J. Phys. A: Math. Gen. 23 (1990) 5577-5599. Printed in the UK

Asymptotic corrections to the Wigner semicircular eigenvalue spectrum of a large real symmetric random matrix using the replica method

Gurjeet S Dhesi† and Raymund C Jones

School of Mathematics and Statistics, University of Birmingham, Birmingham B15 2TT, UK

Received 9 March 1990

Abstract. The replica method has previously been used to calculate the semicircular averaged eigenvalue spectrum of the Gaussian orthogonal ensemble of real symmetric $N \times N$ random matrices in the limit where $N \to \infty$. In this paper we develop a perturbative scheme which, within this same replica framework, is used to calculate the corrections within this semicircular band of eigenvalues to order $1/N$ and $1/N^2$. Comparison is made between these results and previously published work by other authors on the corrections to order $1/N$. A new and straightforward self-consistency argument is presented and used to derive the shape of the averaged eigenvalue spectrum when N is large but finite and the scaling behaviour of this averaged eigenvalue spectrum near the band edges is demonstrated in a straightforward fashion. Some comments are made on the relation of our results to those of field theoretical calculations in zero dimensions.

1. Introduction

In recent years there has been a revival of interest in problems associated with random matrices and their applications in physics. Seminal work was done by Dyson (1962) who pointed out that in studying the highly excited states of large nuclei it would be more profitable to study the statistical properties of such spectra rather than attempt a detailed *ab initio* calculation on a complex system. In such a problem one has readily available only global information on the symmetry of the underlying Hamiltonian which, because of the large number of particles involved, could be represented as a large square matrix (either symmetric or Hermitian). In the absence of other information one then proceeds rather as in statistical mechanics and constructs a statistical ensemble of such matrices subject to the requirement that each member of the ensemble obeys some physically useful symmetry requirement. For this reason much attention has focused on the Gaussian orthogonal ensemble (GOE) of $N \times N$ real symmetric matrices which is invariant under orthogonal transformations. This latter invariance leads to a description of a typical member of the ensemble as a real symmetric $N \times N$ matrix in which each element is a normally distributed random variable with mean zero. It was shown by Wigner (e.g. 1958) that the ensemble averaged eigenvalue density $\rho(\lambda)$ (for which $\rho(\lambda)\,d\lambda$ gives the average number of eigenvalues between λ and $\lambda + d\lambda$) is a semicircle whose radius, as $N \to \infty$, is proportional to the standard deviation associated with a matrix element of an individual member of the ensemble. A useful compendium of early results and papers in this area is to be found in the reprint collection edited by Porter (1965) and in the book by Mehta (1967) whiich has become a standard text

† Present address: Department of Physics, University of Leicester, Leicester, UK.

on important calculations and ideas in random matrix physics. There is a more recent and exhaustive review of work in this area by Brody *et al* (1981).

Use of random matrix ensembles has now found a firm place in areas outside nuclear physics. They have been used in condensed matter physics to describe problems associated with certain models of spin glasses by Kosterlitz *et al* (1976) and more recently in the description of atomic and molecular spectra by Camarda and Georgopulos (1983) and by Mukamel *et al* (1984). Some recent and fascinating work by Bohigas and Giannoni (1983) has shown that the eigenvalue spectra of the quantum counterparts of non-integrable classically chaotic systems show many of the features associated with the GOE described earlier.

The earliest calculations of the averaged eigenvalue density (AED) for the GOE when $N \to \infty$ are to be found in the texts by Porter (1965) and Mehta (1967) and usually rely either on elaborate moment expansions or on the properties of the Hermite polynomials and oscillator wavefunctions which arise naturally from the many integrations against a Gaussian weight which occur in the GOE. However, a radically different method was presented by Edwards and Jones (1976), (referred to hereafter as EJ), for calculating the AED of the GOE this new technique allowed an extension of existing results to the case of a GOE in which each matrix element is allowed to have a finite non-zero mean. The methods used by these latter authors rely on the so-called replica method first used by Edwards (1970) in the study of polymer physics. EJ predicted that under certain circumstances an isolated eigenvalue would appear outside the Wigner semicircular band; this result was confirmed by Jones *et al* (1978) using completely different methods based on techniques used to describe localized excitation modes in an impure lattice. The work on ensembles with a zero mean has been extended by Edwards and Warner (1980) to Hermitian matrices. More recently Jones and Dhesi (1990) have used the replica method to study the spectrum of the random sign ensemble, first studied by Wigner (1955, 1957) and then confirmed the Wigner conjecture (1958), that any ensemble of matrices whose elements are described by any reasonably well behaved probability density function, would yield the Wigner semicircle as $N \to \infty$.

There has also been interest in the behaviour of the AED when N is large but finite. Bronk (1964) studied the band tailing which occurred in the AED of a GOE near the edges of the Wigner band when N was large but finite. Since then there have been a number of papers which addressed the problem of calculating the corrections to the Wigner semicircle which are of order $1/N$. Work by Takano and Takano (1984) uses a direct graphical technique; that by Verbaarschot *et al* (1984) uses a moment calculation and an important paper by Verbaarschot and Zirnbauer (1984), (referred to hereafter as VZ), casts the whole problem into the language of a ϕ^3 field theory. VZ produced an explicit expression for the AED near the Wigner band edge and also calculate the $1/N$ corrections to Wigner's result inside the band. Their results are at variance with those of Takano and Takano (1984).

In this paper we revert to the replica method and assumption of replica symmetry made by EJ. In section 2 we outline the EJ replication method for calculating the AED of a large real symmetric random matrix. In section 3 we set up a diagrammatic expansion and a perturbation theory which will allow us in section 4 to calculate the corrections both of order $1/N$ and of order $1/N^2$ to the Wigner semicircle. Our results to $O(1/N)$ will be shown to agree with those of VZ rather than those of Takano and Takano (1984); the results for the corrections to the semicircle which are of order $1/N^2$ are new. Both sets of corrections are non-vanishing and convergent only inside the Wigner semicircle and away from its band edges: further comments on this are made in section 5. In section 5 we show how a self-consistency condition imposed on our formalism can be used to describe the behaviour of the AED near the band edge and to provide a reasonable description of the rest of the AED when N is large (but finite). Our final results describing the band edges are identical with those of VZ, although they are derived in a very different fashion, and show the expected non-analyticity in N and scaling behaviour at the band edges. We believe that our calculation of the AED near the band edge is significantly simpler than that published in previous work by both Bronk (1984) and VZ. Finally, in section 6 we compare our analytical calculations with some numerical simulations on random matrices and (as an extreme case) with the exact result known for $N=2$.

2. The replica technique

In this section we outline the method developed by EJ for calculating the AED of a real symmetric $N\times N$ matrix $\mathbf{J}$ with eigenvalues $\{J_i\}$. The density $\nu(\lambda)$ of such eigenvalues, chosen as normalized to unity, is given by the expression

$$\nu(\lambda)=N^{-1}\sum\delta(\lambda-J_i). \tag{2.1}$$

If we use the result that $\det(\mathbf{I}\lambda-\mathbf{J})=\Pi\,(\lambda-J_i)$ and give λ the usual infinitesimal imaginary part, $-\mathrm{i}\varepsilon$, then (2.1) can be written as

$$\nu(\lambda)=\frac{1}{N\pi}\,\mathrm{Im}\,\frac{\partial}{\partial\lambda}\ln\det(\mathbf{I}\lambda-\mathbf{J}). \tag{2.2}$$

The method developed by EJ uses the result that

$$\ln x=\lim_{n\to 0}\left(\frac{x^n-1}{n}\right) \tag{2.3}$$

in order to write (2.2) as

$$\nu(\lambda)=-\frac{2}{N\pi}\,\mathrm{Im}\,\frac{\partial}{\partial\lambda}\lim_{n\to 0}\frac{1}{n}\,(\det{}^{-1/2}(\mathbf{I}\lambda-\mathbf{J})^n-1). \tag{2.4}$$

The determinant is now parametrized as a multiple Fresnel integral of the form

$$\det{}^{-1/2}(\mathbf{I}\lambda-\mathbf{J})=\left(\frac{\mathrm{e}^{\mathrm{i}\pi/4}}{\pi^{1/2}}\right)^N\int_{-\infty}^{+\infty}\prod_i \mathrm{d}x_i\,\exp\left(-\mathrm{i}\sum_{i,j}x_i(\mathbf{I}\lambda-\mathbf{J})_{ij}x_j\right). \tag{2.5}$$

We now substitute (2.5) into (2.4) and assume that this latter result holds for integer values of n and may then be continued to $n=0$. We thus obtain the basic result that

$$\nu(\lambda)=-\frac{2}{N\pi}\,\mathrm{Im}\,\frac{\partial}{\partial\lambda}\lim_{n\to 0}\frac{1}{n}\left[\frac{\mathrm{e}^{\mathrm{i}\pi/4}}{\pi^{1/2}}\right]^{Nn}\int_{-\infty}^{+\infty}\prod_{i,\alpha}\mathrm{d}x_i^\alpha \times\left[\exp\left(-\mathrm{i}\sum_{ij,\alpha}x_i^\alpha(\lambda\delta_{ij}-J_{ij})x_j^\alpha\right)-1\right]. \tag{2.6}$$

The integration is now over the Nn variables $\{x_i^\alpha\}$ where the indices i and α range from 1 to N and from 1 to n respectively; the limit $n\to 0$ is to be taken at the end of the calculations.

The averaged density of eigenvalues $\rho(\lambda)$ of an ensemble of real symmetric matrices, from which a typical matrix element J_{ij} has a probability density function (PDF) $p(J_{ij})$, is then obtained by calculating

$$\rho(\lambda)=\int\nu(\lambda;\{J_{ij}\})\prod_{ij}p(J_{ij})\,\mathrm{d}J_{ij} \tag{2.7}$$

3. A diagrammatic expansion for $\rho(\lambda)$ in the GOE

We consider an $N\times N$ real symmetric matrix $\mathbf{J}$; such a matrix is a member of the GOE if its matrix elements $\{J_{ij}\}$ are described by the following probability density function (see, e.g., Porter 1965)

$$p(J_{ij}\}=\begin{cases}\dfrac{\exp-(J_{ij}^2/2\sigma^2)}{\sqrt{2\pi\sigma^2}} & \text{for } i\neq j\\[2ex] \dfrac{\exp-(J_{ij}^2/4\sigma^2)}{\sqrt{4\pi\sigma^2}} & \text{for } i=j.\end{cases} \tag{3.1}$$

For convenience, and by convention, we write $\sigma^2\equiv J^2/N$, where J^2 is a number of order unity.

