues of $J + M$ are identical to those of J provided Eq. (2) is satisfied. Therefore, we can compute the eigenvalues of J rather than $J + M$ (except that we will include different variances for different columns). In addition, imposing the constraint of Eq. (2) does not change the distribution of eigenvalues of J, for large N. These two observations imply that we can compute the eigenvalue spectrum we seek using distributions with different variances but 0 means, without having to impose Eq. (2).

We now proceed to calculate the density of eigenvalues ρ in the complex plane for such a matrix, averaged over the underlying probability distributions. Our calculation follows the approach used in Ref. [9], which begins by considering the quantity $G(\omega) = \text{Tr}[1/(\omega I - J)]$, which is also equal to $\int d^2\lambda \rho/(\omega - \lambda)$, where I is the identity matrix, ω is a complex variable, and the integral in the second expression is over complex λ. As shown in Ref. [9], for example, the real and complex parts of $G(\omega)$ act like the x and y components of an electric field in two dimensions produced by a charge density given by ρ. As a result, ρ can be related to a potential ϕ through Poisson's equation. Finally, the potential can be written in terms of a determinant of a square of the operator $\omega I - J$ which can, in turn, be expressed in terms of a Gaussian integral [9] (see below).

We consider cases in which the density and potential are functions of $|\omega|^2$ so that

$$\rho = \frac{1}{\pi}(|\omega|^2\phi'' + \phi'), \tag{3}$$

where the primes denote derivatives with respect to $|\omega|^2$. We construct a matrix with different variances for different columns by expressing its elements as $J_{ij}\sigma_j$ with the elements of J drawn from a Gaussian distribution with

zero mean and variance $1/N$. This matrix satisfies $\langle (J_{ij}\sigma_j)^2\rangle = \sigma_j^2/N$. The potential we need is determined by

$$\exp(-N\phi) = \int \prod_{i=1}^{N} \frac{dz_i^* dz_i}{\pi} \exp(-N\tilde{Q}), \tag{4}$$

with the average over the distribution of J values given by

$$\exp(-N\tilde{Q}) = \int \prod_{i,j=1}^{N} \sqrt{\frac{N}{2\pi}} dJ_{ij} \exp(-NQ). \tag{5}$$

In our case,

$$Q = \sum_i \left(\frac{|\omega|^2}{\sigma_i^2} + \epsilon_i\right)\frac{z_i^* z_i}{N} + \frac{1}{2}\sum_{i,j,k} J_{ki}A_{ij}J_{kj} - \sum_{k,j} B_{kj}J_{kj},$$
$$A_{ij} = \frac{z_i^* z_j}{N} + \frac{z_j^* z_i}{N} + \delta_{ij}, \qquad B_{kj} = \frac{\omega^* z_k^* z_j}{\sigma_k N} + \frac{\omega z_j^* z_k}{\sigma_k N} \tag{6}$$

and the term involving ϵ_i is introduced to assure convergence of the integrals, with the understanding that all $\epsilon_i \to 0$ from the positive side at the end of the calculation.

To order $1/\sqrt{N}$, the matrix A has all but two of its eigenvalues equal to one. The two exceptions are the eigenvectors z_i and z_i^*, both with eigenvalue $1 + r$. This observation allows us to compute the Gaussian integral over J in Eq. (5) by completing the square [shifting J by $J_{ij} \to J_{ij} - B_{ij}/(1 + r)$] and calculating the determinant of A. The result is

$$\tilde{Q} = \ln(1 + r) + \frac{|\omega|^2\tilde{r}}{1 + r} + r_\epsilon, \quad \text{where } r = \frac{1}{N}\sum_i z_i^* z_i,$$
$$\tilde{r} = \frac{1}{N}\sum_i \frac{z_i^* z_i}{\sigma_i^2} \quad \text{and } r_\epsilon = \frac{1}{N}\sum_i \epsilon_i z_i^* z_i. \tag{7}$$

The potential ϕ is then determined by computing the integrals over z and z^* by saddle-point approximation [14].

In the case of excitatory and inhibitory neurons, there are two variances. Without loss of generality, we choose fN columns from a distribution with variance $\sigma_E^2 = 1/(N\alpha)$ and $(1 - f)N$ columns from a distribution with variance $\sigma_I^2 = 1/N$. In this case, we define r_1 as $1/(N\alpha)$ times the sum of $|z_i|^2$ over all i values corresponding to excitatory columns, and r_2 as $1/N$ times the sum of $|z_i|^2$ over all inhibitory i values. Then,

$$\phi = \ln\left(\frac{1 + r_1 + r_2}{(r_1)^f (r_2)^{1-f}}\right) + \frac{|\omega|^2(\alpha r_1 + r_2)}{1 + r_1 + r_2} + \epsilon_1 r_1 + \epsilon_2 r_2,$$

where this expression is to be evaluated at values of r_1 and r_2 that minimize ϕ. The extra factors in the logarithm in ϕ as compared to $\tilde{Q}$ come from expressing the integrals over z and z^* in spherical coordinates.

