RANDOM MATRICES and the Statistical Theory of Energy Levels

M. L. MEHTA

DEPARTMENT OF PHYSICS AND ASTROPHYSICS
UNIVERSITY OF DELHI
DELHI, INDIA

1967

ACADEMIC PRESS New York and London

ACADEMIC PRESS INC.
111 Fifth Avenue, New York, New York 10003

United Kingdom Edition published by
ACADEMIC PRESS INC. (LONDON) LTD.
Berkeley Square House, London W.1

LIBRARY OF CONGRESS CATALOG CARD NUMBER: 67-23169

PRINTED IN THE UNITED STATES OF AMERICA

Preface

Though random matrices were first encountered in mathematical statistics by Hsu, Wishart, and others, intensive study of their properties in connection with nuclear physics began with the work of Wigner in the 1950's. Much material has accumulated since then, and it was felt that it should be collected. A reprint volume to satisfy this need has been edited by C. E. Porter with a critical introduction (see References); nevertheless, the feeling was that a book containing a coherent treatment of the subject would be welcome.

We make the assumption that the local statistical behavior of the energy levels of a sufficiently complicated system is simulated by that of the eigenvalues of a random matrix. Chapter 1 is a rapid survey of our understanding of nuclear spectra from this point of view. The discussion is rather general, in sharp contrast to the precise problems treated in the rest of the book. In Chapter 2 an analysis of the usual symmetries that a quantum system might possess is carried out, and the joint probability density function for the various matrix elements of the Hamiltonian is derived as a consequence. The transition from matrix elements to eigenvalues is made in Chapter 3 and the standard arguments of classical statistical mechanics are applied in Chapter 4 to derive the eigenvalue density. An unproved conjecture is also stated. In Chapter 5 the method of integration over alternate variables is presented, and an application of the Fredholm theory of integral equations is made to the problem of eigenvalue spacings. The methods developed in Chapter 5 are basic to an understanding of most of the remaining chapters. Chapter 6 deals with the correlations and spacings for less useful cases. A Brownian motion model is described in Chapter 7. Chapters 8 to 11 treat circular ensembles; Chapters 8 to 10 repeat calculations analogous to those of Chapters 4 to 7. The integration method discussed in Chapter 11 originated with Wigner and is being published here for the first time. The theory of non-Hermitian random matrices,

though not applicable to any physical problems, is a fascinating subject and must be studied for its own sake. In this direction an impressive effort by Ginibre is described in Chapter 12. For the Gaussian ensembles the level density in regions where it is very low is discussed in Chapter 13. The investigations of Chapter 16 and Appendices A.29 and A.30 were recently carried out in collaboration with Professor Wigner at Princeton University. Chapters 14, 15, and 17 treat a number of other topics. Most of the material in the appendices is either well known or was published elsewhere and is collected here for ready reference. It was surprisingly difficult to obtain the proof contained in A.21, while A.29, A.30, and A.31 are new.

It is my pleasant duty to thank Professor C. Bloch, Professor F. J. Dyson, and Professor E. P. Wigner to whom I owe so much by way of education and inspiration. I have made use of the cited literature and in particular published as well as unpublished works of E. P. Wigner, F. J. Dyson, and M. Gaudin. I am thankful to the editors of the various publications for allowing me to do so. This book was written in sections at Tata Institute of Fundamental Research, Bombay, the Indian Institute of Technology, Kanpur, Delhi University, Argonne National Laboratory, and Princeton University. The lectures given at the State University of New York at Stony Brook, Long Island, were helpful in the initial stages. I am grateful to all these institutions for their hospitality. My thanks are due to my colleagues H. S. Mani, N. Rosenzweig, and P. K. Srivastava for their critical comments. A few additions and changes were made at almost every stage of the process of publication and I am thankful to the staff of Academic Press for their cooperation.

October, 1967 M. L. MEHTA
Saclay, France

1 / Introduction

1.1. The Need to Study Random Matrices

The experimental nuclear physicists have been and still are collecting vast amounts of data concerning the excitation spectra of various nuclei [Garg *et al.*, 1; Rosen *et al.*, 1]. The ground state and low-lying excited states have been impressively explained in terms of an independent particle model where the nucleons are supposed to move freely in an average potential well [Mayer and Jensen, 1; Kisslinger and Sorenson, 1]. As the excitation energy increases, more and more nucleons are thrown out of the main body of the nucleus, and the approximation of replacing their complicated interactions with an average potential becomes more and more inaccurate. At still higher excitations the nuclear states are so dense and the intermixing is so strong that it is a hopeless task to try to explain the individual states; but when the complications increase beyond a certain point the situation becomes hopeful again, for we are no longer required to explain the characteristics of every individual state but only their average properties, which is much simpler.

The average behavior of the various energy levels is of prime importance in the study of nuclear reactions. In fact, nuclear reactions may be put into two major classes—fast and slow. In the first case a typical reaction time is of the order of the time taken by the incident nucleon to pass through the nucleus. The wavelength of the incident nucleon is much smaller than the nuclear dimensions, and the time it spends inside the nucleus is so short that it interacts with only a few nucleons inside the nucleus. A typical example is the head-on collision with one nucleon in which the incident nucleon hits and ejects a nucleon, thus giving it almost all its momentum and energy. Consequently, in such cases the coherence and interference effects between incoming and outgoing nucleons are strong.

Another extreme is provided by the slow reactions in which the

typical reaction times are two to three orders of magnitude larger. The incident nucleon is trapped and all its energy and momentum are quickly distributed among the various constituents of the target nucleus. It takes a long time before enough energy is again concentrated on a single nucleon to eject it. The compound nucleus lives long enough to forget the manner of its formation, and the subsequent decay is therefore independent of the way in which it was formed.

In the slow reactions, unless the energy of the incident neutron is very sharply defined, a large number of neighboring energy levels of the compound nucleus are involved, hence the importance of an investigation of their average properties, such as the distribution of neutron and radiation widths, level spacings, and fission widths. It is natural that such phenomena, which result from complicated many-body interactions, will give rise to statistical theories. We shall concentrate mainly on the average properties of nuclear levels such as level spacings.

According to quantum mechanics, the energy levels of a system are supposed to be described by the eigenvalues of a Hermitian operator, called the Hamiltonian. The energy-level scheme of a system consists in general of a continuum and a certain, perhaps a large, number of discrete levels. The Hamiltonian of the system should have the same eigenvalue structure and therefore must operate in an infinite dimensional Hilbert space. To avoid the difficulty of working with an infinite dimensional Hilbert space, we make approximations amounting to a truncation keeping only the part of the Hilbert space that is relevant to the problem at hand and either forgetting about the rest or taking its effect in an approximate manner on the part considered. Because we are interested in the discrete part of the energy-level schemes of various quantum systems, we approximate the true Hilbert space by one having a finite, though large, number of dimensions. Choosing a basis in this space, we represent our Hamiltonians by finite dimensional matrices. If we can solve the eigenvalue equation,

$$H\Psi_i = E_i\Psi_i ,$$

we shall get all the eigenvalues and eigenfunctions of the system, and any physical information can then be deduced, in principle, from this knowledge. In the case of the nucleus, however, there are two difficulties. First, we do not know the Hamiltonian and, second, even if we did, it would be far too complicated to attempt to solve it.

Therefore from the very beginning we shall be making statistical hypotheses on H, compatible with the general symmetry properties. Choosing a complete set of functions as basis, we represent the Hamiltonian operators H as matrices. The elements of these matrices are random variables whose distributions are restricted only by the general symmetry properties we might impose on the ensemble of operators. Statistical theory does not predict the detailed level sequence of any one nucleus, but it does describe the general appearance and the degree of irregularity of the level structure that is expected to occur in any nucleus, which is too complicated to be understood in detail.

In classical statistical mechanics a system may be in any one of the many possible states, but one does not ask in which particular state a given system is. Here we shall renounce knowledge of the system itself. As in orthodox statistical mechanics we shall consider an ensemble of Hamiltonians, each of which could describe a different nucleus. There is a reasonable expectation, though no rigorous mathematical proof, that a system under observation will be described correctly by an ensemble average. This expectation is strong because the system might be one of a huge variety of systems, and very few of them will deviate much from a properly chosen ensemble average. On the other hand, our assumption that the ensemble average correctly describes a particular system, say the U^{239} nucleus, is not compelling. In fact, if this particular nucleus turns out to be far removed from the ensemble average, it will show that the U^{239} Hamiltonian possesses specific properties of which we are not aware. This, then, will prompt us to try to discover the nature and origin of these properties [Dyson, 1].

Wigner was the first to propose in this connection the hypothesis alluded to, namely that the local statistical behavior of levels in a simple sequence is identical with the eigenvalues of a random matrix. A simple sequence is one whose levels all have the same spin, parity, and other strictly conserved quantities, if any, which result from the symmetry of the system. The corresponding symmetry requirements are to be imposed on the random matrix. Porter and Rosenzweig were the early workers in the field who analyzed the nuclear experimental data made available by Hughes, Harvey, Rosen, and co-workers and the atomic data compiled by C. E. Moore [1]. They found that the occurrence of two levels close to each other in a simple sequence is a rare event. They also used the computer to generate and diagonalize a large number of random matrices. This Monte Carlo

analysis indicated the correctness of Wigner's hypothesis. In fact, it indicated more; the density and the spacing distribution of eigenvalues of real symmetric matrices are independent of many details of the distribution of individual matrix elements. All that is required is the same distribution for all diagonal elements and that the off-diagonal elements be distributed symmetrically about the zero mean and have the same mean square deviation. This independence is to be expected as well in the case of complex Hermitian or self-dual quaternion matrices, but apart from this numerical evidence and a few heuristic arguments of Wigner no rigorous derivation of this fact has yet been found. The case of the Gaussian distributions of matrix elements is still the only one treated analytically by Hsu, Mehta, Gaudin, Dyson, Bronk, Ginibre, and others, and we have described these developments in great detail in the following pages. From a group-theoretical analysis Dyson [5] found that an irreducible ensemble of matrices, invariant under a symmetry group G, necessarily belongs to one of three classes, named by him orthogonal, unitary, and symplectic. We shall not go into these elegant group-theoretical arguments but shall devote enough space to the study of the circular ensembles introduced by Dyson. It is remarkable that standard thermodynamics can be applied to obtain certain results which otherwise would be difficult to derive. A theory of the Brownian motion of matrix elements has also been created by Dyson thus rederiving a few known results. However, it remains largely a curiosity.