An expression for the AED is then straightforwardly obtained by substituting the PDF given by (3.1) into (2.7) and carrying out the Gaussian integrals over the $\{J_{ij}\}$: the result is that

$$\rho(\lambda)=-\frac{2}{N\pi}\,\mathrm{Im}\,\frac{\partial}{\partial\lambda}\lim_{n\to0}\frac{1}{n}\left\{\left(\frac{e^{i\pi/4}}{\pi^{1/2}}\right)^{Nn}\int_{-\infty}^{+\infty}\prod_{i,\alpha}dx_i^\alpha \times\exp\left[-i\lambda\sum_{i\alpha}(x_i^\alpha)^2-\frac{J^2}{N}\sum_{i,j}\left(\sum_\alpha x_i^\alpha x_j^\alpha\right)^2\right]-1\right\}. \tag{3.2}$$

It will prove convenient to rearrange the terms in the exponent by writing

$$\sum_{i,j}\left[\sum_\alpha x_i^\alpha x_j^\alpha\right]^2\equiv\sum_\alpha\left[\sum_i (x_i^\alpha)^2\right]^2+\sum_{\substack{\alpha\neq\beta\\ i,j}} x_i^\alpha x_j^\alpha x_i^\beta x_j^\beta$$

whence (3.2) becomes

$$\rho(\lambda)=-\frac{2}{N\pi}\,\mathrm{Im}\,\frac{\partial}{\partial\lambda}\lim_{n\to0}\left\{\left(\frac{e^{i\pi/4}}{\pi^{1/2}}\right)^{Nn}\int_{-\infty}^{+\infty}\prod_{i,\alpha}dx_i^\alpha \times\exp\left[-i\lambda\sum_{i\alpha}(x_i^\alpha)^2-\frac{J^2}{N}\sum_\alpha\left(\sum_i (x_i^\alpha)^2\right)^2\right] \times\exp\left[-\frac{J^2}{N}\sum_{i,j;\alpha\neq\beta} x_i^\alpha x_j^\alpha x_i^\beta x_j^\beta\right]. \tag{3.3}$$

This expression is very similar to equation (3.3) of EJ and hence may be handled in the fashion described there: since we shall ultimately be considering the case where N is large but $n\to0$, we first seek the terms in the exponent which are of order Nn. The second term which is the exponent of (3.3), contains such a term, but it was argued in EJ that the third term in the exponent (3.3) (i.e. the term containing $\Sigma_{\alpha\neq\beta}$) has zero mean but a square which is of order n. A careful diagrammatic analysis by Edwards and Warner (1980) confirms that in the limit $N\to\infty$ and $n\to0$ the term with $\alpha\neq\beta$ gives a contribution which is of order $n1$ (rather than $O(nN)$) and so in the limit $N\to\infty$ may be neglected. Thus the AED, $\rho_0(\lambda)$, in this limit $N\to\infty$ is obtained by evaluating

$$\rho_0(\lambda)=-\frac{2}{N\pi}\,\mathrm{Im}\,\frac{\partial}{\partial\lambda}\lim_{n\to0}\frac{1}{n}\left\{\left(\frac{e^{i\pi/4}}{\pi^{1/2}}\right)^{Nn}\int_{-\infty}^{+\infty}\prod_{i,\alpha}dx_i^\alpha \times\exp\left[-\sum_\alpha\left(+i\lambda\sum_i(x_i^\alpha)^2+\frac{J^2}{N}\left(\sum_i(x_i^\alpha)^2\right)^2\right)\right]-1\right\}. \tag{3.4}$$

It is shown in Jones and Dhesi (1990) that if we make the assumption of replica symmetry by replacing each of the n variables $\{x_i^\alpha\}$, $\alpha = 1 \ldots n$, by a single variable x_i then (3.4) becomes

$$\rho_0(\lambda) = -\frac{2}{N\pi} \operatorname{Im} \frac{\partial}{\partial \lambda} \lim_{n \to 0} \frac{1}{n} \left[\left(\frac{e^{i\pi/4}}{\pi^{1/2}} \right)^{Nn} H^n - 1 \right] \tag{3.5}$$

where

$$H \equiv \int_{-\infty}^{+\infty} \prod_i dx_i \exp\left[-i\lambda \sum_i x_i^2 - \frac{J^2}{N} \left(\sum_i x_i^2 \right)^2 \right]. \tag{3.6}$$

This latter integral is most easily evaluated by introducing polar coordinates in the N-dimensional space of the $\{x_i\}$ whence a straightforward saddle point integration and use of the basic idenity (2.3) gives

$$\rho_0(\lambda) = \begin{cases} \dfrac{1}{2\pi J^2} (4J^2 - \lambda^2)^{1/2} & \text{for } |\lambda| < 2J \\ 0 & \text{for } |\lambda| > 2J. \end{cases} \tag{3.7}$$

This latter is, of course, the well known Wigner semicircle for the AED and holds when $N \to \infty$.

Thus we see that the first two terms in the exponent of (3.3) yield the $N \to \infty$ limit for $\rho(\lambda)$ which is the Wigner semicircle (3.7). Our aim is to use (3.3) to develop the systematic corrections to $\rho_0(\lambda)$ to $O(1/N)$ and $O(1/N^2)$ and compare our calculations to $O(1/N)$ with those of other authors. We shall see that there are problems with such an expansion. Loosely, we shall construct an expansion of (3.3) by expanding out the final exponential in a series in $(J^2/N)\sum_{\alpha \neq \beta} \ldots$ (since we have already argued that this latter term does not contribute in the limit $N \to \infty$).

Our starting point is to use the auxiliary field (or Hubbard–Stratanovitch) identity (see, e.g., Sherrington 1971) in the form used by EJ, i.e.

$$\exp\left[-\frac{J^2}{N} \left(\sum_i (x_i^\alpha)^2 \right)^2 \right] \\ \equiv \left[\frac{N}{2\pi} \right]^{1/2} \frac{\lambda}{(2J^2)^{1/2}} \int_{-\infty}^{+\infty} ds^\alpha \exp\left[-\frac{\lambda^2}{4J^2} N(s^\alpha)^2 - i\lambda s^\alpha \sum_i (x_i^\alpha)^2 \right]. \tag{3.8}$$

We use this in (3.3) to obtain the expression

$$\rho(\lambda) = -\frac{2}{N\pi} \operatorname{Im} \frac{\partial}{\partial \lambda} \lim_{n \to 0} \frac{1}{n} \left[\frac{e^{i\pi/4}}{\pi^{1/2}} \right]^{Nn} \left(\frac{N}{2\pi} \right)^{n/2} \left(\frac{\lambda}{\sqrt{2J^2}} \right) \\ \times \int_{-\infty}^{+\infty} \prod_\alpha ds^\alpha \exp\left[-\frac{N\lambda^2}{4J^2} \sum_\alpha (s^\alpha)^2 \right]_I \tag{3.9}$$

where

$$I = \int_{-\infty}^{+\infty} \prod_{i,\alpha} dx_i^\alpha \exp\left[-i\lambda \sum_{i,\alpha} (1 + s^\alpha)(x_i^\alpha)^2 \right] \\ \times \left[1 - \frac{J^2}{N} \sum_{\substack{\alpha \neq \beta \\ i,j}} x_i^\alpha x_j^\alpha x_i^\beta x_j^\beta + \frac{1}{2!} \left(-\frac{J^2}{N} \right)^2 \sum_{\substack{\alpha \neq \beta, \gamma \neq \delta \\ i,j,k,l}} \ldots \right] \tag{3.10}$$

and the series in this last expression is simply the expansion of the third exponential in (3.3).

For convenience we write $I \equiv \Sigma_{k=0}^{\infty} I^{(k)}$, where $I^{(k)}$ is the contribution to (3.10) from the term in $(-J^2/N)^k$. At this stage we make the assumption of replica symmetry which (following our earlier remarks and the arguments in EJ) consists of replacing s^α by the variable s which we assume independent of α.

Clearly

$$I^{(0)} \equiv \int_{-\infty}^{+\infty} \prod_{i,\alpha} \mathrm{d}x_i^\alpha \exp\left[-\mathrm{i}\lambda \sum_{i,\alpha} (1+s^\alpha)(x_i^\alpha)^2\right]$$

is easily evaluated and has the value

$$\prod_{i,\alpha} \left[\frac{\pi}{\mathrm{i}\lambda(1+s^\alpha)}\right]^{1/2}$$

which simply becomes

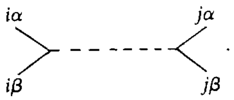
$$\left[\frac{\pi}{\mathrm{i}\lambda(1+s)}\right]^{Nn/2}$$

under the assumption of replica symmetry.

We regard (3.10) as providing a perturbation expansion for I which may be evaluated by integrating (or 'averaging') successive terms of the series against the (complex) Gaussian exponential. The averaged eigenvalue spectrum can be calculated by substituting a suitable approximation for I back into (3.9); in this latter expansion we shall be dividing by n, taking the limit as $n \to 0$ and then using the basic identity (2.3) to produce a non-zero contribution to $\rho(\lambda)$ when $n \to 0$. It is then clear that in every order of the perturbation series (3.10) we must retain the term which is linear in n. Higher powers of n will yield no contribution to $\rho(\lambda)$. Direct calculation will demonstrate this point explicitly. Keeping track of the higher-order terms in (3.10) is plainly tedious so we use a diagrammatic technique to aid the calculation. We represent the vertex

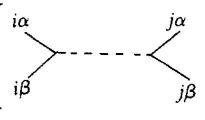
$$\sum_{\substack{\alpha \neq \beta \\ i,j}} x_i^\alpha x_j^\alpha x_i^\beta x_j^\beta$$

by the diagram



The first order contribution to I, $I^{(1)}$, can thus be represented as

$$I^{(1)} = \frac{J^2}{N} \left\{ \quad \right\}_x \qquad (3.11)$$



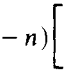
where the brackets, $\{\}_x$, denote the average against the Gaussian weight in (3.10).