The equations $\partial\phi/\partial r_1 = \partial\phi/\partial r_2 = 0$ that determine r_1 and r_2 are

$$-\epsilon_1 = \frac{1 + \alpha|\omega|^2}{1 + r_1 + r_2} - \frac{f}{r_1} - \frac{|\omega|^2(\alpha r_1 + r_2)}{(1 + r_1 + r_2)^2}$$
$$\text{and} \quad -\epsilon_2 = \frac{1 + |\omega|^2}{1 + r_1 + r_2} - \frac{1 - f}{r_2} - \frac{|\omega|^2(\alpha r_1 + r_2)}{(1 + r_1 + r_2)^2}. \tag{8}$$

Here the convergence factors have reduced to two, ϵ_1 and ϵ_2, and these equations must be solved in the limit where these both go to zero from the positive side.

Equations (8) have two solutions depending on whether $|\omega|^2$ is greater than (outside solution) or less than (inside solution) a critical value $1 - f + f/\alpha$. It is easiest to express the solutions by writing $r_1 = qr_2$. Then, outside the circle

$$r_2 = \frac{\alpha(1 - f)}{(|\omega|^2 - 1)\alpha + (\alpha - 1)f} \quad \text{and} \quad q = \frac{f}{\alpha(1 - f)}, \tag{9}$$

which gives the potential

$$\phi = 1 + \ln\left(\frac{\alpha^f}{(f)^f(1 - f)^{1-f}}\right) + \ln(|\omega|^2), \tag{10}$$

and a zero density, $\rho = 0$. Inside the circle,

$$r_2 \to \infty \quad \text{and}$$
$$q = \frac{(1 - \alpha)|\omega|^2 + 2f - 1}{2(1 - f)} + \frac{\sqrt{[(1 - \alpha)|\omega|^2 - 1]^2 + 4f(1 - \alpha)|\omega|^2}}{2(1 - f)} \tag{11}$$

and

$$\phi = \ln\left(\frac{1 + q}{q^f}\right) + \frac{|\omega|^2(q\alpha + 1)}{q + 1}. \tag{12}$$

The density of eigenvalues, ρ, is obtained from this using Eq. (3) and

$$\phi' = q'\left(\frac{1}{q+1} - \frac{f}{q} + \frac{\alpha|\omega|^2}{q+1} - \frac{|\omega|^2(\alpha q + 1)}{(q+1)^2}\right) + \frac{\alpha q + 1}{q + 1}$$
$$\text{and} \quad \phi'' = q''\left(\frac{1}{q+1} - \frac{f}{q} + \frac{\alpha|\omega|^2}{q+1} - \frac{|\omega|^2(\alpha q + 1)}{(q+1)^2}\right) + 2q'\left(\frac{\alpha}{q+1} - \frac{\alpha q + 1}{(q+1)^2}\right) + (q')^2\left(\frac{-1}{(q+1)^2} + \frac{f}{q^2} - \frac{2\alpha|\omega|^2}{(q+1)^2} + \frac{2|\omega|^2(\alpha q + 1)}{(q+1)^3}\right) \tag{13}$$

where, from Eq. (11),

$$q' = \frac{1-\alpha}{2(1-f)} \times \left(1 + \frac{(1-\alpha)|\omega|^2 - 1 + 2f}{\sqrt{[(1-\alpha)|\omega|^2 - 1]^2 + 4f(1-\alpha)|\omega|^2}}\right)$$

$$\text{and } q'' = \frac{(1-\alpha)^2}{2(1-f)\sqrt{[(1-\alpha)|\omega|^2 - 1]^2 + 4f(1-\alpha)|\omega|^2}} \times \left(1 - \frac{[(1-\alpha)|\omega|^2 - 1 + 2f]^2}{[(1-\alpha)|\omega|^2 - 1]^2 + 4f(1-\alpha)|\omega|^2}\right). \quad (14)$$

Figure 2 compares the analytic expression we have computed for the density of eigenvalues with numerical results for $N = 1000$. The numerical densities were calculated by counting the number of eigenvalues in successive concentric rings. All of the eigenvalues are located inside a circle of radius $\sqrt{1 - f + f/\alpha}$ because $\rho = 0$ for the outside solution. Note that, since we are considering the case $N\sigma_E^2 = 1/\alpha$ and $N\sigma_I^2 = 1$, the square of this radius can also be written as $N[f\sigma_E^2 + (1-f)\sigma_I^2]$. *Thus, the square of the radius of the circle containing the eigenvalues is N times the average of the variances of the excitatory and inhibitory distributions.* The distributions are characterized by a high-density central region with a radius given approximately by $\min(\sigma_E, \sigma_I)\sqrt{N}$, which is equal to 1 in the cases shown in Fig. 2.