The physical properties of metals depend characteristically on their excitation spectra. In bulk metal at high temperatures the electronic energy levels lie very near to one another and are broad enough to overlap and form a continuous spectrum. As the sample gets smaller, this spectrum becomes discrete, and as the temperature decreases the widths of the individual levels decrease. If the metallic particles are minute enough and at low enough temperatures, the spacings of the electronic energy levels may eventually become much larger than the other energies, such as the level widths and the thermal energy kT. Under such conditions the thermal and the electromagnetic properties of the fine metallic particles may deviate considerably from those of the bulk metal. This circumstance has already been noted by Fröhlich [1] and proposed by him as a test of quantum mechanics. Because it is difficult to control the shapes of such small particles while they are being experimentally produced, the electronic energy levels are seen to be random and the theory for

the eigenvalues of random matrices may be useful in their study.

Random matrices are also encountered in other branches of physics. For example, glass may be considered as a collection of random nets, that is, a collection of particles with random masses exerting random mutual forces, and it is of interest to determine the distribution of frequencies of such nets [Dyson, 6]. A one-dimensional model of glass is considered in Chapter 14.

1.2. A Summary of Statistical Facts about Nuclear Energy Levels

1.2.1. Level Density

As the excitation energy increases, the nuclear energy levels occur on the average at smaller and smaller intervals. In other words, level density increases with the excitation energy. The first question we might ask is how fast does this level density increase for a particular nucleus and what is the distribution of these levels with respect to spin and parity? This is an old problem treated by Bethe [1]. Even a simple model in which the nucleus is taken as a degenerate Fermi gas with equidistant single-particle levels gives an adequate result. It amounts to determining the number of partitions $\lambda(n)$ of a positive integer n into smaller positive integers ν_1, ν_2,...

$$n = \nu_1 + \nu_2 + \cdots, \qquad \nu_1 > 0, \nu_2 > 0, \ldots .$$

For large n this number, according to the Hardy–Ramanujan [1] formula, is given by

$$\lambda(n) \sim \exp[(\tfrac{1}{3}\theta\pi^2 n)^{1/2}],$$

where θ is equal to 1 or 2 according to whether the ν_i are all different or whether some of them are allowed to be equal. With a slight modification due to later work [Lang and Lecouteur, 1; Cameron, 1], Bethe's result gives the level density as

$$\rho(E, j, \pi) \propto (2j+1)(E-\Delta)^{-5/4} \exp\left[-\frac{1}{2\sigma^2} j(j+1)\right] \exp[2a(E-\Delta)^{1/2}],$$

where E is the excitation energy, j is the spin, and π is the parity. The dependence of the parameters σ, a, and Δ on the neutron and proton numbers is complicated and only imperfectly understood.

However, for any particular nucleus a few measurements will suffice to determine them all; the formula will then remain valid for a wide range of energy that contains thousands and even millions of levels.

1.2.2. Distribution of Neutron Widths

An excited level may decay in many ways; for example, by neutron ejection or by giving out a quantum of radiation. These processes are characterized by the corresponding decay widths of the levels. The neutron reduced widths $\Gamma_n^0 = \Gamma_n/E^{1/2}$, in which Γ_n is the neutron width and E is the excitation energy of the level, show large fluctuations from level to level. From an analysis of the then available data Scott [1] and later Thomas and Porter [1] concluded that they had a χ^2-distribution with $\nu = 1$ degree of freedom:

$$P(x) = [\Gamma(\tfrac{1}{2}\nu)]^{-1}\, e^{-(1/2)\nu x}(\tfrac{1}{2}\nu x)^{(1/2)\nu-1} \cdot \tfrac{1}{2}\nu = (2\pi x)^{-1/2}\, e^{-(1/2)x},$$

where $P(x)\,dx$ is the probability that a certain reduced width will lie in an interval dx around the value x. This indicates a Gaussian distribution for the reduced width amplitude

$$\left(\frac{2}{\pi}\right)^{1/2} \exp[-\tfrac{1}{2}(\sqrt{x})^2]\, d(\sqrt{x})$$

expected from the theory. In fact, the reduced width amplitude is proportional to the integral of the product of the compound nucleus wave function and the wave function in the neutron-decay channel over the channel surface. If the contributions from the various parts of the channel surface are supposed to be random and mutually independent, their sum will have a Gaussian distribution with zero mean.

1.2.3. Radiation and Fission Widths†

The total radiation width is almost a constant for particular spin states of a particular nucleus. The total radiation width is the sum of partial radiation widths

$$\Gamma = \sum_{i=1}^{m} \Gamma_i\,.$$

† Bohr [1].

If we assume that each of these $\Gamma_i/\bar{\Gamma}_i$ has a χ^2-distribution with one degree of freedom like the neutron widths and all the $\bar{\Gamma}_i$ are the same, then $\Gamma/\bar{\Gamma}$ will have a χ^2-distribution with m degrees of freedom. Even if the $\bar{\Gamma}_i$ are different, we have

$$\Gamma = \sum_i \Gamma i$$

and

$$\overline{(\Gamma - \bar{\Gamma})^2} = 2 \sum_i (\bar{\Gamma}_i)^2,$$

so that for m large $\Gamma/\bar{\Gamma}$ has a narrow distribution. It is difficult to measure the partial radiation widths.

Little is known about the fission-width distributions. Some known fission widths of U^{235} have been analyzed [Bohr, 1] and a χ^2-distribution with 2 to 3 degrees of freedom has been found to give a satisfactory fit.

From now on we shall no longer consider neutron, radiation, or fission widths.

1.2.4. Level Spacings

Let us regard level density as a function of the excitation energy as known and consider an interval of energy δE centered at E. This interval is much smaller compared with E, whereas it is large enough to contain many levels; that is,

$$E \gg \delta E \gg D,$$

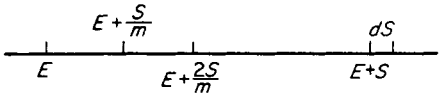
where D is the mean distance between neighboring levels. How are the levels distributed in this interval? Although the level density varies strongly from nucleus to nucleus, the fluctuations in the precise positions of the levels seem not to depend on the nucleus and not even on the excitation energy. As the density of the levels is nearly constant in this interval, we might think that they occur at random positions without regard to one another, the only condition being that their density be a given constant. However, such is not the case. It is true that nuclear levels with different spin and parity or atomic levels with different sets of good quantum numbers seem to have no influence on each other. However, levels with the same set of good quantum numbers show a large degree of regularity. For instance, they rarely occur close together.

A more detailed analysis of the experimental data regarding the above quantities as well as the strength functions may be found in Garg *et al.* [1] and Rosen *et al.* [1].

1.3. Definition of a Suitable Function for the Study of Level Correlations

To distinguish between various possibilities we define the distribution of spacings. Let $E_1, E_2, \ldots, E_n$ be the positions of the successive levels in the interval $\delta E (E_1 \leqslant E_2 \leqslant \cdots)$ and let $S_1, S_2, \ldots$ be their distances apart $S_i = E_{i+1} - E_i$. The average value of S_i is the mean spacing D. We define the relative spacings $t_i = S_i/D$. The probability density function $p(t)$ is defined by the condition that $p(t)\,dt$ is the probability that any t_i will have a value between t and $t + dt$.

For the simple case in which the positions of the energy levels are not correlated the probability that any E_i will fall between E and $E + dE$ is independent of E and is simply $\rho\,dE$, where $\rho = D^{-1}$ is the average number of levels in a unit interval of energy. Let us determine the probability of a spacing S; that is, given a level at E, what is the probability of having no level in the interval $(E, E + S)$ and one level in dS at $E + S$. For this we divide the interval S into m equal parts. Because the levels are independent, the probability of having no level in $(E, E + S)$ is the product of the



probabilities of having no level in any of these m parts. If m is large, so that S/m is small, we can write this as

$$\left(1 - \rho \frac{S}{m}\right)^m \xrightarrow[m\to\infty]{} e^{-\rho S}$$

Moreover, the probability of having a level in dS at $E + S$ is $\rho\,dS$. Therefore, given a level at E, the probability that there is no level in $(E, E + S)$ and one level in dS at $E + S$ is

$$e^{-\rho S} \rho\, dS$$

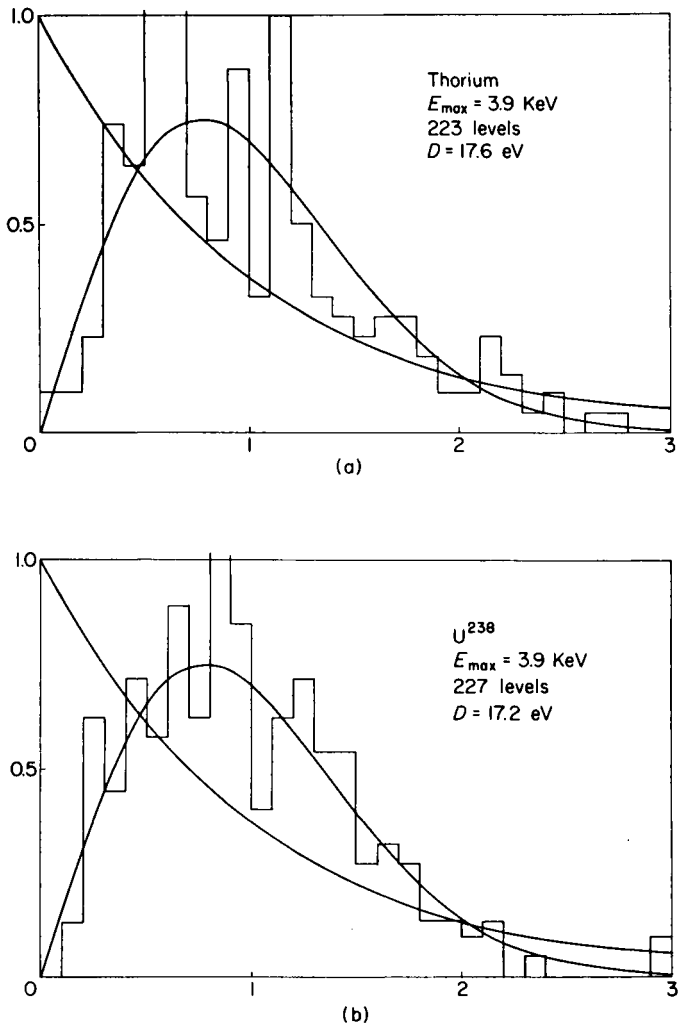


FIG. 1.1. Summary of the experimental data on nuclear level spacings for the elements Th and U^{238}. (*a*) Histogram of the observed density of level spacings as a function of $t = S/D$; the spacing is measured in units of the mean level spacing for thorium. (*b*) The same histogram for the nucleus U^{238}. The two solid curves correspond to the random and orthogonal cases. For details, see (1.1), (5.84), and (5.105). From Garg *et al.* [1].

or in terms of the variable $t = S/D = \rho S$

$$p(t)\,dt = e^{-t}\,dt. \tag{1.1}$$

This is known as the Poisson distribution or the spacing rule for random levels.