Since the vertex contains the restriction that $\alpha \neq \beta$, the usual rules of integration against a Gaussian yield a non-zero contribution to (3.10) only when $i = j$. This we represent by joining the 'legs' labelled $i\beta$ and $j\beta$. There is only one way of doing this, i.e. we say that the symmetry factor, S, of the resulting graph is 1. The resulting integral is then

$$\int_{-\infty}^{+\infty}\prod_{i,\gamma}\mathrm{d}x_i^\gamma(x_i^\alpha)^2(x_i^\beta)^2\exp\left[-\mathrm{i}\lambda\sum_{k,\gamma}(1+s)(x_k^\gamma)^2\right] = \left[\frac{\pi}{\mathrm{i}\lambda(1+s)}\right]^{Nn/2}\left[\frac{1}{2\mathrm{i}\lambda(1+s)}\right]\left[\frac{1}{2\mathrm{i}\lambda(1+s)}\right] \tag{3.12}$$

we thus may write

$$I^{(1)} = -\frac{J^2}{N}\left[\frac{\pi}{\mathrm{i}\lambda(1+s)}\right]^{Nn/2} N(n^2-n)\left[\quad\right](S\equiv 1). \tag{3.13}$$



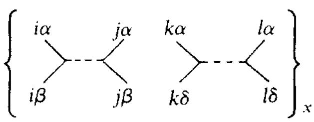
We identify a single continuous line resulting from the contraction of two sets of indices with a propagator $\Gamma(s,\lambda)$ defined by

$$\Gamma(s,\lambda) \equiv \text{———} \equiv \frac{1}{2\mathrm{i}\lambda(1+s)}.$$

The normalization factor $(\pi/\mathrm{i}\lambda(1+s))^{Nn/2}$, is simply the value of $I^{(0)}$. Thus $I^{(1)} \equiv -I^{(0)}J^2(n^2-n)\Gamma^2(s,\lambda)$ and is nominally of order N^0.

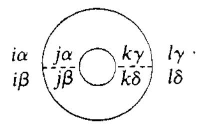
The next term of the expansion, $I^{(2)}$, can be represented as

$$I^{(2)} = \frac{1}{2!}\left[-\frac{J}{N}\right]^2 \left\{\quad\right\}_x . \tag{3.14}$$



There are three distinct types of contribution to (3.14). Consider the set of contractions which leads to a diagram of the form

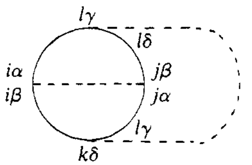
$$\sum_{\substack{i,j,k,l\\ \alpha\neq\beta,\gamma\neq\delta}} \tag{3.15}$$



There are four distinct ways of contracting so as to form such a diagram, i.e. the symmetry factor of (3.15) is 4. The contraction sets $i=l$, $j=k$, $\alpha=\gamma$, $\beta=\delta$ but maintains the constraints $\alpha\neq\beta$ and $\gamma\neq\delta$. If the summations are performed, we see that (3.15) is of order $N^2(n^2-n)$ and gives to $I^{(2)}$ a contribution of order $N^0(n^2-n)$.

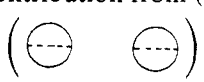
Equation (3.14) also contains contractions which lead to a diagram of the form

$$\sum_{\substack{i,j,k,l\\ \alpha\neq\beta,\gamma\neq\delta}} \tag{3.16}$$



This also has a symmetry factor of 4.

The contraction sets $i=l$, $l=j$ (twice) and $k=i$; also $\alpha=\gamma$, $\delta=\beta$, with the constraint that $\alpha\neq\beta$ and $\delta\neq\gamma$. Performing the summations, we see that (3.16) is of order $N(n^2-n)$ and gives to $I^{(2)}$ a contribution of order $N^{-1}(n^2-n)$, i.e. lower by a factor N than the contribution from (3.15). It is easily seen that an unlinked diagram such as



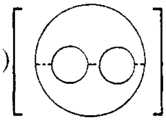
will give to (3.14) a contribution whose n-dependence is of order $(n^2-n)^2$. In the light of our previous remarks, such a term is of order n^2 as $n \to 0$ and makes no contribution to $\rho(\lambda)$ and so may be omitted. This result is general and holds to all orders in perturbation theory.

Thus for small n we have

$$I^{(2)} = \frac{4.1!}{2!}(-J^2)^2 I^{(0)}(n^2-n)\Gamma^4(s,\lambda) + \frac{4.1!}{2!}(-J^2)^2 I^{(0)} \frac{(n^2-n)}{N}\Gamma^4(s,\lambda). \tag{3.17}$$

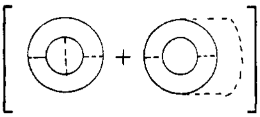
A similar explicit calculation for the third-order term, $I^{(3)}$, in I shows that the contributions of order $N^0(n^2-n)$ and $N^{-1}(n^2-n)$ are

$$I^{(3)} = \frac{1}{3!}\left(-\frac{J^2}{N}\right)^3 I^{(0)} N^3 (n^2-n)$$



$$\otimes$$

$$+\frac{1}{3!}\left(-\frac{J^2}{N}\right)^3 I^{(0)} N^2 (n^2-n)$$



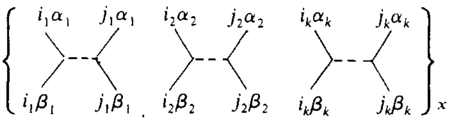
$$\otimes \tag{3.18}$$

where the numerical factors post multiplying each square bracket are the symmetry factors for the graphs within the bracket. From this we see that

$$I^{(3)} = \frac{4^2 \cdot 2!}{3!}(-J^2)^3 I^{(0)}(n^2-n)\Gamma^6(s,\lambda) + \frac{4^2 2!}{3!} 2(-J^2)^3 I^{(0)} \frac{(n^2-n)}{N}\Gamma^6(s,\lambda). \tag{3.19}$$

The general result giving the contribution to $I^{(k)}$ is established in the same way:

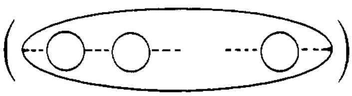
$$I^{(k)} = \frac{1}{k!}\left(-\frac{J^2}{N}\right)^k$$



$$. \tag{3.20}$$

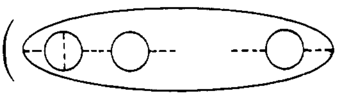
Again we perform the contractions as before and retain terms in $I^{(k)}$ of order $N^0(n^2-2n)$ and $N^{-1}(n^2-n)$; this gives

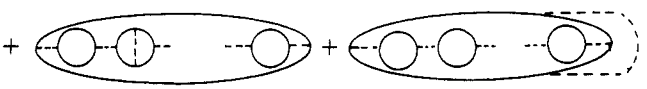
$$I^{(k)} = \frac{1}{k!}\left(-\frac{J^2}{N}\right)^k I^{(0)} N^k (n^2-n)$$



$$S_0$$

$$+\frac{1}{k!}\left(-\frac{J^2}{N}\right)^k I^{(0)} N^{k-1} (n^2-n)$$





$$\ldots + \ldots \Big) S_{-1}. \tag{3.21}$$

Each of the symmetry factors S_0 and S_{-1} has the value $4^{k-1}(k-1)!$ (cf (3.17) and (3.19) for $k=2$ and $k=3$ respectively) and there are $(k-1)$ diagrams of order $N^{-1}(n^2-n)$ each of which makes an identical contribution.

From this we see that

$$I^{(k)}=4^{k-1}\frac{(k-1)!}{k!}(-J^2)^kI^{(0)}(n^2-n)\Gamma^{2k}(s,\lambda)$$
$$+4^{k-1}\frac{(k-1)!}{k!}(k-1)(-J^2)^kI^{(0)}\frac{(n^2-n)}{N}\Gamma^{2k}(s,\lambda) \quad (3.22)$$

where, anticipating the final result, we have omitted the contribution of terms which are proportional to higher powers of (n^2-n) and (N^{-1}).

Finally, we obtain the result that

$$I=I^{(0)}\left\{1+n\left[J^2\Gamma^2-\frac{4}{2}(J^2\Gamma^2)^2+\frac{4^2}{3}(J^2\Gamma^2)^3-\frac{4^3}{4}(J^2\Gamma^2)^4+\dots\right]\right.$$
$$\left.+\frac{n}{N}\left[-\frac{4}{2}(J^2\Gamma^2)^2+\frac{4^2 2}{3}(J^2\Gamma^2)^3-\frac{4^3 3}{4}(J^2\Gamma^2)^4+\frac{4^4 4}{5}(J^2\Gamma^2)^5+\dots\right]\right\}. \quad (3.23)$$

Again, only terms in N^0n and $N^{-1}n$ have been retained; the series can be summed in closed form and its value is $I=I^{(0)}[1+nP]$ where

$$P=\frac{1}{4}\left\{\ln(1+4J^2\Gamma^2)-\frac{1}{N}\left[\ln(1+4J^2\Gamma^2)-\frac{4J^2\Gamma^2}{(1+4J^2\Gamma^2)}\right]\right\}. \quad (3.24)$$

Up to terms linear in n, we may write $I=I^{(0)}\exp(nP)$. Under the assumption of replica symmetry, this may be substituted back into (3.9) to give

$$\rho(\lambda)=-\frac{2}{N\pi}\operatorname{Im}\frac{\partial}{\partial\lambda}\lim_{n\to 0}\frac{1}{n}\left(\frac{e^{i\pi/4}}{\pi^{1/2}}\right)^{Nn}$$
$$\times\left\{\int_{-\infty}^{+\infty}ds\left(\frac{N}{2\pi}\right)^{1/2}\frac{\lambda}{\sqrt{2J^2}}e^{-N\lambda^2s^2/4J^2}\left(\frac{\pi}{i\lambda(1+s)}\right)^{N/2}e^{P(s,\lambda)}\right\}^n. \quad (3.25)$$

The basic identity (2.3) is now used once again to reconstruct the logarithm, and we finally obtain

$$\rho(\lambda)=-\frac{2}{N\pi}\operatorname{Im}\frac{\partial}{\partial\lambda}\ln\left\{\lambda\exp\left(-\frac{N}{2}\ln\lambda\right)\right.$$
$$\left.\times\int_{-\infty}^{+\infty}ds\exp\left(-Ng(s,\lambda)+f(s,\lambda)+\frac{1}{N}h(s,\lambda)\right)\right\} \quad (3.26)$$

where

$$g(s,\lambda)=\frac{\lambda^2s^2}{4J^2}+\frac{1}{2}\ln i(1+s) \quad (3.27a)$$

$$f(s,\lambda)=\frac{1}{4}\ln\left(1-\frac{J^2}{\lambda^2(1+s)^2}\right) \quad (3.27b)$$

and

$$h(s,\lambda)=-\frac{1}{4}\ln\left(1-\frac{J^2}{\lambda^2(1+s)^2}\right)+\frac{1}{4}\frac{J^2}{(J^2-\lambda^2(1+s)^2)}. \quad (3.27c)$$

Equations (3.26), (3.27a), (3.27b) and (3.27c) are the basic results from which we shall derive the corrections to the Wigner semicircle of orders N^{-1} and N^{-2}.

