Notes on the enumeration of RNA secondary structures by matrix models

Vittorio Erba

Last updated: October 24, 2019

Abstract

Modeling and prediction of RNA folding structure is an open and challenging problem in biophysics. In these notes, I will introduce the Matrix Models approach to the problem first defined in [1], and describe in detail the topological characterization of RNA folding structures in the easiest setting possible, as presented in [2]. Finally, I will review the comparison results between this analytical characterization and the observed topology of RNA folding structures.

Contents

1 Motivation: RNA secondary structures

An RNA molecule can be modeled as a chain of L nucleotides. A folding configuration (tipically called *secondary structure* in the literature) of an RNA molecule is determined by a pairing of its nucleotides, so that paired nucleotides lie at short distance one from the other in the embedding 3D space. Formally, a secondary structure can be described as a map $s : [L] \to [L]$ such that $s(i) = j \iff s(j) = i$ for all $1 \leq i, j \leq L$; in this case, the pair of nucleotides (i, j) is matched (notice that there may be unpaired nucleotides such that $s(i) = i$). Here $[L]$ denotes the set of integers from 1 to L.

In nature, the secondary structure of an RNA molecule is determined, at equilibrium, by the interaction between nucleotides and by the characteristics of the backbone, that favor stacking of paired nucleotides and prevent nucleotides that are too near to each other to be paired due to its rigidity. In these notes, we will neglect all backbone considerations, and restrict our attention to the pair-pair interaction.

Define $V_{ij} = \exp(-\beta\epsilon_{ij})$ to be the Gibbs-Boltzmann weight for the interaction ϵ_{ij} between the two nucleotides i, j at inverse temperature β. Then, the partition function is a sum over the

number of pairings of all possible pairing configurations:

$$\begin{aligned} Z_L &= 1 + \sum_{1 \text{ pairing}} V_1 + \sum_{2 \text{ pairings}} V_1 V_2 + \dots \\ &= 1 + \sum_{i<j} V_{ij} + \sum_{i<j<k<l} (V_{ij}V_{kl} + V_{ik}V_{jl} + V_{il}V_{jk}) + \dots \end{aligned} \tag{1}$$

where in the first line V_i denotes the energy of the i-th pairing in the sum.

A graphical representation of this sum is given in Figure 1 of [2], where a RNA chain is represented on the x axis and its paired nucleotides are drawn as arcs between the nucleotides on the upper half-plane. In this representation, one can easily see that different secondary structures are characterized by a different number of crossing between the pairing arcs: if the number of crossings is non-null, then we say that the secondary structure contains a *pseudoknot*.

We would like to characterize pseudoknots by their *topological genus*, i.e. the number of handles that we must glue to a 2-sphere to be able to draw the arc diagram of the pseudoknot without crossings between arcs. For example, 0 crossing diagrams have genus 0, 1 crossing diagrams have genus 1 etc..., but diagrams with a higher number of crossings c can have arbitrary genus g in $1 \leq g \leq c$, as different crossings may benefit from the same handle.

2 Topological expansion for the enumeration of RNA secondary structures

2.1 The partition function as a GUE average

One can recognize that the expression for the partition function in Equation 1 strictly resembles the solution, due to Wick's theorem, of the following integral over L $n \times n$ hermitian matrices ϕ_l:

$$\begin{aligned} Z_L(n) &= \frac{\int \prod_{l=1}^{L} d\phi_l \exp\left[-\frac{n}{2}\sum_{i,j=1}^{L}(V^{-1})_{ij}\operatorname{Tr}(\phi_i\phi_j)\right] \frac{1}{n}\operatorname{Tr}\left(\prod_{l=1}^{L}(1+\phi_l)\right)}{\int \prod_{l=1^L} d\phi_l \exp\left[-\frac{n}{2}\sum_{i,j=1}^{L}(V^{-1})_{ij}\operatorname{Tr}(\phi_i\phi_j)\right]} \\ &= \left\langle \frac{1}{n}\operatorname{Tr}\left(\prod_{l=1}^{L}(1+\phi_l)\right)\right\rangle \end{aligned} \tag{2}$$

where brackets denote the average over the matrix measure in question, that is a L dimensional, coupled GUE measure with covariance matrix $\frac{n}{2}(V^{-1})_{ij}$.