That (1.1) is not correct for nuclear levels of the same spin and parity or for atomic levels of the same parity and orbital and spin angular momenta is clearly seen by a comparison with the empirical evidence (Figures 1.1 and 1.2).

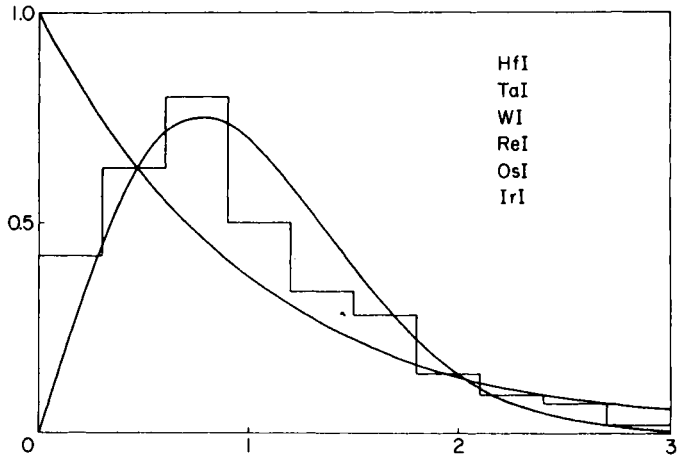


FIG. 1.2. Plot of the density of spacings between odd parity atomic levels of a group of elements in the region of osmium. The levels in each element were separated according to the angular momentum, and separate histograms were constructed for each level series, and then combined. The elements and the number of contributed spacings are HfI, 74; TaI, 180; WI, 262; ReI, 165; OsI, 145; IrI, 131, which lead to a total of 957 spacings. The solid curves correspond to the random and orthogonal cases; (1.1), (5.84), and (5.105). From Porter and Rosenzweig [1].

1.4. Wigner Surmise

When the experimental situation was not yet conclusive, Wigner [3] proposed the following rules for spacing distributions:

1. In the sequence of levels with the same spin and parity, called a simple sequence, the probability density function for a spacing is given by

$$p_W(t) = \frac{\pi}{2}\,t \exp\left(-\frac{\pi}{4}\,t^2\right), \qquad t = \frac{S}{D}. \tag{1.2}$$

2. The levels with different spin and parity are not correlated. The function $p(t)$ for a mixed sequence may be obtained by randomly superimposing the constituent simple sequences (cf. Appendix A.22).

Two simple arguments give rise to Rule 1. As pointed out by Wigner [3] and by Landau and Smorodinsky [1], it is reasonable to expect that, given a level at E, the probability that another level will lie around $E + S$ is proportional to S for very small S. Now, if we extrapolate this to all S's and, in addition, assume that the probabilities in various intervals of length S/m obtained by dividing S into m equal parts are mutually independent, we arrive at

$$p(t)\,dt = \lim_{m\to\infty} \prod_{r=0}^{m-1} \left(1 - \frac{tr}{m}\frac{1}{m}a\right) at\,dt$$

$$= at\,e^{-(1/2)at^2}\,dt. \tag{1.3}$$

The constant a can be determined by the condition that the average value of $t = S/D$ is unity:

$$\int_0^\infty t\,p(t)\,dt = 1. \tag{1.4}$$

Let us, at this point, define the n-point correlation function $R_n(E_1, \ldots, E_n)$ so that $R_n\,dE_1\,dE_2 \cdots dE_n$ is the probability of finding

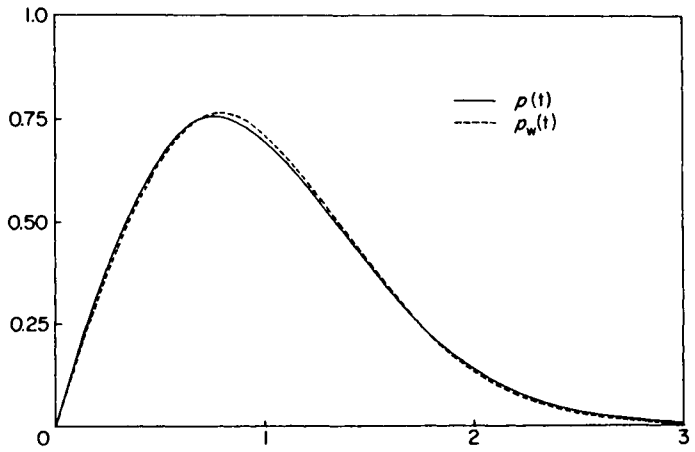


FIG. 1.3. The probability density functions $p(t)$ and $p_W(t)$; (1.2), (5.84), and (5.105)

a level in each of the intervals $(E_1, E_1 + dE_1),\ldots, (E_n, E_n + dE_n)$, all other levels being unobserved. The two simple arguments of Wigner given in the derivation of Rule 1 are equivalent to the following. The two-point correlation function $R_2(E_1, E_2)$ is linear in the variable $|E_1 - E_2|$, and three and higher order correlation functions are negligibly small.

We shall see in Chapter 5 that both arguments are inaccurate, whereas Rule 1 is very near the correct result (Figure 1.3). It is surprising that the two errors made so nearly compensate each other.

1.5. Electromagnetic Properties of Small Metallic Particles

Consider small metallic particles at low temperatures. The number of electrons in a volume V is $n \approx 4\pi p_0^3 V/3h^3$, where p_0 is the Fermi momentum and h is Planck's constant. The energy of an excitation near the Fermi surface is $E_0 \approx p_0^2/2m^*$, where m^* is the effective mass of the electron. The level density at zero excitation is therefore $\sigma = dn/dE_0 \approx 4\pi p_0 V m^*/h^3$, and the average level spacing is the inverse of this quantity $\Delta \approx \sigma^{-1}$. For a given temperature we can easily estimate the size of the metallic particles for which $\Delta \gg kT$, where k is Boltzmann's constant and T is the temperature in degrees Kelvin. For example, a metallic particle of size $10^{-6} - 10^{-7}$ cm contains $10^4 - 10^5$ electrons and, at $T \approx 10°$K, $\Delta \approx 1$ eV, whereas $kT \approx 10^{-3}$ eV. It is possible to produce particles of this size experimentally and then to sort them out according to their size (e.g., by centrifuging and sampling at a certain radial distance). Thus we have a large number of metallic particles, each of which has a different shape and therefore a different set of electronic energy levels but the same average level spacing, for the volumes are equal. It would be desirable if we could separate (e.g., by applying a nonuniform magnetic field) particles containing an odd number of conduction electrons from those containing an even number. The energy-level schemes for these two types of particle have very different properties (see Chapters 2 and 3).

Given the position of the electronic energies, we can calculate the partition function in the presence of a magnetic field and then use thermodynamic relations to derive various properties such as electronic specific heat and spin paramagnetism. Fröhlich [1] assumed that the energies were equally spaced and naturally obtained the result that all physical quantities decrease exponentially at low temperatures

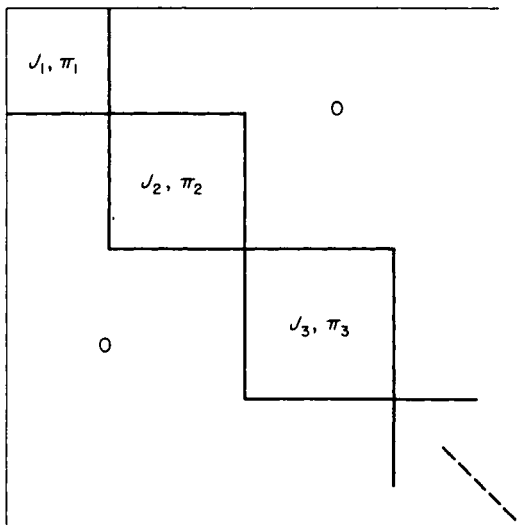
as $e^{-\Delta/kT}$ for $1 \ll \Delta/kT$. Kubo [1] repeated the calculation with the assumption that the energies were random without correlations and that their spacings therefore follow a Poisson law. He arrived at a linear law for the specific heat $\sim kT/\Delta$. The constants are different for particles containing an odd number of electrons from those containing an even number. For spin paramagnetism even the dependence on temperature is different for the two sets of particles. Instead of Fröhlich's equal spacing rule or Kubo's Poisson law, it would perhaps be better to adopt the point of view of Gorkov and Eliashberg [1], which may be justified as follows. The energies are the eigenvalues of a fixed Hamiltonian with random boundary conditions. We may incorporate these boundary conditions into the Hamiltonian by the use of fictitious potentials.

In contrast to nuclear spectra, we have the possibility of realizing in practice all three ensembles considered in various sections of this book. They apply in particular when (a) the number of electrons (in each of the metallic particles) is even and there is no external magnetic field, (b) the number of electrons (in each of the metallic particles) is odd and there is no external magnetic field, (c) there is an external magnetic field $H \gg \Delta/\mu$, where μ is the magnetic moment of the electron.

2 / Gaussian Ensembles. The Joint Probability Density Function for the Matrix Elements

2.1. Preliminaries

In the mathematical model our systems are characterized by their Hamiltonians, which in turn are represented by Hermitian matrices. Let us look into the structure of these matrices. The low-lying energy levels (eigenvalues) are far apart and each may be described by a different set of quantum numbers. As we go to higher excitations, the levels draw closer, and because of their mutual interference most of the approximate quantum numbers lose their usefulness, for they are no longer exact. At still higher excitations the interference is so great that some quantum numbers may become entirely meaningless. However, there may be certain exact integrals of motion, such as total spin or parity, and the quantum numbers corresponding to them



are conserved whatever the excitation may be. If the basic functions are chosen to be the eigenfunctions of these conserved quantities, all Hamiltonian matrices of the ensemble will reduce to the form of diagonal blocks. One block will correspond uniquely to each set of exact quantum numbers. The matrix elements lying outside these blocks will all be zero, and levels belonging to two different blocks will be statistically uncorrelated. As to the levels corresponding to the same block, the interactions are so complex that any regularity resulting from partial diagonalization will be washed out.