4. Calculation of $\rho(\lambda)$ to order N^{-1} and N^{-2}

We now show that (3.26) will yield an expression for $\rho(\lambda)$ to $O(N^{-2})$ away from the band edges ($\lambda = \pm 2J$) by constructing an asymptotic expansion for the integral in (3.26) in powers of N^{-1}. We rewrite (3.26) as

$$\rho(\lambda) = -\frac{2}{N\pi}\,\mathrm{Im}\,\frac{\partial}{\partial\lambda}\left\{\ln\lambda\,\exp\left(-\frac{N}{2}\ln\lambda\right) + W_N(\lambda)\right\} \tag{4.1}$$

where

$$W_N(\lambda) = \ln\left[\int_{-\infty}^{+\infty} \mathrm{d}s\,\exp\left(-Ng(s,\lambda) + f(s,\lambda) + \frac{1}{N}h(s,\lambda)\right)\right]. \tag{4.2}$$

In appendix A, we show that for large N, a saddle point evaluation of the integral of the form $E_N \equiv \int_c \mathrm{e}^{NG(s)}\,\mathrm{d}s$ yields the approximate result that

$$E_N \approx \left(-\frac{2\pi}{NG''(\bar{s})}\right)^{1/2} \mathrm{e}^{NG(\bar{s})}\left[1 + \frac{1}{2N}\left(\frac{1}{4}\frac{G^{(\mathrm{iv})}(\bar{s})}{(G''(\bar{s}))^2} - \frac{5}{12}\frac{G'''(\bar{s})}{(G''(\bar{s}))^3}\right) + O(N^{-2})\right] \tag{4.3}$$

where $\bar{s}$ is determined by the saddle point condition $[\partial G(s)/\partial s]_{s=\bar{s}} = 0$.

We write

$$G = -g(s,\lambda) + (1/N)f(s,\lambda) + (1/N^2)h(s,\lambda) \tag{4.4}$$

and use (4.3) and (4.2) to determine an expansion of $\rho(\lambda)$ in powers of N^{-1}.

Let us denote by s_0 the saddle point associated with $g(s,\lambda)$ and which is defined by $(\partial g/\partial s)_{s=s_0} = 0$. The presence of the terms in $f(s,\lambda)$ and $h(s,\lambda)$ will cause a shift in the saddle point of G from s_0 to $\bar{s}$ where, to order N^{-1}, we can use Taylor's theorem to write

$$\bar{s} = s_0 + \frac{1}{N}\frac{f'(s_0)}{g''(s_0)} + O(N^{-2}). \tag{4.5}$$

In order to determine the requisite approximation to $W_N(\lambda)$ and hence $\rho(\lambda)$, we replace $\bar{s}$ in (4.3) by the approximate expression given in (4.5) and then use definition (4.4) of G in order to expand out (4.3) in powers of N^{-1}. The calculation is lengthy and given in appendix B where it is shown that

$$\begin{aligned} W_N(\lambda) = {} & \frac{1}{2}\ln\left(-\frac{2\pi}{N}\right) - Ng(s_0) - \frac{1}{2}\ln g''(s_0) + f(s_0) \\ & + \frac{1}{N}\left[h(s_0) + \left(-\frac{1}{8}\frac{g^{(\mathrm{iv})}(s_0)}{(g''(s_0))^2} + \frac{5}{24}\frac{(g'''(s_0))^2}{(g''(s_0))^3} + \frac{f''(s_0)}{2g''(s_0)}\right.\right. \\ & \left.\left. + \frac{(f'(s_0))^2}{2g''(s_0)} - \frac{f'(s_0)g'''(s_0)}{2(g''(s_0))^2}\right)\right]. \end{aligned} \tag{4.6}$$

For convenience of notation, we have not written the explicit dependence on λ the functions f, g and h in (4.6).

From the definition, (3.27a) it is easy to see that $g(s,\lambda)$ has conjugate saddle points at $\frac{1}{2}\{[-1 \pm \mathrm{i}[(4J^2/\lambda^2) - 1]^{1/2}\}$ and it is argued in EJ that the contour chosen for the saddle point integration may only be deformed to pass through one of these saddle points and that the lower saddle point leads to a physically reasonable positive AED $\rho(\lambda)$. Thus following EJ we choose

$$s_0=\frac{1}{2}\left[-1-\mathrm{i}\left(\frac{4J^2}{\lambda^2}-1\right)^{1/2}\right]. \tag{4.7}$$

Substituting (4.6) into (4.1) finally yields

$$\rho(\lambda)=\rho_0(\lambda)+\rho_1(\lambda)+\rho_2(\lambda)+\mathrm{O}(N^{-3}) \tag{4.8}$$

where

$$\rho_0(\lambda)=-\frac{2}{N\pi}\operatorname{Im}\frac{\partial}{\partial\lambda}\left(-\frac{N}{2}\ln\lambda-Ng(s_0)\right) \tag{4.9a}$$

$$\rho_1(\lambda)=-\frac{2}{N\pi}\operatorname{Im}\frac{\partial}{\partial\lambda}(\ln\lambda-\tfrac{1}{2}\ln g''(s_0)+f(s_0)) \tag{4.9b}$$

$$\rho_2(\lambda)=-\frac{2}{N\pi}\operatorname{Im}\frac{\partial}{\partial\lambda}\left(\frac{1}{N}h(s_0)+\frac{1}{N}Q(s_0)\right) \tag{4.9c}$$

and

$$Q(s_0)=-\frac{1}{8}\frac{g^{(\mathrm{iv})}(s_0)}{(g''(s_0))^2}+\frac{5}{24}\frac{(g'''(s_0))^2}{(g''(s_0))^3}+\frac{f''(s_0)}{2g''(s_0)}+\frac{(f'(s_0))^2}{2g''(s_0)}-\frac{f'(s_0)g'''(s_0)}{2(g''(s_0))^2}. \tag{4.9d}$$

Equations (4.8) and (4.9) give the required expansion of $\rho(\lambda)$ in powers of N^{-1}. It should be emphasized that although the diagrammatic calculations in section 3 yield the argument of the exponential in (3.26) to order N^{-1}, the analysis of this section shows that this will give an expansion of $\rho(\lambda)$ which is correct to order N^{-2}. However because of convergence problems associated with the series (3.23) when $s=s_0$ and λ is close to $\pm 2J$, we should not expect this method to yield sensible corrections to $\rho_0(\lambda)$ near the edges of the semicircular band of eigenvalues at $|\lambda|=2J$. Thus we expect that for large but finite N, (4.8) will yield a good asymptotic approximation to the AED only for $\lambda^2<4J^2$, i.e. inside the semicircular band of eigenvalues. In section 5 we shall consider the shape of the AED $\rho(\lambda)$ near $|\lambda|=2J$ when N is large but finite.

We now explicitly evaluate $\rho(\lambda)$. The contributions $\rho_0(\lambda)$ is obtained from (3.27a), (4.9a) and (4.7) and gives the result

$$\rho_0(\lambda)=-\frac{2}{N\pi}\operatorname{Im}\left[-\frac{1}{4J^2}(\lambda+\mathrm{i}\sqrt{4J^2-2\lambda^2})\right].$$

It will prove useful to define a Green function $\mathscr{G}_0(\lambda)$ associated with $\rho_0(\lambda)$ by the relation

$$\mathscr{G}_0(\lambda)=\frac{1}{2J^2}(\lambda+\mathrm{i}\sqrt{4J^2-2\lambda^2}) \tag{4.10}$$

so that formally $\rho_0(\lambda)\equiv 1/\pi\operatorname{Im}\mathscr{G}_0(\lambda)$. From these we see that

$$\rho_0(\lambda)=\begin{cases}\dfrac{1}{2\pi J^2}\sqrt{4J^2-2\lambda^2} & |\lambda|<2J\\ 0 & |\lambda|>2J\end{cases} \tag{4.11}$$

i.e. in the limit $N\to\infty$, we once again have the Wigner semicircular band of eigenvalues, normalized so that $\int_{-\infty}^{+\infty}\rho_0(\lambda)\,\mathrm{d}\lambda=1$.