Let's check this claim by explicitly applying Wick's theorem. First, let's identify better the observable we would like to average:

$$\begin{aligned} &\left\langle \frac{1}{n}\operatorname{Tr}\left(\prod_{l=1}^{L}(1+\phi_l)\right)\right\rangle \\ &= n^{-1}\left\langle \operatorname{Tr}(\mathbb{I}) + \sum_i \operatorname{Tr}(\phi_i) + \sum_{i<j}\operatorname{Tr}(\phi_i\phi_j) + \dots + \sum_{i_1<\dots<i_L}\operatorname{Tr}(\phi_{i_1}\dots\phi_{i_L})\right\rangle \\ &= 1 + \sum_{i<j}\langle\operatorname{Tr}(\phi_i\phi_j)\rangle + \sum_{i<j<k<l}\langle\operatorname{Tr}(\phi_i\phi_j\phi_k\phi_l)\rangle + \dots \end{aligned} \tag{3}$$

where all terms with an odd number of matrices are null as we are considering a GUE ensamble with null mean.

The propagator is (latin letters run in $[L]$ over different matrices, greek letters run in $[n]$ over the elements of a single matrix)

$$\langle \phi_{a,\alpha\beta}\phi_{b,\gamma\delta}\rangle = \sum_{i,j} O^T_{ai}O_{jb}\langle\psi_{i,\alpha\beta}\psi_{j,\gamma\delta}\rangle \tag{4}$$

where O is the rotation that diagonalizes the covariance matrix, i.e.

$$\begin{aligned}(V^{-1})_{ab} &= \sum_k O^T_{ak}\lambda_k O_{kb} \\ \psi_i &= \sum_k O_{ik}\phi_k\end{aligned} \tag{5}$$

In the ψ coordinates, the L GUE integrals decouple and one obtains

$$\langle \phi_{a,\alpha\beta}\phi_{b,\gamma\delta}\rangle = \sum_{i,j} O^T_{ai}O_{jb}\,\langle \psi_{i,\alpha\beta}\psi_{j,\gamma\delta}\rangle = n^{-1}\sum_i O^T_{ai}O_{ib}\lambda_i^{-1}\delta_{\alpha\gamma}\delta_{\beta\delta} = n^{-1}V_{ij}\delta_{\alpha\gamma}\delta_{\beta\delta}\,. \tag{6}$$

Now, expansion of the traces and application of Wick's theorem give that

$$Z_L(n) = 1 + \sum_{i<j} V_{ij} + \sum_{i<j<k<l}\left(V_{ij}V_{kl} + n^{-2}V_{ik}V_{jl} + V_{il}V_{jk}\right) + \dots \tag{7}$$

so that the $Z_L = Z_L(n=1)$. This procedure not only recovers the wanted partition function, but also introduces a new parameter n that refines the enumeration; in fact, associated to each power g of n^{-2}, one has the secondary structures whose arc diagram has genus g.

Is g really the genus? It is! This matrix theory is analogous to a t'Hooft model, where the N^{-2} expansion selects precisely the genus of each Feynman diagram. A more precise treatment of the parallel between the t'Hooft topological expansion and the topology of arc diagrams is given in [1, 3].

2.2 Hubbard-Stratonovich simplifications

To simplify the problem, assume that $V = a\mathbb{I} + v\mathbb{U}$, where $\mathbb{U}_{ij} = 1$ is the matrix with all entries equal to 1 as done in [2] (the general case is more difficult, and it's treated in [1]).