We shall assume that such a basis has already been chosen and restrict our attention to one of the diagonal blocks, an $(N \times N)$ Hermitian matrix in which N is a large but fixed positive integer. Because nuclear spectra contain at least hundreds of levels with the same spin and parity, we are interested in the limit of very large N.

With these preliminaries, the matrix elements may be supposed to be random variables and allowed the maximum statistical independence permitted under symmetry requirements. To specify precisely the correlations among various matrix elements we need a careful analysis of the consequences of time-reversal invariance.

2.2. Time-Reversal Invariance†

We begin by recapitulating the basic notions of time-reversal invariance. From physical considerations, the time-reversal operator is required to be antiunitary [Wigner, 1] and can be expressed, as any other antiunitary operator, in the form

$$T = KC, \tag{2.1}$$

where K is a fixed unitary operator and the operator C takes the complex conjugate of the expression following it. Thus a state under time reversal transforms to

$$\psi^R = T\psi = K\psi^*, \tag{2.2}$$

ψ^* being the complex conjugate of ψ. From the condition

$$(\phi, A\psi) = (\psi^R, A^R\phi^R)$$

for all pairs of states ψ, ϕ, and (2.2), we deduce that under time reversal an operator A transforms to

$$A^R = KA^TK^{-1}, \tag{2.3}$$

† Sections 2.2 to 2.5 are based largely on an article by Dyson [1].

where A^T is the transpose of A. A is said to be self-dual if $A^R = A$. A physical system is invariant under time reversal if its Hamiltonian is self-dual, that is, if

$$H^R = H. \tag{2.4}$$

When the representation of the states is transformed by a unitary transformation, $\psi \to U\psi$, T transforms according to

$$T \to UTU^{-1} = UTU^{+} \tag{2.5}$$

or K transforms according to

$$K \to UKU^T. \tag{2.6}$$

Because operating twice with T should leave the physical system unchanged, we have

$$T^2 = \alpha \cdot 1, \qquad |\alpha| = 1, \tag{2.7}$$

where 1 is the unit operator; or

$$T^2 = KCKC = KK^*CC = KK^* = \alpha \cdot 1, \tag{2.8}$$

But K is unitary:

$$K^*K^T = 1.$$

From these two equations we get

$$K = \alpha K^T = \alpha(\alpha K^T)^T = \alpha^2 K.$$

Therefore

$$\alpha^2 = 1 \quad \text{or} \quad \alpha = \pm 1, \tag{2.9}$$

so that the unitary matrix K is either symmetric or antisymmetric. In other words, either

$$KK^* = 1 \tag{2.10}$$

or

$$KK^* = -1. \tag{2.11}$$

These alternatives correspond, respectively, to an integral or a half-odd integral total angular momentum of the system measured

in units of $\hbar$ [Wigner, 1], for the total angular momentum operator $\mathbf{J} = (J_1, J_2, J_3)$ must transform as

$$J_l^R = -J_l, \qquad l = 1, 2, 3. \tag{2.12}$$

For brevity we call the two possibilities the even-spin and odd-spin case, respectively.

2.3. Gaussian Orthogonal Ensemble

Suppose now that the even-spin case holds and (2.10) is valid. Then a unitary operator U will exist such that (cf. Appendix A.23)

$$K = UU^T. \tag{2.13}$$

By (2.6) a transformation $\psi \to U^{-1}\psi$ performed on the states ψ brings K to unity. Thus in the even-spin case the representation of states can always be chosen so that

$$K = 1. \tag{2.14}$$

After one such representation is found, further transformations $\psi \to R\psi$ are allowed only with R a real orthogonal matrix so that (2.14) remains valid. The consequence of (2.14) is that self-dual matrices are symmetric. In the even spin case every system invariant under time reversal will be associated with a real symmetric matrix H if the representation of states is suitably chosen. For even-spin systems with time-reversal invariance the Gaussian orthogonal ensemble E_{1G}, defined below, is therefore appropriate.

Definition 2.1: The Gaussian orthogonal ensemble E_{1G} is defined in the space T_{1G} of real symmetric matrices by two requirements:

1. The ensemble is invariant under every transformation

$$H \to W^T H W \tag{2.15}$$

of T_{1G} into itself, where W is any real orthogonal matrix.

2. The various elements H_{kj}, $k \leqslant j$, are statistically independent.

These requirements, expressed in the form of equations, read as follows:

1. The probability $P(H)\,dH$ that a system of E_{1G} will belong to the volume element $dH = \prod_{k\leqslant j} dH_{kj}$ is invariant under real orthogonal transformations:

$$P(H')\,dH' = P(H)\,dH, \tag{2.16}$$

where

$$H' = W^T H W \tag{2.17}$$

and

$$W^T W = W W^T = 1. \tag{2.18}$$

2. This probability density function $P(H)$ is a product of functions, each of which depends on at most a single variable:

$$\prod_{k\leqslant j} \tag{2.19}$$

Suppose, next, that we are dealing with a system invariant under space rotations. The spin may now be even or odd. The Hamiltonian matrix H which represents the system commutes with every component of $\mathbf{J}$. If we use the standard representation of the J matrices with J_1 and J_3 real and J_2 pure imaginary, (2.12) may be satisfied by the usual choice [Wigner, 1]

$$K = e^{i\pi J_2} \tag{2.20}$$

for K. With this choice of K, H and K commute and H^R reduces to H^T. Thus a rotation-invariant system is represented by a real symmetric matrix H, and once again the ensemble E_{1G} is appropriate.

2.4. Gaussian Symplectic Ensemble†

In this section we discuss a system to which E_{1G} does not apply, a system with odd-spin, invariant under time reversal, but having no rotational symmetry. In this case (2.11) holds, K cannot be diagonalized by any transformation of the form (2.6), and there is no integral of the motion by which the double-valuedness of the time-reversal operation can be trivially eliminated.

† Dyson [1].

Every antisymmetric unitary operator can be reduced by a transformation (2.6) to the standard canonical form (cf. Appendix A.23)

$$Z = \begin{bmatrix} 0 & -1 & 0 & 0 & \cdots \\ 1 & 0 & 0 & 0 & \cdots \\ 0 & 0 & 0 & -1 & \cdots \\ 0 & 0 & 1 & 0 & \cdots \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \end{bmatrix} \equiv \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} + \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} + \cdots \tag{2.21}$$

which consists of (2×2) blocks

$$\begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$$

along the leading diagonal; all other elements of Z are zero. We assume that the representation of states is chosen so that K is reduced to this form. The number of rows and columns of all matrices must now be even, for otherwise K would be singular in contradiction to (2.11). It is convenient to denote the order of the matrices by $2N$ instead of N. After one such representation is chosen, for which $K = Z$, further transformations $\psi \to B\psi$ are allowed, only with B a unitary $(2N \times 2N)$ matrix for which

$$Z = BZB^T. \tag{2.22}$$

Such matrices B form precisely the N-dimensional symplectic group [Weyl, 1], usually denoted by $Sp(N)$.

It is well known [Chevalley, 1; Dieudonné, 1] that the algebra of the symplectic group can be expressed most conveniently in terms of quaternions. We therefore introduce the standard quaternion notation for (2×2) matrices,

$$e_1 = \begin{bmatrix} i & 0 \\ 0 & -i \end{bmatrix}, \qquad e_2 = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}, \qquad e_3 = \begin{bmatrix} 0 & -i \\ -i & 0 \end{bmatrix}, \tag{2.23}$$

with the usual multiplication table

$$e_1^2 = e_2^2 = e_3^2 = -1, \tag{2.24}$$

$$e_1e_2 = -e_2e_1 = e_3\,, \qquad e_2e_3 = -e_3e_2 = e_1\,, \qquad e_3e_1 = -e_1e_3 = e_2\,. \tag{2.25}$$

Note that in (2.23), as well as throughout the rest of this book, i is the ordinary imaginary unit and not a quaternion unit. The matrices e_1, e_2, and e_3, together with the (2×2) unit matrix

$$1 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix},$$

form a complete set, and any (2×2) matrix with complex elements can be expressed linearly in terms of them with

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} = \frac{1}{2}(a+d)\,1 - \frac{i}{2}(a-d)\,e_1 - \frac{1}{2}(b-c)\,e_2 + \frac{i}{2}(b+c)\,e_3\,. \tag{2.26}$$

All the $(2N \times 2N)$ matrices will be considered as cut into N^2 blocks of (2×2) and each (2×2) block expressed in terms of quaternions. In general, a $(2N \times 2N)$ matrix with complex elements thus becomes an $(N \times N)$ matrix with complex quaternion elements. In particular, the matrix Z is now

$$Z = e_2 I, \tag{2.27}$$

where I is the $(N \times N)$ unit matrix. It is easy to verify that the rules of matrix multiplication are not changed by this partitioning.

Let us add some definitions. We call a quaternion "real" if it is of the form

$$q = q^{(0)} + \mathbf{q} \cdot \mathbf{e}, \tag{2.28}$$

with $q^{(0)}$, $q^{(1)}$, $q^{(2)}$, and $q^{(3)}$. Thus a real quaternion does not correspond to a (2×2) matrix with real elements. Any complex quaternion has a

$$\bar{q} = q^{(0)} - \mathbf{q} \cdot \mathbf{e}, \tag{2.29}$$

which is distinct from its

$$q^* = q^{(0)*} + \mathbf{q}^* \cdot \mathbf{e}. \tag{2.30}$$

A quaternion with $q^* = q$ is real; one with $q^* = -q$ is pure imaginary; and one with $\bar{q} = q$ is a scalar. By applying both types of conjugation together, we obtain the

$$q^+ = \bar{q}^* = q^{(0)*} - \mathbf{q}^* \cdot \mathbf{e}. \tag{2.31}$$

A quaternion with $q^+ = q$ is Hermitian and corresponds to the ordinary notion of a (2×2) Hermitian matrix; one with $q^+ = -q$

is anti-Hermitian. The conjugate (Hermitian conjugate) of a product of quaternions is the product of their conjugates (Hermitian conjugates) taken in the reverse order:

$$\overline{(q_1 q_2 \cdots q_n)} = \bar{q}_n \cdots \bar{q}_2 \bar{q}_1 , \tag{2.32}$$

$$(q_1 q_2 \cdots q_n)^+ = q_n{}^+ \cdots q_2{}^+ q_1{}^+ . \tag{2.33}$$

Now consider a general $(2N \times 2N)$ matrix A which is to be written as an $(N \times N)$ matrix Q with quaternion elements q_{kj}; $k, j = 1, 2, \ldots, N$. The standard matrix operations on A are then reflected in Q in the following way:

Transposition,

$$(Q^T)_{kj} = -e_2 \bar{q}_{jk} e_2 . \tag{2.34}$$

Hermitian conjugation,

$$(Q^+)_{kj} = q_{jk}^+ . \tag{2.35}$$

Time reversal,

$$(Q^R)_{kj} = e_2 (Q^T)_{kj} e_2^{-1} = \bar{q}_{jk} . \tag{2.36}$$

The matrix Q^R is called the "dual" of Q. A "self-dual" matrix is one with $Q^R = Q$.