The first-order correction to $\rho(\lambda)$ (i.e. the correction of order N^{-1}) is now obtained from (4.2*b*), (3.27*a*), (3.27*b*) and (4.7). We define a first-order correction $\mathscr{G}_1(\lambda)$ to the Green function $\mathscr{G}_0(\lambda)$ by the relation.

$$\rho_1(\lambda)\equiv\frac{1}{\pi}\operatorname{Im}\mathscr{G}_1(\lambda)\equiv-\frac{2}{N\pi}\operatorname{Im}\frac{\partial}{\partial\lambda}(\ln\lambda-\tfrac{1}{2}\ln g''(s_0)+f(s_0)) \tag{4.12}$$

Explicit calculation of (4.12) shows that

$$\mathscr{G}_1(\lambda)=\frac{1}{2N(\lambda^2-4J^2)}[\lambda+(\lambda^2-4J^2)^{1/2}] \tag{4.13}$$

and we see straightforwardly from this, that

$$\rho_1(\lambda)=\begin{cases}\dfrac{1}{N}\left\{\dfrac{1}{2}[\delta(\lambda+2J)+\delta(\lambda-2J)]-\dfrac{1}{2\pi(4J^2-\lambda^2)^{1/2}}\right\} & |\lambda|<2J\\ 0 & |\lambda|>2J.\end{cases} \tag{4.14}$$

Thus to order N^{-1} the total averaged eigenvalue spectrum is given by

$$\rho(\lambda)=\rho_0(\lambda)+\rho_1(\lambda)$$
$$=\begin{cases}\dfrac{1}{2\pi J^2}(4J^2-\lambda^2)^{1/2}\left(1+\dfrac{1}{N}\dfrac{J^2}{(\lambda^2-4J^2)}\right)\\ \quad+\dfrac{1}{4N}(\delta(\lambda+2J)+\delta(\lambda-2J)) & |\lambda|\leqslant 2J\\ 0 & |\lambda|<2J.\end{cases} \tag{4.15}$$

We comment first on this result: it is easily verified that $\int_{-2J}^{+2J}\rho_1(\lambda)\,d\lambda=0$, so that to this order of approximation the spectrum remains correctly normalized to unity—the integrated contributions from the isolated delta functions at each band edge exactly cancelling part of the density function which diverges as $-(2J-\lambda)^{-1/2}$ near the upper band edge. We have already noted that we do not expect such a perturbation calculation to converge near the band edges. Although our result for $\rho_1(\lambda)$ is not in itself new, the method of calculation is novel and it is of interest because two recently published calculations of $\rho_1(\lambda)$ disagree with each other. All existing calculations predict the presence of a single eigenvalue at the top and bottom end of the Wigner semicircular band, however recent work by Takano and Takano (1984) based on earlier work by Takano *et al* (1983) which uses a direct graphical technique, predicts that the contribution to $\rho_1(\lambda)$, additional to the delta functions at the band edges, should be $-(\lambda^2-J^2)/2\pi NJ^2(4\lambda^2-\lambda^2)^{1/2}$ rather than the contribution $-1/2\pi N(4J^2-\lambda^2)^{1/2}$ which we have found. However Verbaarschot *et al* (1984), using a moment calculation, and Verbaarschot and Zirnbauer (1984) (vz) by casting the problem as a ϕ^3 field theory, both find a result which is in complete agreement with our own.

The second-order correction, of order N^{-2}, must be obtained from (4.9*c*) and (4.9*d*). Some lengthy algebra shows that we may write $\rho_2(\lambda)=\pi^{-1}\operatorname{Im}\mathscr{G}_2(\lambda)$ where $\mathscr{G}_2(\lambda)$ is a second-order correction to the Green function $\mathscr{G}_0(\lambda)$ and is given by

$$\mathscr{G}_2(\lambda)=\frac{1}{N^2}\left(-\frac{1}{4}\frac{iJ^2\lambda^2}{(4J^2-\lambda^2)^{5/2}}+\frac{1}{2}\frac{J^2\lambda}{(4J^2-\lambda^2)}+\frac{5}{4}\frac{iJ^2}{(4J^2-\lambda^2)^{3/2}}\right.$$
$$\left.-\frac{1}{2}\frac{\lambda}{(4J^2-2\lambda^2)}-\frac{i}{2}\frac{1}{(4J^2-\lambda^2)^{1/2}}\right). \tag{4.16}$$

The appropriate correction to the AED is then

$$\rho_2(\lambda)=\frac{1}{4N^2}(\delta(\lambda+2J)+\delta(\lambda-2J))$$
$$-\frac{1}{4\pi N^2}\left(\frac{J^2\lambda^2}{(4J^2-\lambda^2)^{5/2}}-\frac{5J^2}{(4J^2-\lambda^2)^{3/2}}+\frac{2}{(4J^2-\lambda^2)^{1/2}}\right). \tag{4.17}$$

This result is new. Once more we see that (4.17) displays the expected divergences when $|\lambda|=2J$, as did the first-order correction. However, it is easily verified that (6.17) is not correctly normalized. This is again unsurprising since we have no reason to expect that an expansion in powers of N^{-1} of a function, which, near $|\lambda|=2J$, is not analytic in N, should preserve the correct normalization properties when the integrals are dominated by the divergences at the band edges. Our final result is that to order N^{-2}, the averaged eigenvalue spectrum, $\hat{\rho}(\lambda)$, is

$$\hat{\rho}(\lambda)=\rho_0(\lambda)+\rho_1(\lambda)+\rho_2(\lambda)$$
$$=\frac{1}{2\pi J^2}(4J^2-\lambda^2)^{1/2}\left[1+\frac{1}{N}\left(1+\frac{1}{N}\right)\frac{J^2}{(\lambda^2-4J^2)}\right.$$
$$\left.+\frac{1}{N^2}\frac{5}{2}\frac{J^4}{(\lambda^2-4J^2)^2}+\frac{1}{N^2}\frac{1}{2}\frac{J^4\lambda^2}{(\lambda^2-4J^2)^3}\right]$$
$$+\frac{1}{4N}\left(1+\frac{1}{N}\right)(\delta(\lambda+2J)+\delta(\lambda-2J)). \tag{4.18}$$

This is the result we have sought to calculate. In section 6 we shall compare approximations discussed here with the results of numerical simulations of certain random matrices. It should be noted that our perturbative scheme yields non-zero corrections to $\rho_0(\lambda)$ only inside the Wigner semicircle band.

5. A self-consistent calculation of $\rho(\lambda)$ and the band edge problem

We have seen in section 4 that attempts to calculate $\rho(\lambda)$ in powers of N^{-1} using the perturbation expansion developed here, must diverge at the band edges—but these are precisely the places where we expect the AED to be small, even for finite N. The problem here is subtle and one is remined of the difficulties encountered in critical phenomena where naive expansions in the bare coupling constant of the problem always diverge at the critical point itself. The vanishing of $\rho(\lambda)$ at $|\lambda|=2J$ as $N\to\infty$ is rather similar to such problems of critical phenomena: in these latter problems it is necessary to rescale at the critical point and in the same spirit, we shall attempt to describe the band edges by using a scaling procedure and an 'effective action' to describe the problem. In doing so we shall be led to a self-consistency problem which will yield an approximate (and finite) expression for $\rho(\lambda)$ which gives a reasonable description of the AED for large finite N for all λ.

We now cast our problem into a form very similar to that of a problem in statistical mechanics or field theory (see, for example, Amit (1978) or Ramond (1981)). The averaged eigenvalue density is first formally written as

$$\rho(\lambda)=\frac{2}{N\pi}\operatorname{Im}\frac{\partial}{\partial\lambda}\ln\int e^{A(s,\lambda)}\,ds. \tag{5.1}$$

The saddle point of the integrand at $s = s_c$ is defined by the condition

$$\left(\frac{\partial}{\partial s} A(s, \lambda)\right)_{s=s_c} = 0$$

and the function $A(s, \lambda)$ is chosen so that the expression

$$\rho(\lambda) = -\frac{2}{N\pi} \operatorname{Im} \frac{\partial}{\partial \lambda} \ln e^{A(s_c)} \tag{5.2}$$

yields the correct AED to order N^{-1} i.e. it gives the approximation of (4.15). By comparison with (4.9*a*) and (4.9*b*) we see that we must choose

$$A(s, \lambda) = -\frac{N}{2} \ln \lambda - Ng(s, \lambda) + \ln \lambda + f(s, \lambda) - \tfrac{1}{2} \ln\left(\frac{\partial^2 g(s, \lambda)}{\partial s^2}\right) \tag{5.3}$$

where $g(s, \lambda)$ and $f(s, \lambda)$ are defined by (3.27*a*) and (3.27*b*) respectively.

In a field theoretical context, $A(s, \lambda)$ would be called the effective action to order N^{-1} and its saddle point s_c is known as the classical solution of this effective action to order N^{-1}. The saddle point condition is

$$\left\{-Ng'(s, \lambda) + f'(s, \lambda) - \frac{1}{2} \frac{g'''(s, \lambda)}{g''(s, \lambda)}\right\}_{s=s_c} = 0 \tag{5.4}$$

and the prime indicates a partial derivative with respect to s. From the definitions (3.27*a*) and (3.27*b*) respectively we may verify that $f'(s, \lambda) = \frac{1}{4} g'''(s, \lambda)/g''(s, \lambda)$ so that the saddle point condition becomes

$$\left\{+Ng'(s, \lambda) + \frac{1}{4} \frac{g'''(s, \lambda)}{g''(s, \lambda)}\right\}_{s=s_c} = 0. \tag{5.5}$$

Using the definition of $g(s, \lambda)$ we see straightforwardly that

$$-\frac{\lambda}{J} s_c - \frac{1}{\lambda/J - (-\lambda s_c/J)} = \frac{1}{N}\left(\frac{(\lambda/J - (-\lambda s_c/J))^{-3}}{1 - (\lambda/J - (-\lambda s_c/J))^{-2}}\right). \tag{5.6}$$

This equation cannot of course be solved for s_c in a simple closed form for arbitrary values of the ratio λ/J. However when $N \to \infty$, we readily see that the solution of (5.6) tends to the value s_0 given in equation (4.7).

Now an attempt to solve (5.6) iteratively will certainly yield $s_c = s_0 +$ (terms of order N^{-1}) and this latter correction of order $1/N$ will yield corrections to the semicircle in a self-consistent manner. Within the spirit of an effective action field theory and by analogy with (4.7) and (4.10) (in which we show that $\mathcal{G}_0(\lambda) = -(\lambda/J^2)s_0$) we solve (5.6) for s_c and then define a self-consistently determined propagator $\mathcal{G}(\lambda)$ by the relation $\mathcal{G}(\lambda) = -(\lambda/J^2)s_c$. We note that when $N \to \infty$, our previous remarks imply that $\mathcal{G}(\lambda)$ tends to $\mathcal{G}_0(\lambda)$.