First of all, one has to ensure that V is a positive definite matrix, so that the GUE measure is well defined. Noticing that V is a RS (Replica Symmetric) matrix, one can compute the determinant (see Appendix C):

$$\det V = (a + Lv)a^{L-1} \tag{8}$$

so that for any v, $a > 0$ ensures positive definitness. Notice that no diagonal term of V is present in the expression for $Z_L(n)$, so the result is independent from a. To sum it up, for any v, let's choose a random positive a: this grants the convergence of our integrals, and does not alter our partition function.

In the following, we will need the inverse of this matrix. It has the same structure $V^{-1} = c\mathbb{I} - d\mathbb{U}$ with $c = a^{-1}$,$d = \frac{cv^2}{Lvc+1}$, and the minus sign grants that both coefficients can be taken positive.

Under this simplifying assumptions, one can rewrite the matrix integrals using an Hubbard-Stratonovich (H-S) transformation. First, rewrite the exponent of the GUE measure by explicitly using the structure of V_{ij}

$$\begin{aligned}\sum_{i,j}\mathrm{Tr}\,(\phi_i\phi_j)\,(V^{-1})_{ij} &= (c-d)\sum_i \mathrm{Tr}\left(\phi_i^2\right) - d\sum_{i\neq j}\mathrm{Tr}\,(\phi_i\phi_j) \\ &= c\sum_i \mathrm{Tr}\left(\phi_i^2\right) - d\,\mathrm{Tr}\left(\left(\sum_i \phi_i\right)^2\right),\end{aligned} \tag{9}$$

so that the second term can be expanded using an H-S transformation. This gives

$$
\begin{aligned}
&Z_L(n) = \\
&= \frac{\int d\tau \prod_{l=1}^{L} d\phi_l \exp\left[-\frac{nc}{2}\sum_{i=1}^{L}\mathrm{Tr}\left(\phi_i^2\right) - \frac{1}{2nd}\mathrm{Tr}\left(\tau^2\right) + \mathrm{Tr}\left(\tau\sum_i \phi_i\right)\right] \frac{1}{n}\mathrm{Tr}\left(\prod_{l=1}^{L}\left(\mathbb{I}+\phi_l\right)\right)}{\int d\tau \prod_{l=1}^{L} d\phi_l \exp\left[-\frac{nc}{2}\sum_{i=1}^{L}\mathrm{Tr}\left(\phi_i^2\right) - \frac{1}{2nd}\mathrm{Tr}\left(\tau^2\right) + \mathrm{Tr}\left(\tau\sum_i \phi_i\right)\right]} \\
&= \frac{\int d\tau \exp\left[-\frac{\mathrm{Tr}\left(\tau^2\right)}{2nd}\right] \frac{1}{n}\mathrm{Tr}\left(\prod_{i=1}^{L}\int d\phi_l \exp\left[-\frac{nc}{2}\mathrm{Tr}\left(\phi_l^2\right) + \mathrm{Tr}\left(\tau\phi_l\right)\right]\left(\mathbb{I}+\phi_l\right)\right)}{\int d\tau \exp\left[-\frac{\mathrm{Tr}(\tau^2)}{2nd}\right]\prod_{i=1}^{L}\int d\phi_l \exp\left[-\frac{nc}{2}\mathrm{Tr}\left(\phi_l^2\right) + \mathrm{Tr}\left(\tau\phi_l\right)\right]} \\
&= \frac{\int d\tau \exp\left[-\frac{\mathrm{Tr}\left(\tau^2\right)}{2nd} + \frac{L\,\mathrm{Tr}\left(\tau^2\right)}{2nc}\right] \frac{1}{n}\mathrm{Tr}\left(\left(\mathbb{I}+\frac{\tau}{nc}\right)^L\right)}{\int d\tau \exp\left[-\frac{\mathrm{Tr}(\tau^2)}{2nd} + \frac{L\,\mathrm{Tr}(\tau^2)}{2nc}\right]} \\
&= \frac{\int d\sigma \exp\left[-\frac{n}{2v}\mathrm{Tr}\left(\sigma^2\right)\right] \frac{1}{n}\mathrm{Tr}\left(\left(\mathbb{I}+\sigma\right)^L\right)}{\int d\sigma \exp\left[-\frac{n}{2v}\mathrm{Tr}\left(\sigma^2\right)\right]}
\end{aligned}
\tag{10}
$$