The usefulness of quaternion algebra is a consequence of the simplicity of (2.35) and (2.36). In particular, it is noteworthy that the time-reversal operator K does not appear explicitly in (2.36) as it did in (2.3). By (2.35) and (2.36) the condition

$$Q^R = Q^+ \tag{2.37}$$

is necessary and sufficient for the elements of Q to be real quaternions. When (2.37) holds, we call Q "quaternion real."

A unitary matrix B that satisfies (2.22) is automatically quaternion real. In fact, it satisfies the conditions

$$B^R = B^+ = B^{-1}, \tag{2.38}$$

which define the symplectic group. The matrices H which represent the energy operators of physical systems are Hermitian as well as self-dual:

$$H^R = H, \qquad H^+ = H, \tag{2.39}$$

hence are also quaternion real.

$$q_{jk}^{+} = \bar{q}_{jk} = q_{kj} \tag{2.40}$$

or $q_{jk}^{(0)}$ Thus the number of real independent parameters that define a $(2N \times 2N)$ self-dual Hermitian matrix is

$$\tfrac{1}{2}N(N+1) + \tfrac{1}{2}N(N-1)\cdot 3 = N(2N-1).$$

From this notational excursion, let us come back to the point. Systems having odd-spin, invariance under time-reversal, but no rotational symmetry, must be represented by self-dual, Hermitian Hamiltonians. Therefore the Gaussian symplectic ensemble, as defined below, should be appropriate for their description.

Definition 2.2: The Gaussian symplectic ensemble E_{4G} is defined in the space T_{4G} of self-dual Hermitian matrices by the following properties:

1. The ensemble is invariant under every orthomorphism

$$H \to W^{R}HW \tag{2.41}$$

of T_{4G} into itself, where W is any symplectic matrix.

2. Various linearly independent components of H are also statistically independent.

These requirements put in the form of equations read as follows:

1. The probability $P(H)\,dH$ that a system E_{4G} will belong to the volume element

$$dH \equiv \prod_{k\leqslant j} dH_{kj}^{(0)} \prod_{\lambda=1}^{3} \prod_{k<j} dH_{kj}^{(\lambda)} \tag{2.42}$$

is invariant under symplectic transformations; that is,

$$P(H')\,dH' = P(H)\,dH \tag{2.43}$$

if

$$H' = W^{R}HW, \tag{2.44}$$

where

$$WZW^T = Z \tag{2.45}$$

2. The probability density function $P(H)$ is a product of functions each of which depends on a single variable:

$$P(H) = \prod_{k \leqslant j} f_{kj}^{(0)}(H_{kj}^{(0)}) \prod_{\lambda=1}^{3} \prod_{k<j} f_{kj}^{(\lambda)}(H_{kj}^{(\lambda)}). \tag{2.46}$$

2.5. Gaussian Unitary Ensemble

For completeness we discuss briefly a much simpler ensemble, the Gaussian unitary ensemble E_{2G} which applies to systems without invariance under time reversal. Such systems are easily created in principle by putting an ordinary atom or nucleus, for example, into an externally generated magnetic field. The external field is not affected by the time-reversal operation. However, for the unitary ensemble to be applicable, the splitting of levels by the magnetic field must be at least as large as the average level spacing in the absence of the magnetic field. The magnetic field must, in fact, be so strong that it will completely "mix up" the level structure that would exist in zero field; for otherwise our random hypothesis cannot be justified. This state of affairs could never occur in nuclear physics. In atomic or molecular physics a practical application of the unitary ensemble may perhaps be possible.

A system without time-reversal invariance has a Hamiltonian that may be an arbitrary Hermitian matrix not restricted to be real or self-dual. Thus we are led to the following definition.

Definition 2.3: The Gaussian unitary ensemble E_{2G} is defined in the space of Hermitian matrices by the following properties:

1. The probability $P(H)\,dH$ that a system of E_{2G} will belong to the volume element

$$dH = \prod_{k \leqslant j} dH_{kj}^{(0)} \prod_{k<j} dH_{kj}^{(1)}, \tag{2.47}$$

where $H_{kj}^{(0)}$ and $H_{kj}^{(1)}$ are real and imaginary parts of H_{kj}, is invariant under every automorphism

$$H \to U^{-1}HU \tag{2.48}$$

of T_{2G} into itself, where U is any unitary matrix.

2. Various linearly independent components of H are also statistically independent.

In mathematical language these requirements are

1. $$P(H')\,dH' = P(H)\,dH, \tag{2.49}$$

$$H' = U^{-1}HU, \tag{2.50}$$

where U is any unitary matrix.

2. $P(H)$ is a product of functions, each of which depends on a single variable:

$$P(H) = \prod_{k\leqslant j} f_{kj}^{(0)}(H_{kj}^{(0)}) \prod_{k<j} f_{kj}^{(1)}(H_{kj}^{(1)}). \tag{2.51}$$

2.6. Joint Probability Density Function for Matrix Elements

We now come to the question of the extent to which we are still free to specify the joint probability density function $P(H)$. It will be seen that the two postulates of invariance and statistical independence elaborated above fix uniquely the functional form of $P(H)$.

The postulate of invariance restricts $P(H)$ to depend only on a finite number of traces of the powers of H. We state this fact as a lemma [Weyl, 1].

Lemma 2.1. *All the invariants of an $(N \times N)$ matrix H under nonsingular similarity transformations A,*

$$H \to H' = AHA^{-1},$$

can be expressed in terms of the traces of the first N powers of H.

Proof: Because all invariants are symmetric functions of the eigenvalues λ_k, $k = 1, 2, \ldots, N$, of H, and

$$\operatorname{tr} H^j = \sum_{k=1}^{N} \lambda_k{}^j \equiv t_j, \quad \text{say},$$

we need to show that any symmetric function of λ_k can be expressed in terms of the first N of the t_j. Let the secular equation which determines the λ_k be

$$\det[H - \lambda I] \equiv (-\lambda)^N + \sigma_1(-\lambda)^{N-1} + \cdots + \sigma_N = 0,$$

so that, given the coefficients

$$\sigma_j = \sum_{1 \leqslant k_1 < \cdots < k_j \leqslant N} \lambda_{k_1} \lambda_{k_2} \cdots \lambda_{k_j}, \qquad 1 \leqslant j \leqslant N,$$

all the eigenvalues λ_k are uniquely determined except for their order. Thus any symmetric function of the λ_k can be expressed in terms of the basic functions $\sigma_1, \sigma_2, \ldots, \sigma_N$. To show that the t_j form another such basis it will then be sufficient to express σ_j in terms of t_j. This is achieved by the equation (Appendix A.1)

$$\sigma_r = (r!)^{-1} \det[a_{kj}]_{k,j=1,2,\ldots,r}; \qquad 1 \leqslant r \leqslant N, \tag{2.52}$$

where

$$a_{kj} = \begin{cases} t_{j-k+1}, & \text{if } k \leqslant j, \\ j, & \text{if } k = j+1, \\ 0, & \text{if } k > j+1. \end{cases} \tag{2.53}$$

Incidently, we note that

$$\det[a_{kj}]_{k,j=1,2,\ldots,r} \equiv 0, \quad \text{if} \quad r > N, \tag{2.54}$$

which expresses the traces of all the powers of H in terms of those of the first N powers.

The postulate of statistical independence excludes everything except the traces of the first two powers, and these, too, may occur only in an exponential. To see this we will need the following lemma.

Lemma 2.2. *If three continuous and differentiable functions $f_k(x)$; $k = 1, 2, 3$, satisfy the equation*

$$f_1(xy) = f_2(x) + f_3(y), \tag{2.55}$$

they are necessarily of the form $a \ln x + b_k$ with $b_1 = b_2 + b_3$.

Proof: Differentiating (2.55) with respect to x, we have

$$f_1'(xy) = \frac{1}{y} f_2'(x),$$

which, on integration with respect to y, gives

$$\frac{1}{x} f_1(xy) = f_2'(x) \ln y + \frac{1}{x} g(x), \tag{2.56}$$

where $g(x)$ is still arbitrary. Substituting $f_1(xy)$ from (2.56) into (2.55), we get

$$x f_2'(x) \ln y + g(x) - f_2(x) = f_3(y). \tag{2.57}$$

Therefore the left-hand side of (2.57) must be independent of x; this is possible only if

$$x f_2'(x) = a \qquad \text{and} \qquad g(x) - f_2(x) = b_3,$$

that is, only if

$$f_2(x) = a \ln x + b_2 = g(x) - b_3,$$

where a, b_2, and b_3 are arbitrary constants.

Now (2.57) gives

$$f_3(y) = a \ln y + b_3$$

and finally (2.55) gives

$$f_1(xy) = a \ln(xy) + (b_2 + b_3). \qquad \text{Q.E.D.}$$

Let us now examine the consequences of the statistical independence of the various components of H. Consider the particular transformation

$$H = \grave{U}^{-1} H' U, \tag{2.58}$$

where

$$U = \begin{bmatrix} \cos\theta & \sin\theta & 0 \cdots 0 \\ -\sin\theta & \cos\theta & 0 \cdots 0 \\ 0 & 0 & 1 \cdots 0 \\ \cdots & \cdots & \cdots \\ 0 & 0 & 0 \cdots 1 \end{bmatrix} \tag{2.59}$$

or, in quaternion notation (provided N is even),

$$U = \begin{bmatrix} \cos\theta - e_2 \sin\theta & 0 \cdots 0 \\ 0 & 1 \cdots 0 \\ \cdots & \cdots \\ 0 & 0 \cdots 1 \end{bmatrix}. \tag{2.60}$$

This U is, at the same time, orthogonal, symplectic, and unitary.