Using this ansatz we see that (5.6) can be written as

$$\mathcal{G}(\lambda) - \frac{1}{J^2[(\lambda/J^2) - \mathcal{G}(\lambda)]} = \frac{1}{N}\left(\frac{[\lambda/J^2 - \mathcal{G}(\lambda)]^{-3}}{J^2 - [(\lambda/J^2) - \mathcal{G}(\lambda)]^{-2}}\right). \tag{5.7}$$

We may regard (5.7) as a self-consistency condition which determines the propagator $\mathcal{G}(\lambda)$. From it, we may calculate an averaged eigenvalue density $\bar{\rho}(\lambda)$ by the relation

$$\bar{\rho}(\lambda) = (1/\pi) \operatorname{Im} \mathcal{G}(\lambda). \tag{5.8}$$

This is our basic expression which determines $\bar{\rho}(\lambda)$ and in section 6 we shall compare the results obtained from (5.7) with those given by numerical simulation of a random matrix ensemble. It should be noted that with $J^2=1$, (5.7) has exactly the same form as equation (3.23) of VZ who derive their self-consistency condition by treating the band edge problem as a one loop correction to a ϕ^3 field theory.

The shape of the spectrum near the band edge can now be derived using arguments closely parallel to those given by VZ. Close to the upper band edge at $\lambda=2J$ we may write $\lambda/J=2+\delta$, where δ is small. Equation (4.5) shows that the saddle point is shifted from s_0 to s_c where s_c-s_0 is of order N^{-1} and s_0 has the value $-\frac{1}{2}$ at the upper band edge (see (4.7)). In the same spirit, near the upper band edge we may write $-(\lambda/J)s_c=1+p$, where p is small. When this ansatz is substituted with (5.6) it yields the result that

$$p^2-\delta(1+p)=\frac{1}{2N[p-\frac{1}{2}p^2-\frac{1}{2}\delta^2-\delta(1-p)]} \tag{5.9}$$

which is identical to the self-consistency problem obtained by VZ in their equation (3.25). The remainder of our calculations closely follows the work of these authors. We define scaled variables $\bar{p}$ and $\bar{\delta}$ which are to be independent of N, by the relations $p\equiv\bar{p}\,(2N)^{-\alpha}$ and $\delta\equiv\bar{\delta}(2N)^{-\beta}$, with α and β both positive. Since we seek the behaviour of $\rho(\lambda)$ close to the band edge when N is very large we must rewrite (5.9) in terms of the scaled variables $\bar{p}$ and $\bar{\delta}$; by considering (5.9) for N very large we find that $\alpha=\beta/2=\frac{1}{3}$ so that in this limit we have $\bar{p}=p(2N)^{-1/3}$ and $\bar{\delta}=\delta(2N)^{-2/3}$. When N is large and $\bar{\delta}\ll 1$, the self-consistency condition (5.9) then reduces to

$$\bar{p}^2-\bar{\delta}=(\bar{p})^{-1}. \tag{5.10}$$

This cubic equation determines $\bar{p}(\bar{\delta})$ and has the approximate solution

$$\bar{p}(\bar{\delta})=\bar{p}(0)+\bar{\delta}\left[\frac{\mathrm{d}\bar{p}}{\mathrm{d}\bar{\delta}}\right]_{\bar{\delta}=0}. \tag{5.11}$$

When $\bar{\delta}=0$, we choose the root of (5.10) for which $\bar{p}(0)=\mathrm{e}^{2\pi\mathrm{i}/3}$ and $[\mathrm{d}\bar{p}/\mathrm{d}\bar{\delta}]_{\bar{\delta}=0}=\frac{1}{3}\,\mathrm{e}^{-2\pi\mathrm{i}/3}$. It is easily verified that the remaining two roots lead either to a vanishing or to a negative AED at the band edge. The AED near the band edge is then determined by writing

$$\rho\left(\lambda=2J+2J\frac{\bar{\delta}}{(2N)^{2/3}}\right)=\frac{1}{\pi}\,\mathrm{Im}\,\mathscr{G}\left(\lambda=2J+2J\frac{\bar{\delta}}{(2N)^{2/3}}\right). \tag{5.12}$$

Thus

$$\begin{aligned}\rho\left(\lambda=2J+2J\frac{\bar{\delta}}{(2N)^{2/3}}\right)&=\frac{1}{\pi J}\,\mathrm{Im}\left[-\frac{\lambda}{J}s_c\right]_{\lambda=2J+2J\delta}\\&=\frac{1}{\pi J}\,\mathrm{Im}\left[1+\frac{\bar{p}(\bar{\delta})}{(2N)^{1/3}}\right].\end{aligned} \tag{5.13}$$

Use of (5.11) and the approximate solution for $\bar{p}(\bar{\delta})$ immediately shows that

$$\rho\left(\lambda=2J+2J\frac{\bar{\delta}}{(2N)^{2/3}}\right)=\frac{1}{\pi J(2N)^{1/3}}\left(\frac{\sqrt{3}}{2}-\frac{\sqrt{3}}{6}\,\bar{\delta}\right) \tag{5.14}$$

which agrees precisely with the result of VZ.

It is interesting to note that for the particular case $N=2$, and $J^2=0.5$, (5.14) yields $\bar{\rho}(\lambda=2J)=0.246$ as compared with the exact result (Porter and Rosenzweig (1960), see also section 6) of $\rho(\lambda = 2J) = 0.137$. For such small values of N we should not, of course, expect good numerical agreement, but expression (5.14) which is clearly non-analytic in N shows quite explicitly that any attempt to determine the AED in inverse powers of N must fail at the band edges precisely because of this non-analytic behaviour.

These problems have been encountered before by Zeigler (1982) who sought to calculate the density of states of a disordered electronic system with n orbitals per site: when $n \to \infty$ a semicircular band is again found. Attempts to calculate the asymptotic corrections (in inverse powers of n) to this semicircle using perturbation theory give sensible non-zero corrections inside the semicircular band, divergence at the band edge, and no correction outside the band. Close to the band edges however, a sophisticated field theoretical calculation shows a scaling behaviour which is non-analytic in n for a hypercubic lattice of dimension $d(<2)$ when n is large. Indeed, Ziegler's density function behaves as $n^{-1/3}$ for a system in which $d = 0$. This agreement is most satisfactory since our whole calculational technique uses 'functions' $\{x_i^\alpha\}$ which do not depend on any parameter, i.e. we are in effect working with a form of field theory in zero dimensions.

The behaviour of the eigenvalue spectrum near the band edges seems first to have been described by Bronk (1964) who gives an explicit form for the density function near the band edge for an ensemble of complex Hermitian matrices, but who does not quote such an expression for the Gaussian orthogonal ensemble. However, we note that the density function (5.14) decays to zero when $\bar{\delta}$ is of order unity. The number of eigenvalues outside the semicircular band is then obtained by estimating the value of $E \equiv N \int_{2J}^{\infty} \rho(\lambda)\, d\lambda$. It is easy to see from (5.14) that $E \sim N(N^{-1/3})(N^{-2/3})$, i.e. E is a constant independent of N itself. This is in accord with Bronk's calculation which shows that for both real symmetric and Hermitian ensembles of matrices, the number of eigenvalues in the tail of the density function, outside the band edge, is independent of the size N of the matrix.

6. Some numerical results

In this section we compare the closed form approximations given in section 4 and section 5 for the AED of a Gaussian orthogonal ensemble of $N \times N$ matrices with the results produced by numerical simulation of these ensembles. For a given N we have diagonalized a large sample of matrices and calculated an AED in the form of a histogram for the case $J^2 = 0.5$ (for convenience). We have included all eigenvalues between $\lambda = \pm 2.5$. In figure 1 we show the AED produced by diagonalizing a sample of 250 matrices for the case $N = 100$ (see Jones and Dhesi 1990) and it is easily seen, as a check, that the AED of such an ensemble is well described by the Wigner semicircle (3.7). In the case $N = 2$ it has been shown by Porter and Rosenzweig (1960) that the AED can be calulated exactly; when $J^2 = 0.5$ their result has the form

$$\rho_{N=2}(\lambda) = \frac{1}{\sqrt{(2\pi)}} e^{-\lambda^2}(e^{-\lambda^2} + \sqrt{\pi}\, \lambda\, \mathrm{erf}(\lambda)). \tag{6.1}$$

In figure 2 we compare the exact result (6.1) with the simulations for a sample of size 1.5×10^5 and again with the Wigner semicircle. This simply provides another check on the accuracy of the numerical simulation and clearly shows that the Wigner semicircle

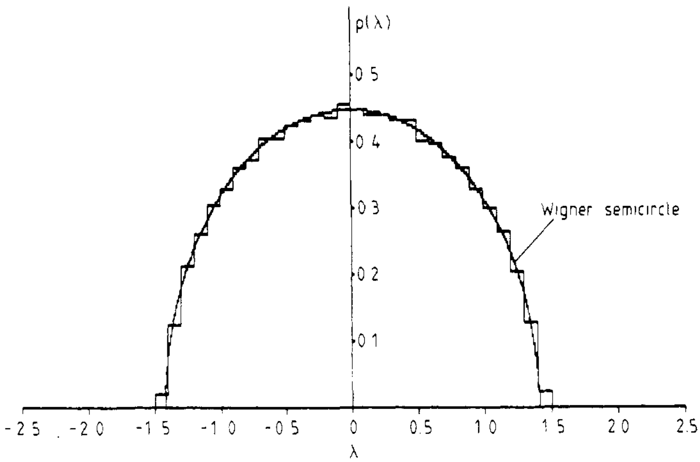


Figure 1. $N = 100$. The diagram compares the Wigner semicircle with the numerical simulations shown on the histogram (sample size = 250).