where in the last line a change of variable was performed ($\tau = \sigma nc$), and the relation between L, v, c, d was exploited to simplify the variance. The normalization factor (denominator) $\tilde{A}(n)$ is an easy Gaussian integral that evaluates to

$$
\tilde{A}(n) = \sqrt{2}^{\,n}\left(\sqrt{\frac{\pi v}{n}}\right)^{n^2}. \tag{11}
$$

2.3 The spectral density of GUE

Equation 10 can be further simplified by expanding the power using the Newton binomial and expressing the GUE integral in the eigenvalue measure:

$$
\begin{aligned}
Z_L(n) &= \frac{\int d\sigma \exp\left[-\frac{n}{2v}\mathrm{Tr}\left(\sigma^2\right)\right] \frac{1}{n}\mathrm{Tr}\left(\left(\mathbb{I}+\sigma\right)^L\right)}{\int d\sigma \exp\left[-\frac{n}{2v}\mathrm{Tr}\left(\sigma^2\right)\right]} \\
&= \frac{\sum_{k=0}^{L}\binom{L}{k}\int d\sigma \exp\left[-\frac{n}{2v}\mathrm{Tr}\left(\sigma^2\right)\right] \frac{1}{n}\mathrm{Tr}\left(\sigma^L\right)}{\int d\sigma \exp\left[-\frac{n}{2v}\mathrm{Tr}\left(\sigma^2\right)\right]} \\
&= \frac{\sum_{k=0}^{L}\binom{L}{k}\int \prod_{i=1}^{n} d\sigma_i \Delta(\sigma)^2 \exp\left[-\frac{n}{2v}\sum_i \sigma_i^2\right] \frac{1}{n}\sum_i \sigma_i^L}{\int \prod_{i=1}^{n} d\sigma_i \Delta(\sigma)^2 \exp\left[-\frac{n}{2v}\sum_i \sigma_i^2\right]}
\end{aligned}
\tag{12}
$$

where $\{\sigma_i\}$'s are the eigenvalues of the matrix σ and $\Delta(\sigma)$ is the Vandermonde determinant of the eigenvalues. Finally, one has:

$$
\begin{aligned}
Z_L(n) &= \frac{\sum_{k=0}^{L}\binom{L}{k}\int \prod_{i=1}^{n} d\sigma_i \Delta(\sigma)^2 \exp\left[-\frac{n}{2v}\sum_i \sigma_i^2\right] \frac{1}{n}\sum_i \int d\lambda \delta\left(\lambda-\sigma_i\right)\lambda^L}{\int \prod_{i=1}^{n} d\sigma_i \Delta(\sigma)^2 \exp\left[-\frac{n}{2v}\sum_i \sigma_i^2\right]} \\
&= \int d\lambda (1+\lambda)^L \frac{\int \prod_{i=1}^{n} d\sigma_i \Delta(\sigma)^2 \exp\left[-\frac{n}{2v}\sum_i \sigma_i^2\right] \frac{1}{n}\sum_i \delta\left(\lambda-\sigma_i\right)}{\int \prod_{i=1}^{n} d\sigma_i \Delta(\sigma)^2 \exp\left[-\frac{n}{2v}\sum_i \sigma_i^2\right]} \\
&= \int d\lambda (1+\lambda)^L \rho_n(\lambda)
\end{aligned}
\tag{13}
$$

where $\rho_n(\lambda)$ is the spectral density of the GUE ensamble.