Differentiation of (2.58) with respect to θ gives

$$\begin{aligned} \frac{\partial H}{\partial \theta} &= \frac{\partial U^T}{\partial \theta} H' U + U^T H' \frac{\partial U}{\partial \theta} \\ &= \frac{\partial U^T}{\partial \theta} U H + H U^T \frac{\partial U}{\partial \theta}, \end{aligned} \tag{2.61}$$

and by substituting for U, U^T, $\partial U/\partial\theta$, and $\partial U^T/\partial\theta$ from (2.59) or (2.60) we get

$$\frac{\partial H}{\partial\theta} = AH + HA^T, \tag{2.62}$$

where

$$A = \frac{\partial U^T}{\partial\theta}\,U = \begin{bmatrix} 0 & -1 & 0\cdots 0 \\ 1 & 0 & 0\cdots 0 \\ 0 & 0 & 0\cdots 0 \\ \cdot & \cdot & \cdot\ \cdot\ \cdot \\ 0 & 0 & 0\cdots 0 \end{bmatrix} \tag{2.63}$$

or, in quaternion notation, A is diagonal.

$$A = \begin{bmatrix} e_2 & 0\cdots 0 \\ 0 & 0\cdots 0 \\ \cdot & \cdot\ \cdot\ \cdot \\ 0 & 0\cdots 0 \end{bmatrix}. \tag{2.64}$$

If the probability density function

$$P(H) = \prod f_{kj}^{(\alpha)}(H_{kj}^{(\alpha)}) \tag{2.65}$$

is invariant under the transformation U, its derivative with respect to θ must vanish; that is

$$\sum \frac{1}{f_{kj}^{(\alpha)}}\,\frac{\partial f_{kj}^{(\alpha)}}{\partial H_{kj}^{(\alpha)}}\,\frac{\partial H_{kj}^{(\alpha)}}{\partial\theta} = 0. \tag{2.66}$$

Let us write this equation explicitly, say, for the unitary case. Equations (2.62) and (2.66) give

$$\begin{aligned} &\left\{\left[-\frac{1}{f_{11}^{(0)}}\frac{\partial f_{11}^{(0)}}{\partial H_{11}^{(0)}} + \frac{1}{f_{22}^{(0)}}\frac{\partial f_{22}^{(0)}}{\partial H_{22}^{(0)}}\right][2H_{12}^{(0)}] + \frac{1}{f_{12}^{(0)}}\frac{\partial f_{12}^{(0)}}{\partial H_{12}^{(0)}}[H_{11}^{(0)} - H_{22}^{(0)}]\right\} \\ &\quad + \sum_{i=3}^{N}\left[-\frac{1}{f_{1k}^{(0)}}\frac{\partial f_{1k}^{(0)}}{\partial H_{1k}^{(0)}}H_2^{(0)} + \frac{1}{f_{2k}^{(0)}}\frac{\partial f_{2k}^{(0)}}{\partial H_{2k}^{(0)}}H_{1k}^{(0)}\right] \\ &\quad + \sum_{i=3}^{N}\left[-\frac{1}{f_{1k}^{(1)}}\frac{\partial f_{1k}^{(1)}}{\partial H_{1k}^{(1)}}H_{2k}^{(1)} + \frac{1}{f_{2k}^{(1)}}\frac{\partial f_{2k}^{(1)}}{\partial H_{2k}^{(1)}}H_{1k}^{(1)}\right] = 0. \end{aligned} \tag{2.67}$$

The braces at the left-hand side of this equation depend on mutually exclusive sets of variables and their sum is zero. Therefore each must be a constant; for example,

$$-\frac{H_{2k}^{(0)}}{f_{1k}^{(0)}}\frac{\partial f_{1k}^{(0)}}{\partial H_{1k}^{(0)}}+\frac{H_{1k}^{(0)}}{f_{2k}^{(0)}}\frac{\partial f_{2k}^{(0)}}{\partial H_{2k}^{(0)}}=C_k^{(0)}. \tag{2.68}$$

On dividing both sides of (2.68) by $H_{1k}^{(0)}H_{2k}^{(0)}$ and applying the Lemma 2.2, we conclude that the constant $C_k^{(0)}$ must be zero, that is,

$$\frac{1}{H_{1k}^{(0)}}\frac{1}{f_{1k}^{(0)}}\frac{\partial f_{1k}^{(0)}}{\partial H_{1k}^{(0)}}=\frac{1}{H_{2k}^{(0)}}\frac{1}{f_{2k}^{(0)}}\frac{\partial f_{2k}^{(0)}}{\partial H_{2k}^{(0)}}=\text{constant}$$

$$=-2a, \text{ say}, \tag{2.69}$$

which on integration gives

$$f_{1k}^{(0)}[H_{1k}^{(0)}]=\exp\{-a[H_{1k}^{(0)}]^2\}. \tag{2.70}$$

In the other two cases we also derive a similar equation. Now because the off-diagonal elements come only as squares in the exponential and all invariants are expressible in terms of the traces of powers of H, the function $P(H)$ is an exponential that contains traces of at most the second power of H.

Because $P(H)$ is required to be invariant under more general transformations than we have here considered, one might think that the form of $P(H)$ is further restricted. This, however, is not so, for

$$P(H)=\exp(-a\,\mathrm{tr}\,H^2+b\,\mathrm{tr}\,H+C)$$

$$=e^C\prod_{k\leqslant j}\exp\{-a[H_{kj}^{(0)}]^2+bH_{jj}^{(0)}\}\prod_{k<j,\lambda}\exp\{-a[H_{kj}^{(\lambda)}]^2\} \tag{2.71}$$

is already a product of functions, each of which depends on a separate variable. Moreover, and real, a must be real and positive and b and c must be real.

Therefore we have proved the following theorem [Porter and Rosenzweig, 1; Wishart, 1].

Theorem 2.1. *In all the above three cases the form of $P(H)$ is automatically restricted to*

$$P(H)=\exp(-a\,\mathrm{tr}\,H^2+b\,\mathrm{tr}\,H+c), \tag{2.72}$$

where a is real and positive and b and c are real.

In the foregoing discussion we have emphasized the postulate of statistical independence of various components of H even at the risk of frequent repetitions. This statistical independence is important in restricting $P(H)$ to the simple form (2.72), and hence makes the subsequent analytical work tractable. However, it lacks a clear physical motivation and therefore looks somewhat artificial.

The main objection to the assumption of statistical independence, leading to (2.72), is that all values of $H_{kj}^{(\lambda)}$ are not equally weighted and therefore do not correspond to all "interactions" being "equally probable." By a formal change Dyson [1–6] has defined his "circular ensembles," which are esthetically more satisfactory to some people and equally easy to work with. We shall come to them in Chapters 8 to 11. On the other hand, Rosenzweig [1] has emphasized the "fixed strength" ensemble. Others [Leff, 1; Fox and Kahn, 1] have arbitrarily tried the so-called "generalized" ensembles. A brief review of these topics is given in Chapter 17.

3 / Gaussian Ensembles. The Joint Probability Density Function for the Eigenvalues†

3.1. Orthogonal Ensemble

The joint probability density function (abbreviated j.p.d.f. later in the chapter) for the eigenvalues $\theta_1, \theta_2, \ldots, \theta_N$ can be obtained from (2.72) by expressing the various components of H in terms of the N eigenvalues θ_j and other mutually independent variables, p_μ, say, which together with the θ_j form a complete set. In an $(N \times N)$ real symmetric matrix the number of independent real parameters which determine all H_{kj} is $\frac{1}{2}N(N+1)$. We may take these as H_{kj} with $k \leqslant j$. The number of extra parameters p_μ needed is therefore

$$l = \tfrac{1}{2}N(N+1) - N = \tfrac{1}{2}N(N-1). \tag{3.1}$$

Because

$$\operatorname{tr} H^2 = \sum_1^N \theta_j^2, \qquad \operatorname{tr} H = \sum_1^N \theta_j, \tag{3.2}$$

the probability that the N roots and the $\frac{1}{2}N(N-1)$ parameters will occur in unit intervals around $\theta_1, \ldots, \theta_N$ and $p_1, p_2, \ldots, p_l$ is, according to (2.72),

$$\Delta(\theta_1, \ldots, \theta_N ; p_1, \ldots, p_l) = \exp\left(-a\sum_1^N \theta_j^2 + b\sum_1^N \theta_j + c\right) J(\theta, p), \tag{3.3}$$

where J is the Jacobian

$$J(\theta, p) = \left| \frac{\partial(H_{11}, H_{12}, \ldots, H_{NN})}{\partial(\theta_1, \ldots, \theta_N, p_1, \ldots, p_l)} \right|. \tag{3.4}$$

Hence the j.p.d.f. of the eigenvalues θ_j can be obtained by integrating (3.3) over the parameters $p_1, \ldots, p_l$. It is usually possible to choose

† This chapter is based largely on Wigner's article [6].

these parameters so that the Jacobian (3.4) becomes a product of a function of θ_j and a function of p_μ. If this is the case, the integration provides the required j.p.d.f. as a product of the exponential in (3.3), the aforementioned function of the θ_j and a constant. The constant can then be absorbed in c in the exponential.

To define the parameters p_μ [Wigner, 6] we recollect that any real symmetric matrix H can be diagonalized by a real orthogonal matrix [Wigner, 2]:

$$H = U\Theta U^{-1} \tag{3.5}$$

$$= U\Theta U^T, \tag{3.5'}$$

where Θ is the diagonal matrix with diagonal elements $\theta_1, \theta_2, \ldots, \theta_N$ arranged in some order, say, $\theta_1 \leqslant \theta_2 \leqslant \cdots \leqslant \theta_N$, and U is a real orthogonal matrix

$$UU^T = U^TU = 1, \tag{3.6}$$

whose columns are the normalized eigenvectors of H. They are, or may be chosen to be, mutually orthogonal. To define U completely we must in some way fix the phases of the eigenvectors—for instance by requiring that the first nonvanishing component be positive. Thus U depends on $\frac{1}{2}N(N-1)$ real parameters and may be chosen to be the U_{kj}, $k > j$. If H has multiple eigenvalues, further conditions are needed to fix U completely. It is not necessary to specify them, for they apply only in regions of lower dimensionality which are irrelevant to the probability density function. At any rate, the $\frac{1}{2}N(N-1)$ parameters p_μ are supposed to characterize the U which is subject to the preceding conditions. Once this is done, the matrix H, which completely determines the Θ and the U subject to the preceding conditions, also determines the θ_j and the p_μ uniquely. Conversely, the θ_j and p_μ completely determine the U and Θ, and hence by (3.5) all the matrix elements of H.