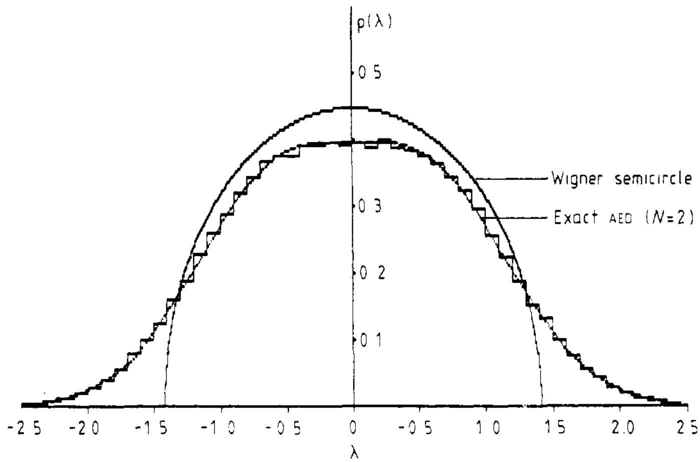


Figure 2. $N = 2$. The diagram compares the exact analytical result for the AED with the Wigner semicircle and the numerical simulations shown on the histogram (sample size = 150 000).

does not well describe the AED of an ensemble of small matrices: the semicircle clearly gives too large an AED near $\lambda = 0$ but then vanishes for $|\lambda| > \sqrt{2}$, whereas the true AED has a small but finite tail which extends to infinity in both directions. Similar features are readily observable for $N = 3$, 5 and 10 although we have not reproduced the relevant data in this section. Figures 3, 4, 5 and 6 contain the numerical AED for $N = 2$, 3, 5 and 10 respectively and compare them with the approximation schemes developed in sections 4 and 5. In section 4 we developed an expansion for $\rho(\lambda)$ in inverse powers of N and the resulting approximation as far as terms in N^{-2} is given by the expression for $\hat{\rho}(\lambda)$ in (4.18). We have pointed out that $\hat{\rho}(\lambda)$ diverges at the Wigner band edges $|\lambda| = \sqrt{2}$ where such an expansion cannot be expected to converge. The self-consistent approximation $\bar{\rho}(\lambda)$, developed in section 5, is expected to give a better overall

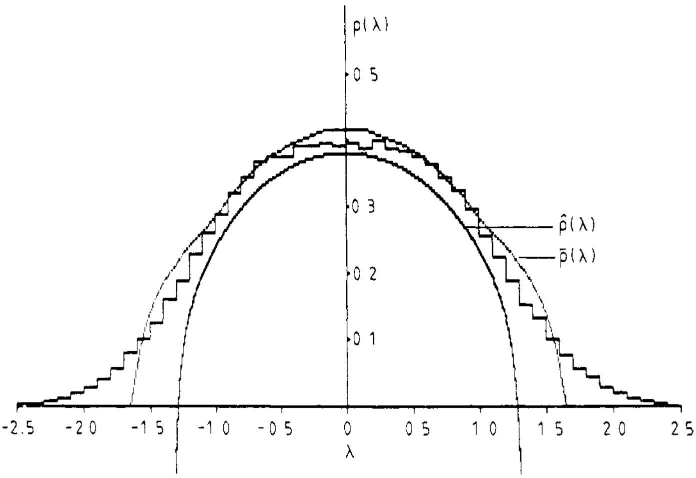


Figure 3. $N = 2$. The diagram compares the self-consistent theory, $\bar{\rho}(\lambda)$, with the perturbative result $\hat{\rho}(\lambda)$ and the numerical simulations shown on the histogram (sample size = 150 000).

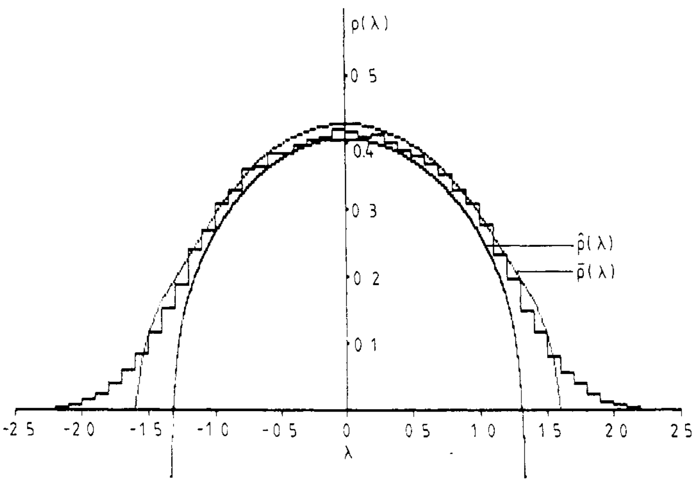


Figure 4. $N = 3$. The diagram compares the self-consistent theory, $\bar{\rho}(\lambda)$, with the perturbative result $\hat{\rho}(\lambda)$ and the numerical simulations shown on the histogram (sample size = 150 000).

description of the AED. It is obtained by solving (5.7) numerically (for a given value of λ) for its real and imaginary parts and then using (5.8) to calculate the self-consistent approximations $\bar{\rho}(\lambda)$. On figures 3, 4, 5 and 6 we display, for $N = 2$, 3, 5 and 10, both the self-consistent approximation $\bar{\rho}(\lambda)$ and the perturbative approximation $\hat{\rho}(\lambda)$ (shorn of both delta functions). Some general points may be noted. In general the perturbative approximation, $\hat{\rho}(\lambda)$ lies closer to the numerical AED near to $\lambda = 0$, where it still nevertheless underestimates the AED by a small quantity which decreases as N increases. By comparison the self-consistent approximation $\bar{\rho}(\lambda)$ always lies above $\hat{\rho}(\lambda)$ close to $\lambda = 0$. This is consistent with our earlier remarks that $\hat{\rho}(\lambda)$ is expected to be a good

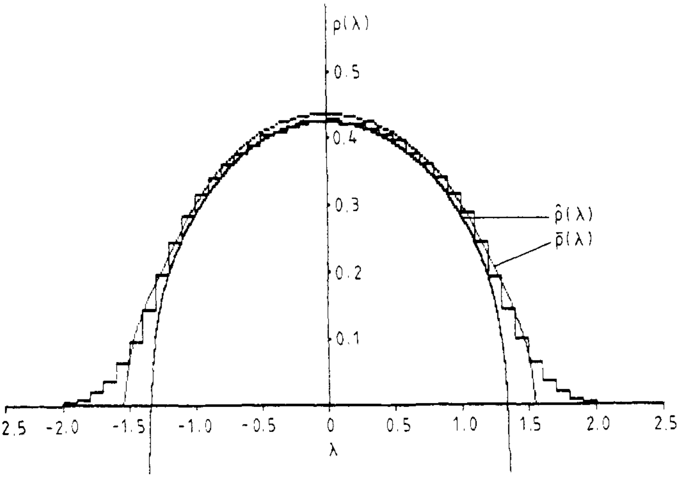


Figure 5. $N = 5$. The diagram compares the self-consistent theory, $\bar{\rho}(\lambda)$, with the perturbative result $\hat{\rho}(\lambda)$ and the numerical simulations shown on the histogram (sample size = 150 000).

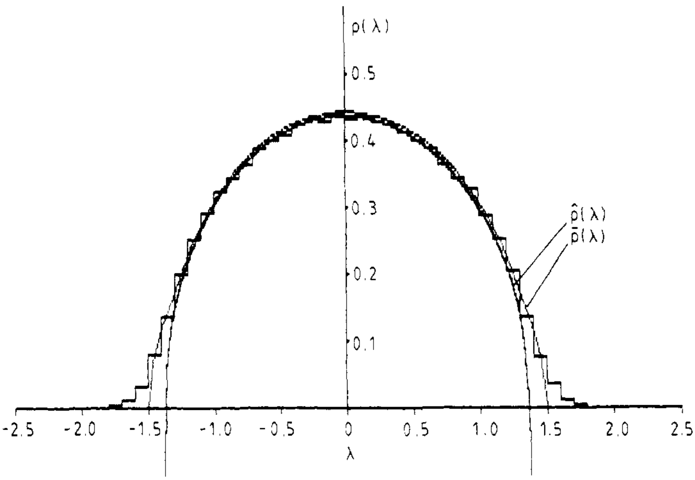


Figure 6. $N = 10$. The diagram compares the self-consistent theory, $\bar{\rho}(\lambda)$, with the perturbative result $\hat{\rho}(\lambda)$ and the numerical simulations shown on the histogram (sample size = 30 000).

approximation to the AED only near the centre of the band of eigenvalues. However close to the edges of the Wigner band at $|\lambda| = \sqrt{2}$, $\hat{\rho}(\lambda)$ diverges negatively and is strictly zero for $|\lambda| = \sqrt{2}$.

In this region the self-consistent calculation $\bar{\rho}(\lambda)$ plainly gives a better description of the behaviour of the AED which is free of all divergences. This approximation, $\bar{\rho}(\lambda)$, clearly exhibits a tail of states which extends outside the Wigner band edges but which is finite in extent—whereas the numerical AED has a finite tail for all λ and does not cut off anywhere. We see again that as N increases, the points where $\bar{\rho}(\lambda)$ vanishes become ever closer to the Wigner band edges, whilst near $\lambda = 0$ both approximations approach the numerical AED. Although we have not reproduced the data here, it may be seen that when $N=20$, $\bar{\rho}(\lambda)$, $\hat{\rho}(\lambda)$ and the Wigner AED are virtually indistinguishable. We should not of course expect precise numerical agreement between the histograms and our approximations $\bar{\rho}(\lambda)$ and $\hat{\rho}(\lambda)$ for very small values of N; it is interesting to note that even when for $N=2$ the self-consistent calculation $\bar{\rho}(\lambda)$, overestimates the exact AED by only 5% at $\lambda=0$, whilst at $\lambda=0$ the perturbative result, $\hat{\rho}(\lambda)$, underestimates the exact value by about 3.9%. From the histograms and detailed study of the values of $\bar{\rho}(0)\hat{\rho}(0)$, we find that when $N=3$ the overestimate in $\bar{\rho}(0)$ is about 2.7% and the underestimate in $\hat{\rho}(0)$ is about 2.5%. For $N=10$, the corresponding figures are 2.2% and 0.9%. Near the upper band edge of the Wigner semicircle at $\lambda=\sqrt{2}$, $\hat{\rho}(\sqrt{2})$ diverges negatively and the start of this divergence is visible on figures 3 to 6. However the self-consistent result for $\bar{\rho}(\sqrt{2})$ can been in all cases to be close to the AED generated by numerical simulation. However quantitative agreement between $\bar{\rho}(\lambda)$ and the AED is not good for values of λ greater than about 1.5 where the tail of states (described in section 6) predicted by $\bar{\rho}(\lambda)$ does not extend to sufficiently large values of λ.