The spectral density can be exactly computed through the orthogonal polynomial technique, as sketched in Appendix B, and we will use it in the form given in [4]:

$$
\rho_n(\lambda) = \frac{e^{-\frac{n\lambda^2}{2v}}}{\sqrt{2\pi n v}} \sum_{k=0}^{n-1} \binom{n}{k+1} \frac{H_{2k}\left(\sqrt{\frac{n}{2v}}\lambda\right)}{2^k k!}. \tag{14}
$$

Moreover, one can analytically sum the exponential generating function of $Z_L(n)$ to obtain

$$G(t,n,v) = \sum_{L\geq 0} Z_L(n)\frac{t^L}{L!} = \int d\lambda \rho_N(\lambda) e^{t(1+\lambda)}\,. \tag{15}$$

So, in conclusion, one has to evaluate the known integral

$$\begin{aligned}\int d\lambda\, H_{2k}\left(\sqrt{\frac{n}{2v}}\lambda\right) e^{-\frac{n}{2v}\lambda^2+t\lambda} &= \sqrt{\frac{2v}{n}}\int dx\, H_{2k}\left(x\right) e^{-x^2+t\sqrt{\frac{2v}{n}}x} \\ &= \sqrt{\frac{2v\pi}{n}}\left(t\sqrt{\frac{2v}{n}}\right)^{2k}\exp\left[\frac{2vt^2}{n}\right]\end{aligned} \tag{16}$$

giving

$$\begin{aligned}G(t,n,v) &= \frac{1}{n}\sum_{k=0}^{n-1}\binom{n}{k+1}\frac{v^k t^{2k}}{k! n^k}\exp\left[\frac{2v}{n}t^2+t\right] \\ &= \exp\left[\frac{2v}{n}t^2+t\right]\frac{1}{n}L_{n-1}^{(1)}\left(-\frac{vt^2}{n}\right)\end{aligned} \tag{17}$$

where we used the definition of the generalized Laguerre polynomial $L_j^{(1)}(z)$.

Now, expanding the function $G(t,n,v)$ in powers of t^L, n^{-2g} and v^k allows to enumerate the secondary structures of and RNA molecule with L nucleotides, k pairings of nucleotides and topological genus g, giving the wanted result.

3 So, what about the genus of RNA sequences?

The expansion in L of $G(t,n,v)$ recovers an expression for $Z_L(n)$

$$\frac{Z_L(n)}{Z_L(1)} = \sum_{g\geq 0} a_{L,g} N^{-2g} \tag{18}$$

where we set $v=1$ as we are not interested in the specific number of pairings, and the normalization $Z_L(1)$ grants that the $a_{L,g}$ are a probability distribution in g.

There are two main points to be made:

- at fixed L, the average genus $\langle g\rangle_L$ scales numerically as $\sim 0.23L$, and the measures is quite concentrated, so that most diagrams are not planar;
- at fixed L, there is a maximum value for the genus, that is $g \leq \frac{L}{4}$.

In [5], the genus distribution in real RNA dataset was studied, finding that even for large values of L, the genus of a sequence typically remains much smaller then the typical values given above.

In [6], the genus distribution is studied numerically by Monte Carlo simulations for a model of self avoiding walk on the 2D and 3D lattices, with short-range attractive interaction. The model is known to have a phase transition between a compact globular phase and a swollen phase which can have very complicated features. The genus distribution is different in the two phases, with typical values much larger in the compact phase then in the swollen phase. Moreover, the typical values are more in line with real RNA molecules, while still being extensive in L.

In [7] a new Hamiltonian for RNA folding is discussed, with penalty terms for the genus and for the crossing number. They discuss a renormalization procedure that allow to collapse complicated diagrams into simpler one, but show no analytical or numerical results.

A On Gaussian integrals over GUE matrices

In this appendix, we collect results on Gaussian integrals over hermitian matrices.