Differentiating (3.6), we get

$$\frac{\partial U^T}{\partial p_\mu} U + U^T \frac{\partial U}{\partial p_\mu} = 0, \tag{3.7}$$

and because the two terms in (3.7) are the Hermitian conjugates of each other,

$$S_\mu = U^+ \frac{\partial U}{\partial p} = -\frac{\partial U^+}{\partial p_\mu} U \tag{3.8}$$

is an antisymmetric matrix.

Also from (3.5) we have

$$\frac{\partial H}{\partial p_\mu} = \frac{\partial U}{\partial p_\mu} \Theta U^T + U\Theta \frac{\partial U^T}{\partial p_\mu}. \tag{3.9}$$

On multiplying (3.9) by U^T on the left and by U on the right, we get

$$U^T \frac{\partial H}{\partial p} U = S^{(\mu)}\Theta - \Theta S^{(\mu)}. \tag{3.10}$$

In terms of its components, (3.10) reads

$$\sum_{j,k} \frac{\partial H_{jk}}{\partial p_\mu} U_{j\alpha} U_{k\beta} = S^{(\mu)}_{\alpha\beta}(\theta_\beta - \theta_\alpha). \tag{3.11}$$

In a similar way, by differentiating (3.5) with respect to θ_γ,

$$\sum_{j,k} \frac{\partial H_{jk}}{\partial \theta_\gamma} U_{j\alpha} U_{k\beta} = \frac{\partial \Theta_{\alpha\beta}}{\partial \theta_\gamma} = \delta_{\alpha\beta}\,\delta_{\alpha\gamma}. \tag{3.12}$$

The matrix of the Jacobian in (3.4) can be written in the partitioned form as

$$[J(\theta, p)] = \begin{bmatrix} \dfrac{\partial H_{jj}}{\partial \theta_\gamma} & \dfrac{\partial H_{jk}}{\partial \theta_\gamma} \\ \dfrac{\partial H_{jj}}{\partial p_\mu} & \dfrac{\partial H_{jk}}{\partial p_\mu} \end{bmatrix}. \tag{3.13}$$

The two columns in (3.13) correspond to N and $\frac{1}{2}N(N-1)$ actual columns: $1 \leqslant j < k \leqslant N$. The two rows in (3.13) correspond again to N and $\frac{1}{2}N(N-1)$ actual rows: $\gamma = 1, 2, \ldots, N$; $\mu = 1, 2, \cdots, \frac{1}{2}N(N-1)$. If we multiply the $[J]$ in (3.13) on the right by the

$$\tfrac{1}{2}N(N+1) \times \tfrac{1}{2}N(N+1)$$

matrix written in the partitioned form as

$$[V] = \begin{bmatrix} (U_{j\alpha}U_{j\beta}) \\ (U_{j\alpha}U_{k\beta}) \end{bmatrix}, \tag{3.14}$$

in which the two rows correspond to N and $\frac{1}{2}N(N-1)$ actual rows, $1 \leqslant j < k \leqslant N$, and the column corresponds to $\frac{1}{2}N(N+1)$ actual columns, $1 \leqslant \alpha \leqslant \beta \leqslant N$, we get by using (3.11) and (3.12)

$$[J][V] = \begin{bmatrix} \delta_{\alpha\beta}\,\delta_{\alpha\gamma} \\ S^{(\mu)}_{\alpha\beta}(\theta_\beta - \theta_\alpha) \end{bmatrix}. \tag{3.15}$$

The two rows on the right-hand side correspond to N and $\frac{1}{2}N(N-1)$ actual rows and the column corresponds to $\frac{1}{2}N(N+1)$ actual columns. Taking the determinant on both sides of (3.15), we have

$$J(\theta, p) \det V = \prod_{\alpha<\beta} (\theta_\beta - \theta_\alpha) \det \begin{bmatrix} \delta_{\alpha\beta}\, \delta_{\alpha\nu} \\ S^{\mu}_{\alpha\beta} \end{bmatrix}$$

or

$$J(\theta, p) = \prod_{\alpha<\beta} | \theta_\beta - \theta_\alpha | f(p), \tag{3.16}$$

where $f(p)$ is independent of the θ_j and depends only on the parameters p_μ.

By inserting this result in (3.3) and integrating over the variables p_μ we get the j.p.d.f. for the eigenvalues of the matrices of an orthogonal ensemble

$$P(\theta_1, ..., \theta_N) = \exp\Big[-\sum_1^N (a\theta_j^2 - b\theta_j - c)\Big] \prod_{j<k} | \theta_k - \theta_j |, \tag{3.17}$$

where c is some new constant. Moreover, if we shift the origin of the θ to $b/2a$ and change the energy scale everywhere by a constant factor $\sqrt{2a}$, we may replace θ_j with $(1/\sqrt{2a})\, x_j + b/2a$. By this formal change (3.17) takes the simpler form

$$P_{N1}(x_1, ..., x_N) = C_{N1} \exp\Big(-\frac{1}{2}\sum_1^N x_j^2\Big) \prod_{j<k} | x_j - x_k |, \tag{3.18}$$

where C_{N1} is a constant.

3.2. Symplectic Ensemble

As the analysis is almost identical in all three cases, we have presented the details for one particular ensemble—the orthogonal one. Here and in the following discussion we indicate briefly the modifications necessary to arrive at the required j.p.d.f. in the other two cases.

Corresponding to the result that a real symmetric matrix can be diagonalized by a real orthogonal matrix, we have the following:

Theorem 3.1. *Given a quaternion-real, self-dual matrix H, there exists a symplectic matrix U such that*

$$H = U\Theta U^{-1} = U\Theta U^{R}, \tag{3.19}$$

where Θ is diagonal, real, and scalar (cf. Appendix A.23).

$$\begin{bmatrix} \theta_j & 0 \\ 0 & \theta_j \end{bmatrix} \tag{3.20}$$

along the main diagonal. Thus the eigenvalues of H consist of The Hamiltonian of any system which is invariant under time reversal, which has odd spin, and no rotational symmetry satisfies the conditions of Theorem 3.1. All energy levels of such a system will be doubly degenerate. This is the Kramer's degeneracy [Kramer, 1], and Theorem 3.1 shows how it appears naturally in the quaternion language.

Apart from the N eigenvalues θ_j, the number of real independent parameters p_μ needed to characterize an $N \times N$ quaternion-real, self-dual matrix H is

$$l = 4 \cdot \tfrac{1}{2}N(N-1) = 2N(N-1). \tag{3.21}$$

Equations 3.2 and 3.3 are replaced, respectively, by

$$\operatorname{tr} H^2 = 2\sum_1^N \theta_j^2, \qquad \operatorname{tr} H = 2\sum_1^N \theta_j \tag{3.22}$$

and

$$\Delta(\theta_1, \ldots, \theta_N ; p_1, \ldots, p_l) = \exp\left[-\sum_1^N (2a\theta_j^2 - 2b\theta_j - c)\right] J(\theta, p), \tag{3.23}$$

where $J(\theta, p)$ is now given by

$$J(\theta, p) = \frac{\partial(H_{11}^{(0)}, \ldots, H_{NN}^{(0)}, H_{12}^{(0)}, \ldots, H_{12}^{(3)}, \ldots, H_{N-1,N}^{(0)}, \ldots, H_{N-1,N}^{(3)})}{\partial(\theta_1, \ldots, \theta_N, p_1, \ldots, p_{2N(N-1)})}. \tag{3.24}$$

Equation 3.5 is replaced by (3.19); (3.6), (3.7), (3.8), (3.9), and (3.10) are valid if Note that these equations are now in the quaternion language, and we need to separate the four quaternion parts of modified (3.19). For this we let

$$H_{jk} = H_{jk}^{(0)} + H_{jk}^{(1)} e_1 + H_{jk}^{(2)} e_2 + H_{jk}^{(3)} e_3, \tag{3.25}$$

$$S_{\alpha\beta}^{(\mu)} = S_{\alpha\beta}^{(0\mu)} + S_{\alpha\beta}^{(1\mu)} e_1 + S_{\alpha\beta}^{(2\mu)} e_2 + S_{\alpha\beta}^{(3\mu)} e_3, \tag{3.26}$$

and write (3.10) and the one corresponding to (3.12) in the form of partitioned matrices:

$$\begin{bmatrix} \frac{\partial H_{jj}^{(0)}}{\partial \theta_\gamma} & \frac{\partial H_{jk}^{(0)}}{\partial \theta_\gamma} & \frac{\partial H_{jk}^{(1)}}{\partial \theta_\gamma} \cdots \frac{\partial H_{jk}^{(3)}}{\partial \theta_\gamma} \\ \frac{\partial H_{jj}^{(0)}}{\partial p_\mu} & \frac{\partial H_{jk}^{(0)}}{\partial p_\mu} & \frac{\partial H_{jk}^{(1)}}{\partial p_\mu} \cdots \frac{\partial H_{jk}^{(3)}}{\partial p_\mu} \end{bmatrix} \cdot \begin{bmatrix} v & w \\ A^{(0)} & B^{(0)} \\ \cdot & \cdot \\ \cdot & \cdot \\ A^{(3)} & B^{(3)} \end{bmatrix}$$

$$= \begin{bmatrix} \rho_{\gamma,\alpha\alpha} & \sigma_{\gamma,\alpha\beta}^{(0)} & \cdots & \sigma_{\gamma,\alpha\beta}^{(3)} \\ \epsilon_{\alpha\alpha}^{(\mu)} & S_{\alpha\beta}^{(0\mu)}(\theta_\beta - \theta_\alpha) & \cdots & S_{\alpha\beta}^{(3\mu)}(\theta_\beta - \theta_\alpha) \end{bmatrix}, \tag{3.27}$$

$$1 \leqslant j < k \leqslant N, \quad 1 \leqslant \alpha < \beta \leqslant N, \quad 1 \leqslant \gamma \leqslant N, \quad 1 \leqslant \mu \leqslant 2N(N-1),$$

where the matrices $\partial H_{jj}^{(0)}/\partial \theta_\gamma$, v, and ρ are $N \times N$, the matrices $\partial H_{jk}^{(\lambda)}/\partial \theta_\gamma$ and $\sigma_{\gamma,\alpha\beta}^{(\lambda)}$, with $\lambda = 0, 1, 2, 3$, are $N \times \frac{1}{2}N(N-1)$, the $A^{(\lambda)}$ are all $\frac{1}{2}N(N-1) \times N$, the $\partial H_{jj}^{(0)}/\partial p_\mu$ and the $\epsilon_{\alpha\alpha}^{(\mu)}$ are $2N(N-1) \times N$, the w is $N \times 2N(N-1)$, the $\partial H_{jk}^{(\lambda)}/\partial p_\mu$ and the $S_{\alpha\beta}^{(\lambda\mu)}$ are $2N(N-1) \times \frac{1}{2}N(N-1)$ and the matrices $B^{(\lambda)}$ are $\frac{1}{2}N(N-1) \times 2N(N-1)$. The matrices ρ and the σ appear as we separate the result of differentiation of (3.19) with respect to θ into quaternion components. Moreover, the matrix ρ does not depend on θ_γ, for Θ depends linearly on the θ_γ. The computation of the matrices v, w, $A^{(\lambda)}$, and $B^{(\lambda)}$ is straightforward, but we do not require them. All we need is to note that they are formed of the various components of U, hence do not depend on θ_γ.