7. Summary

We have shown how the replica method, in the form developed by Edwards and Jones (1976), can be used to develop a form of perturbation theory which for the Gaussian orthogonal ensemble will lead to corrections to the Wigner semicircular AED which are of order $1/N$ and $1/N^2$. This latter result is new. However when we compare our calculation of the $1/N$ corrections to the semicircle with previously published conflicting calculations, we find that our results are in complete agreement with those of Verbaarschot and Zirnbauer (1984), who used a different calculational framework. We have also shown how our version of the replica symmetric $n \to 0$ method yields a very straightforward method of calculating the average eigenvalue density in a self-consistent fashion and which is finite everywhere and gives a good overall picture of the AED of an ensemble in which N is finite. This latter calculation permits a very straightforward method for calculating the scaling behaviour of the AED near the Wigner band edges where perturbation theory fails. Again our results agree with those of VZ. We have presented some numerical data which compares our two approximation schemes for the AED with numerically simulated histrograms of Gaussian orthogonal ensembles of different sized matrices.

The problems associated with using the replica method and of assuming replica symmetry are by now well known: de Almeida and Thouless (1978) have shown the need to break the replica symmetry to attempt a sensible description of a spin glass and their view is supported by the work of Parisi (1979) which proposes a more sophisticated symmetry breaking scheme for the spin glass. In an important recent paper, Verbaarschot and Zirnbauer (1985) have shown that the replica method with replica symmetry fails to give the correct result for the two point correlation function of the Gaussian unitary ensemble; however these authors show that the introduction of mixed bosonic and fermionic variables will yield the correct non-perturbative form for the two point function.

Work is now in progress in which we hope to shed some light on the reasons for the failure of the simple replica symmetric techniques for the two point correction function in the Gaussian orthogonal ensemble.

Acknowledgments

The authors are pleased to acknowledge helpful discussions with Dr J Gunson, Dr G P McCauley and Professor M A Moore. Help with some of the numerical work from Dr G P McCauley is gratefully acknowledged. One of the authors (GSD) wishes to acknowledge the support of an SERC Studentship held whilst this work was carried out.

Appendix A

Consider the integral

$$E_N \equiv \int_C \mathrm{e}^{NG(s)}\,\mathrm{d}s. \qquad \text{(A.1)}$$

The integrand has a saddle point at $s=\bar{s}$ where $G'(\bar{s})=0$. Then defining the quantity

$$-u^2 \equiv \frac{1}{2!}(s-\bar{s})^2 G''(\bar{s}) + \frac{1}{3!}(s-\bar{s})^3 G'''(\bar{s}) + \ldots \qquad \text{(A.2)}$$

we have

$$E_N = \mathrm{e}^{NG(\bar{s})} \int_{-\infty}^{+\infty} \mathrm{e}^{-Nu^2} \frac{\mathrm{d}s}{\mathrm{d}u}\,\mathrm{d}u \qquad \text{(A.3)}$$

The process of reversion of the series (A.2) is lengthy but straightforward and gives

$$\frac{\mathrm{d}s}{\mathrm{d}u} = \alpha + 2\beta u + 3\gamma u^2 + \ldots \qquad \text{(A.4)}$$

where

$$\alpha = \sqrt{\frac{1}{A}} \qquad \beta = \frac{B}{2A^2} \qquad \gamma = \frac{1}{2}\sqrt{-A}\left\{\frac{5B^2}{4A^4} - \frac{C}{A^3}\right\}$$

and

$$A = \frac{G''(\bar{s})}{2!} \qquad B = \frac{G'''(\bar{s})}{3!} \qquad C = \frac{G^{(\mathrm{iv})}(\bar{s})}{4!}.$$

The integrals implied by (A.3) and (A.4) are easily performed and give

$$E_N = \sqrt{\frac{-2\pi}{NG''(\bar{s})}}\,\mathrm{e}^{NG(\bar{s})}\left[\frac{1}{2N}\left(\frac{1}{4}\frac{G^{(\mathrm{iv})}(\bar{s})}{(G''(\bar{s}))^2} - \frac{5}{12}\frac{G'''(\bar{s})}{(G''(\bar{s}))^3}\right) + \mathrm{O}(N^{-2})\right]. \qquad \text{(A.5)}$$

Appendix B

In (A.5) we write

$$G = -g(s,\lambda) + \frac{1}{N} f(s,\lambda) + \frac{1}{N^2} h(s,\lambda). \qquad \text{(B.1)}$$

For convenience, the λ dependence of g, f and h will not be written explicitly. Now $g(s)$ has a saddle point at $s=s_0$ where $g'(s_0)=0$; it is easily seen that $G(s)$ has a saddle point at $s=\bar{s}$ where $G'(s_0)=0$ and

$$\bar{s}=s_0+\frac{1}{N}\frac{f'(s)}{g''(s_0)}+\mathrm{O}(N^2). \tag{B.2}$$

We use (A.5), (B.1) and (B.2) to determine the asymptotic dependence of E_N on N.

Taylor's theorem is used to expand $G(s)$ and its derivatives about $s=s_0$ and we thereby generate a series of corrections to the standard saddle point result in a series of inverse powers of N.

Performing these manipulations yields

$$E_N=\left(\frac{2\pi}{Ng''(s_0)}\right)^{1/2}\exp\left(-Ng(s_0)+f(s_0)+\frac{1}{N}h(s_0)+\mathrm{O}(N^{-2})\right)$$
$$\times\left\{1+\frac{1}{N}\left[-\frac{1}{8}\frac{g^{(\mathrm{iv})}(s_0)}{(g''(s_0))^2}+\frac{5}{24}\frac{(g'''(s_0))^2}{(g''(s_0))^3}+\left(\frac{f''(s_0)+(f'(s_0))^2}{2g''(s_0)}\right)\right.\right.$$
$$\left.\left.-\frac{f'(s_0)(g'''(s_0))}{2(g''(s_0))^2}\right]+\mathrm{O}(N^{-2})\right\}. \tag{B.3}$$

Thus with $W_N(\lambda)\equiv\ln E_N$ we finally have

$$W_n(\lambda)=\frac{1}{2}\ln\left(\frac{2\pi}{N}\right)-Ng(s_0)-\frac{1}{2}\ln g''(s_0)+f(s_0)$$
$$+\frac{1}{N}\left[h(s_0)+\left(-\frac{1}{8}\frac{g^{(\mathrm{iv})}(s_0)}{(g''(s_0))^2}+\frac{5}{24}\frac{(g'''(s_0))^2}{(g''(s_0))^3}+\frac{f''(s_0)}{2g''(s_0)}\right.\right.$$
$$\left.\left.+\frac{(f'(s_0))^2}{2g''(s_0)}-\frac{f'(s_0)g'''(s_0)}{2(g''(s_0))^2}\right)\right]+\mathrm{O}(N^{-2}). \tag{B.4}$$

This is the result used in (4.6) which we we wished to establish.

References

Amit D J 1978 *Field Theory, the Renormalization Group and Critical Phenomena* (New York: McGraw-Hill)

Brody T A, Flores J, French J B, Mello P A, Pandey A and Wong S S M 1981 *Rev. Mod. Phys.* **53** 385-479

Bronk B V 1964 *J. Math. Phys.* **5** 215-20

Bohigas O and Giannoni M J 1983 *Chaotic Motion and Random Matrix Theories* (*Lecture Notes in Physics* **209**) (Berlin: Springer)

Camarda H S and Georgopulas P D 1983 *Phys. Rev. Lett.* **50** 492-5

de Almeida J R L and Thouless D J 1978 *J. Phys. A: Math. Gen.* **13** 983-90

Dyson F J 1962 *J. Math. Phys.* **3** 140-56

Edwards S F 1970 *Proc. 4th Int. Conf. on Amorphous Materials* ed R W Douglas and B Ellis (New York: Wiley)

Edwards S F and Jones R C 1976 *J. Phys. A: Math. Gen.* **9** 1595-603

Edwards S F and Warner M 1980 *J. Phys. A: Math. Gen.* **13** 381-96

Jones R C and Dhesi G S 1990 *Can. J. Phys.* in press

Jones R C, Kosterlitz J M and Thouless D J 1978 *J. Phys. A. Math. Gen.* **11** L45-8

Kosterlitz J M, Thouless D J and Jones R C 1976 *Phys. Rev. Lett.* **36** 1217-20

Mehta M 1967 *Random Matrices and the Statistical Theory of Energy Levels* (New York: Academic)

Mukamel S, Sue J and Pandey A 1984 *Chem. Phys. Lett.* **105** 134-9

Parisi G 1979 *Phys. Rev. Lett.* **43** 1754-6

Porter C E 1965 *Statistical Theories of Spectra: Fluctuations* (New York: Academic)

Porter C E and Rosenzweig N 1960 *Ann. Acad. Sci. Fennicae* A **VI** 3-66
Ramond P 1981 *Field Theory: A Modern Primer* (New York: Benjamin)
Sherrington D 1971 *J. Phys. C: Solid State Phys.* **4** 401-16
Takano F and Takano H 1984 *J. Phys. Soc. Japan* **53** 2943-55
Takano F, Tamaribuchi T and Oguchi T 1983 *J. Phys. Soc. Japan* **52** 3341-8
Verbaarschot J J M and Zirnbauer M R 1984 *Ann. Phys.* **158** 78-119
—— 1985 *J. Phys. A: Math. Gen.* **17** 1093-09
Verbaarschot J J M, Zirnbauer M R and Weidenmüller H A 1984 *Ann. Phys.* **153** 367-88
Wigner E P 1955 *Ann. Math.* **62** 548-64
—— 1957 *Can. Math. Congr. Proc.* Toronto: University of Toronto Press 174-84
—— 1958 *Ann. Math. Lpz.* **67** 226-7
Whittle P 1970 *Probability* (London: Penguin Educational) pp 167-73
Ziegler K 1982 *Z. Phys.* B **48** 293-304