In the following, all matrices are $n \times n$ hermitian. The integration measure is defined as

$$d\sigma = \prod_{i=1}^{n} d\sigma_{ii} \prod_{1 \le i < j \le n} d\sigma_{ij}^{(r)} d\sigma_{ij}^{(i)} \tag{19}$$

where the superscripts (r) and (i) denote real and imaginary parts.

Recall the 1d Gaussian integral:

$$\int dx \exp\left[-\frac{a}{2}x^2 + bx + c\right] = \sqrt{\frac{2\pi}{a}} \exp\left[\frac{b^2}{2a} + c\right] \qquad \forall a, b, c : a > 0\,. \tag{20}$$

A.1 Gaussian integrals and Hubbard-Stratonovich transformations

We compute the following integral:

$$\int dX \exp\left[-\frac{a}{2}\operatorname{Tr}\left(X^2\right) + b\operatorname{Tr}\left(XY\right)\right] \tag{21}$$

over $n \times n$ hermitian matrices X, and with $a > 0$, Y hermitian.

First of all, expand the traces:

$$\begin{aligned} \operatorname{Tr}\left(X^2\right) &= \sum_{\alpha} X_{\alpha\alpha}^2 + 2\sum_{\alpha<\beta}\left((X_{\alpha\beta}^{(r)})^2 + (X_{\alpha\beta}^{(i)})^2\right) \\ \operatorname{Tr}\left(XY\right) &= \sum_{\alpha} X_{\alpha\alpha}Y_{\alpha\alpha} + 2\sum_{\alpha<\beta}\left(X_{\alpha\beta}^{(r)}Y_{\alpha\beta}^{(r)} + X_{\alpha\beta}^{(i)}Y_{\alpha\beta}^{(i)}\right)\,. \end{aligned} \tag{22}$$

The integral is now straightforward as it is factorized in 1d gaussian integrals, giving

$$\int dX \exp\left[-\frac{a}{2}\operatorname{Tr}\left(X^2\right) + b\operatorname{Tr}\left(XY\right)\right] = \sqrt{\frac{\pi}{a}}^{n^2} \sqrt{2}^{n} \exp\left[\frac{b^2}{2a}\operatorname{Tr}\left(Y^2\right)\right]\,. \tag{23}$$

The inversion of this formula gives the Hubbard-Stratonovich transformation used in the main text.

The same line of reasoning allows to compute

$$\int dX \exp\left[-\frac{a}{2}\operatorname{Tr}\left(X^2\right) + b\operatorname{Tr}\left(XY\right)\right] X = \sqrt{\frac{\pi}{a}}^{n^2} \sqrt{2}^{n} \exp\left[\frac{b^2}{2a}\operatorname{Tr}\left(Y^2\right)\right] \frac{b}{a}Y\,. \tag{24}$$

A.2 Wick's theorem

Here we perform an explicit computation for the 4-point function

$$\frac{\int dX \exp\left[-\frac{n}{2}\sum_{i,j}(V^{-1})_{ij}\operatorname{Tr}\left(\phi_i\phi_j\right)\right]\frac{1}{n}\operatorname{Tr}\left(\phi_i\phi_j\phi_k\phi_l\right)}{\int dX \exp\left[-\frac{n}{2}\sum_{i,j}(V^{-1})_{ij}\operatorname{Tr}\left(\phi_i\phi_j\right)\right]} = \frac{1}{n}\sum_{\alpha,\beta,\gamma,\delta}\langle\phi_{i,\alpha\beta}\phi_{j,\beta\gamma}\phi_{k,\gamma\delta}\phi_{l,\delta\alpha}\rangle \tag{25}$$

.

for $1 \le i < j < k < l \le L$, with propagator

$$\langle\phi_{i,\alpha\beta}\phi_{j,\gamma\delta}\rangle = n^{-1}V_{ij}\delta_{\alpha\gamma}\delta_{\beta\delta}\,. \tag{26}$$