Figure 2(a) shows results with $\alpha = 0.06$ for different values of f. The value of f determines how many eigenvalues fall into the high-density central region, with higher central densities for smaller f values. There are two limits in which our results reduce to the usual circle law, $f \to 0$ and $f \to 1$. When $f = 0$, all the elements of the matrix come from a single distribution with variance $1/N$, and the eigenvalues all fall uniformly inside the inner circle of radius 1. When $f = 1$, all the elements come from a distribution with variance $1/(N\alpha)$, and all the eigenvalues lie uniformly inside a circle of radius $\sqrt{1/\alpha}$.

Figure 2(b) shows results with $f = 0.5$ for different values of α. We only consider $\alpha \leq 1$ because $\alpha > 1$ can be reduced to this case by rescaling the matrix and changing $f \to 1 - f$. When $\alpha = 1$, both of the distributions have the same variance, $\sigma_E^2 = \sigma_I^2 = 1/N$, and the density is uniform inside the unit circle.

Our results were obtained using Gaussian distributions, but they apply more generally. For example, they apply even though synaptic strength distributions are non-Gaussian [15] and include a zero-strength δ function due to the sparseness of neuronal connectivity. Neurons do not normally form synaptic connections with themselves, but we did not eliminate diagonal terms in the synaptic matrix for our calculations because their effect is negligible for large N. Finally, if Eq. (2) is imposed but the balance condition on the means is not satisfied, the eigenvalues will be identical to what we have computed except that the eigenvalue at zero will be shifted to $\sqrt{N}[f\mu_E + (1-f)\mu_I]$.

The eigenvalue distributions we have obtained have several implications for neural network dynamics. First and most surprisingly, modifying the mean strengths of excitatory and inhibitory synapses has no effect on stability or small-fluctuation dynamics under balanced conditions. Instead, the key elements in determining the spontaneous dynamics of networks constructed in this way are the widths of the distributions of excitatory and inhibitory synaptic strengths. Furthermore, if these widths are different, fewer eigenvalues will appear at the edge of the eigenvalue circle, meaning that there will be fewer slowly oscillating and long-lasting modes. In summary, the calculations we present suggest that having different cell types with different distributions of synaptic strengths has a large impact on network dynamics, and that the critical element to measure, and the critical element that may be modified by the modulatory and plasticity mechanisms that control neural circuit dynamics, are the variances of the synaptic strength distributions.

We thank Bill Bialek, Michael Eisele, Edward Farhi, Vikram Rajan, Evan Schaffer, and Haim Sompolinksy for helpful comments and suggestions. Research supported by the Swartz Foundation and NIH (5-DP1-OD114-02).

[1] M. L. Mehta, *Random Matrices and the Statistical Theory of Energy Levels* (Academic, New York, 1967).
[2] T. A. Brody, J. Flores, J. B. French, P. A. Mello, A. Pandey, and S. S. M. Wong, Rev. Mod. Phys. **53**, 385 (1981).
[3] H. Sompolinsky, A. Crisanti, and H. J. Sommers, Phys. Rev. Lett. **61**, 259 (1988).
[4] B. Cessac and J. A. Sepulchre, Chaos **16**, 013104 (2006).
[5] H. R. Wilson and J. D. Cowan, Biophys. J. **12**, 1, (1972).
[6] O. Shriki, D. Hansel, and H. Sompolinsky, Neural Comput. **15**, 1809 (2003).
[7] Additional references in: T. P. Vogels, K. Rajan, and L. F. Abbott, Annu. Rev. Neurosci. **28**, 357 (2005).
[8] V. L. Girko, Theory Probab. Appl. **29**, 694 (1984).
[9] H. J. Sommers, A. Crisanti, H. Sompolinsky, and Y. Stein, Phys. Rev. Lett. **60**, 1895 (1988).
[10] M. N. Shadlen and W. T. Newsome, Curr. Opin. Neurobiol. **4**, 569 (1994).
[11] T. W. Troyer and K. D. Miller, Neural Comput. **9**, 971 (1997).
[12] It is actually sufficient to make Ju of order $1/\sqrt{N}$.
[13] We thank Michael Eisele for assistance with this point.
[14] The quadratic integral over fluctuations about the saddle-point is ignored because it produces a term of $O(1/N)$. Note that the saddle-node value of $\tilde{Q}$ we obtain is $O(1)$ as required.
[15] S. Song S., P. J. Sjöström, M. Reigl, S. B. Nelson, and D. B. Chklovskii, PLoS Biol. **3**, 505, (2005).