Wick's theorem allow to expand the 4-point function onto two-point functions, giving

$$\begin{aligned}
&\frac{1}{n}\sum_{\alpha,\beta,\gamma,\delta}\left[\langle\phi_{i,\alpha\beta}\phi_{j,\beta\gamma}\rangle\langle\phi_{k,\gamma\delta}\phi_{l,\delta\alpha}\rangle+\langle\phi_{i,\alpha\beta}\phi_{k,\gamma\delta}\rangle\langle\phi_{j,\beta\gamma}\phi_{l,\delta\alpha}\rangle+\langle\phi_{i,\alpha\beta}\phi_{l,\delta\alpha}\rangle\langle\phi_{j,\beta\gamma}\phi_{k,\gamma\delta}\rangle\right]\\
&=\frac{1}{n^3}\sum_{\alpha,\beta,\gamma,\delta}\left[V_{ij}V_{kl}\delta_{\alpha\gamma}\delta_{\beta\beta}\delta_{\gamma\delta}\delta_{\delta\delta}+V_{ik}V_{jl}\delta_{\alpha\delta}\delta_{\beta\gamma}\delta_{\beta\alpha}\delta_{\gamma\delta}+V_{il}V_{jk}\delta_{\alpha\alpha}\delta_{\beta\delta}\delta_{\beta\delta}\delta_{\gamma\gamma}\right]\\
&=\frac{1}{n^3}\left[V_{ij}V_{kl}n^3+V_{ik}V_{jl}n+V_{il}V_{jk}n^3\right]=V_{ij}V_{kl}+V_{ik}V_{jl}n^{-2}+V_{il}V_{jk}
\end{aligned}\tag{27}$$

B GUE spectral density

In this Appendix, we will compute the spectral density of the GUE ensamble, i.e.

$$\rho_n(\lambda)=\left\langle\frac{1}{n}\sum_i\delta\left(\lambda-\sigma_i\right)\right\rangle\tag{28}$$

where $\{\sigma_i\}$'s are the eigenvalues of the GUE matrix σ. The spectral density can be computed by passing to the eigenvalue measure, and then decoupling the Vandermonde by using the orthogonal polynomials trick:

$$\begin{aligned}
\rho_n(\lambda)&=\frac{\int\prod_{i=1}^n d\sigma_i\Delta(\sigma)^2\exp\left[-\frac{n}{2v}\sum_i\sigma_i^2\right]\frac{1}{n}\sum_i\delta\left(\lambda-\sigma_i\right)}{\int\prod_{i=1}^n d\sigma_i\Delta(\sigma)^2\exp\left[-\frac{n}{2v}\sum_i\sigma_i^2\right]}\\
&=\sqrt{\frac{n}{2v}}\frac{\int\prod_{i=1}^n d\tau_i\Delta(\tau)^2\exp\left[-\sum_i\tau_i^2\right]\delta\left(\sqrt{\frac{n}{2v}}\lambda-\tau_n\right)}{\int\prod_{i=1}^n d\tau_i\Delta(\tau)^2\exp\left[-\sum_i\tau_i^2\right]}\\
&=\sqrt{\frac{n}{2v}}\frac{\int\prod_{i=1}^n d\tau_i\left[\sum_{\vec{i},\vec{j}}\epsilon_{\vec{i}}\epsilon_{\vec{j}}H_{i_1}(\tau_1)\dots H_{i_n}(\tau_n)H_{j_1}(\tau_1)\dots H_{j_n}(\tau_n)\right]\exp\left[-\sum_i\tau_i^2\right]\delta\left(\sqrt{\frac{n}{2v}}\lambda-\tau_n\right)}{\int\prod_{i=1}^n d\tau_i\left[\sum_{\vec{i},\vec{j}}\epsilon_{\vec{i}}\epsilon_{\vec{j}}H_{i_1}(\tau_1)\dots H_{i_n}(\tau_n)H_{j_1}(\tau_1)\dots H_{j_n}(\tau_n)\right]\exp\left[-\sum_i\tau_i^2\right]}\\
&=\sqrt{\frac{n}{2v\pi}}e^{-\frac{n\lambda^2}{2v}}\frac{\sum_{\vec{i}}H_{i_n}^2\left(\sqrt{\frac{n}{2v}}\lambda\right)\epsilon_{\vec{i}}h_{i_1}\dots h_{i_{n-1}}}{\sum_{\vec{i}}\epsilon_{\vec{i}}h_{i_1}\dots h_{i_{n-1}}h_{i_n}}\\
&=\sqrt{\frac{1}{2\pi nv}}e^{-\frac{n\lambda^2}{2v}}\sum_{i=0}^{n-1}\frac{H_i^2\left(\sqrt{\frac{n}{2v}}\lambda\right)}{2^i i!}\,.
\end{aligned}\tag{29}$$

where $H_i(x)$ is the Hermite polynomial in the normalization such that

$$\begin{aligned}
\int dx H_i(x)H_j(x)\frac{\exp-x^2}{\sqrt{\pi}}&=h_i\delta_{ij}\\
h_i&=2^i i!\,.
\end{aligned}\tag{30}$$

An alternative form of the spectral density can be given be using the fact that

$$H_i(x)^2=\sum_{l=0}^{i}\frac{(i!)^2 2^{i-l}}{(l!)^2(i-l)!}H_{2l}(x)\tag{31}$$

and that

$$\sum_{j=0}^{k}\binom{j+r}{j}=\binom{k+r+1}{k}\tag{32}$$

so that

$$\rho_n(\lambda)=\frac{e^{-\frac{n\lambda^2}{2v}}}{\sqrt{2\pi nv}}\sum_{k=0}^{n-1}\binom{n}{k+1}\frac{H_{2k}\left(\sqrt{\frac{n}{2v}}\lambda\right)}{2^k k!}\,.\tag{33}$$

C The determinant of a Replica Symmetric matrix

Let $M(a,b) = a\mathbb{I} + b\mathbb{U}$, where $\mathbb{U}$ is the matrix of ones, and all matrices are $n \times n$. We want to compute the eigenvalues of this matrix. We have that

$$\left[M(a,b)\cdot(v_1\dots v_n)^T\right]_j = av_j + b\sum_{i=1}^{N} v_i \stackrel{!}{=} \lambda v_j \qquad \forall j = 1\dots n \tag{34}$$

so that

$$(a-\lambda)v_j + b\sum_i v_i = 0 \qquad \forall j = 1\dots n\,. \tag{35}$$

Choosing $\lambda = a$ allows to find $n-1$ independent eigenvectors, those with $v_1 = 1$, $v_j = -1$ and all the other entries zero. Choosing v to be the unit vector, one finds $\lambda = a + nb$, so that we complete the eigensystem. Thus, the eigenvalues are a with $n-1$ multiplicity and $a+nb$ with single multiplicity.

References

[1] Henri Orland and A. Zee. RNA folding and large N matrix theory. 620(3):456–476.

[2] Graziano Vernizzi, Henri Orland, and A. Zee. Enumeration of RNA Structures by Matrix Models. 94(16):168103.

[3] G. Vernizzi, H. Orland, and A. Zee. Prediction of RNA pseudoknots by Monte Carlo simulations.

[4] G. Akemann, G. M. Cicuta, L. Molinari, and G. Vernizzi. Compact support probability distributions in random matrix theory. 59(2):1489–1497.

[5] Michael Bon, Graziano Vernizzi, Henri Orland, and A. Zee. Topological Classification of RNA Structures. 379(4):900–911.

[6] G. Vernizzi, P. Ribeca, H. Orland, and A. Zee. The Topology of Pseudoknotted Homopolymers. 73(3):031902.

[7] Graziano Vernizzi, Henri Orland, and A. Zee. Classification and predictions of RNA pseudoknots based on topological invariants. 94(4):042410.