Now we take the determinant on both sides of (3.27). The determinant of the first matrix on the left is the Jacobian (3.24). Because the $\sigma^{(\lambda)}$ are all zero, the determinant of the right-hand side breaks into a product of two determinants:

$$\det[\rho_{\gamma,\alpha\alpha}] \det[S_{\alpha\beta}^{(\lambda\mu)}(\theta_\beta - \theta_\alpha)], \tag{3.28}$$

the first one being independent of the θ_γ, whereas the second is

$$\prod_{\alpha<\beta} (\theta_\beta - \theta_\alpha)^4 \det[S_{\alpha\beta}^{(\lambda\mu)}]. \tag{3.29}$$

Thus

$$J(\theta, p) = \prod_{\alpha<\beta} (\theta_\beta - \theta_\alpha)^4 f(p), \tag{3.30}$$

which corresponds to (3.16).

By inserting (3.30) into (3.23) and integrating over the parameters, we obtain the j.p.d.f.

$$P(\theta_1, ..., \theta_N) = \exp\left(-2a \sum_1^N \theta_j^2 + 2b \sum_1^N \theta_j + c\right) \prod_{j<k} (\theta_j - \theta_k). \tag{3.31}$$

As before, we may shift the origin to make $b = 0$ and change the scale of energy to make $a = 1$. Thus the j.p.d.f. for the eigenvalues of matrices in the symplectic ensemble in its simple form is

$$P_{N4}(x_1, ..., x_N) = C_{N4} \exp\left(-2 \sum_1^N x_j^2\right) \prod_{j<k} (x_j - x_k)^4, \tag{3.32}$$

where C_{N4} is a constant.

3.3. Unitary Ensemble

In addition to the real eigenvalues, the number of real independent parameters p_μ needed to specify an arbitrary Hermitian matrix H completely is $N(N-1)$. Equations 3.2 and 3.3 remain unchanged, but (3.4) is replaced by

$$J(\theta, p) = \frac{\partial(H_{11}^{(0)}, ..., H_{NN}^{(0)}, H_{12}^{(0)}, H_{12}^{(1)}, ..., H_{N-1,N}^{(0)}, H_{N-1,N}^{(1)})}{\partial(\theta_1, ..., \theta_N, p_1, ..., p_{N(N-1)})}, \tag{3.33}$$

where $H_{jk}^{(0)}$ and $H_{jk}^{(1)}$ are the real and imaginary parts of H_{jk}. Equations 3.5 to 3.10 are valid if U^T is replaced by U^+. Instead of (3.11) and (3.12), we now have

$$\sum_{j,k} \frac{\partial H_{jk}}{\partial p_\mu} U_{j\alpha}^* U_{k\beta} = S_{\alpha\beta}^{(\mu)}(\theta_\beta - \theta_\alpha), \tag{3.34}$$

$$\sum_{j,k} \frac{\partial H_{jk}}{\partial \theta_\gamma} U_{j\alpha}^* U_{k\beta} = \frac{\partial \Theta_{\alpha\beta}}{\partial \theta_\gamma} = \delta_{\alpha\beta}\delta_{\alpha\gamma}. \tag{3.35}$$

By separating the real and imaginary parts we may write these equations in as

$$\begin{bmatrix} \dfrac{\partial H_{jj}^{(0)}}{\partial \theta_\gamma} & \dfrac{\partial H_{jk}^{(0)}}{\partial \theta_\gamma} & \dfrac{\partial H_{jk}^{(1)}}{\partial \theta_\gamma} \\ \dfrac{\partial H_{jj}^{(0)}}{\partial p_\mu} & \dfrac{\partial H_{jk}^{(0)}}{\partial p_\mu} & \dfrac{\partial H_{jk}^{(1)}}{\partial p_\mu} \end{bmatrix} \cdot \begin{bmatrix} v & w \\ A^{(0)} & B^{(0)} \\ A^{(1)} & B^{(1)} \end{bmatrix}$$

$$= \begin{bmatrix} \rho_{\gamma,\alpha\alpha} & \sigma^{(0)}_{\gamma,\alpha\beta} & \sigma^{(1)}_{\gamma,\alpha\beta} \\ \epsilon^{(\mu)}_{\alpha\alpha} & S^{(0\mu)}_{\alpha\beta}(\theta_\beta - \theta_\alpha) & S^{(1\mu)}_{\alpha\beta}(\theta_\beta - \theta_\alpha) \end{bmatrix}, \tag{3.36}$$

$$1 \leqslant j < k \leqslant N, \qquad 1 \leqslant \alpha < \beta \leqslant N,$$

$$1 \leqslant \mu \leqslant N(N-1), \qquad 1 \leqslant \gamma \leqslant N,$$

where rts of $S^{(\mu)}_{\alpha\beta}$. The matrices $\partial H^{(0)}_{jj}/\partial\theta_\gamma$, v, and ρ are $N \times N$; the $\partial H^{(\lambda)}_{jk}/\partial\theta_\gamma$ and the $\sigma^{(\lambda)}_{\gamma,\alpha\beta}$ are $N \times \frac{1}{2}N(N-1)$; the $A^{(\lambda)}$ are $\frac{1}{2}N(N-1) \times N$; the $\partial H^{(\lambda)}_{jk}/\partial p_\mu$ and $S^{(\lambda\mu)}_{\alpha\beta}$ are $N(N-1) \times \frac{1}{2}N(N-1)$; the $B^{(\lambda)}$ are $\frac{1}{2}N(N-1) \times N(N-1)$; the $\partial H^{(0)}_{jj}/\partial p_\mu$ and the $\epsilon^{(\mu)}_{\alpha\alpha}$ are $N(N-1) \times N$; and the matrix w is $N \times N(N-1)$. To compute v, w, $A^{(\lambda)}$, ρ, ϵ, $\sigma^{(\lambda)}$, etc., is again straightforward, but we do not need them explicitly. What we want to emphasize is that they are either constructed from the components of U or arise from the differentiation of Θ with respect to θ_j and consequently are all independent of the eigenvalues θ_j. Similarly, $S^{(\mu)}$ is independent of θ_j. One more bit of information we need is that which can easily be verified.

Thus by taking the determinants on both sides of (3.36) and removing the factors $(\theta_\beta - \theta_\alpha)$ we have

$$J(\theta, p) = \prod_{\alpha<\beta} (\theta_\beta - \theta_\alpha)^2 f(p), \tag{3.37}$$

where $f(p)$ is some function of the p_μ.

By inserting (3.37) into (3.3) and integrating over the parameters p_μ we get the j.p.d.f. for the eigenvalues of matrices in the unitary ensemble

$$P(\theta_1, \ldots, \theta_N) = \exp\left(-a\sum_1^N \theta_j^2 + b\sum_1^N \theta_j + c\right) \prod_{j<k} (\theta_j - \theta_k)^2, \tag{3.38}$$

and, as before, by a proper choice of the origin and the scale of energy we have

$$P_{N2}(x_1, \ldots, x_N) = C_{N2} \exp\left(-\sum_1^N x_j^2\right) \prod_{j<k} (x_j - x_k)^2. \tag{3.39}$$

We record (3.18), (3.32), and (3.39) as a theorem.

Theorem 3.2. *The joint probability density function for the eigenvalues of matrices from a Gaussian ensemble is given by*

$$P_{N\beta}(x_1, \ldots, x_N) = C_{N\beta} \exp\left(-\tfrac{1}{2}\beta \sum_1^N x_j^2\right) \prod_{j<k} |x_j - x_k|^\beta, \tag{3.40}$$

where $\beta = 1$ if the ensemble is orthogonal, $\beta = 4$ if it is symplectic, and $\beta = 2$ if it is unitary.

$$\int_{-\infty}^{\infty} \cdots \int P_{N\beta}(x_1, \ldots, x_N)\, dx_1 \cdots dx_N = 1. \tag{3.41}$$

In the following chapters [see (5.36), (6.3), and (6.49)] we calculate $C_{N\beta}$ for the physically interesting cases $\beta = 1, 2$, and 4. For these values of β, $C_{N\beta}$ is given by

$$C_{N\beta}^{-1} = (2\pi)^{(1/2)N\beta - (1/2)N - (1/4)\beta N(N-1)} [\Gamma(1 + \tfrac{1}{2}\beta)]^{-N} \prod_{j=1}^N \Gamma(1 + \tfrac{1}{2}\beta j). \tag{3.42}$$

It is possible to understand the different powers of β that appear in (3.40) by a simple mathematical argument based on counting dimensions. The dimension of the space T_{1G} is $\frac{1}{2}N(N+1)$, whereas the dimension of the subspace T'_{1G}, composed of the matrices in T_{1G} with two equal eigenvalues, is $\frac{1}{2}N(N+1) - 2$. Because of the single restriction, the equality of two eigenvalues, the dimension should normally have decreased by one; as it is decreased by two it indicates a factor in (3.40) linear in $(x_j - x_k)$. Similarly, when $\beta = 2$, the dimension of T_{2G} is N^2, whereas that of T'_{2G} is $N^2 - 3$. When $\beta = 4$, the dimension of T_{4G} is $N(2N-1)$, whereas that of T'_{4G} is $N(2N-1) - 5$ (see Appendix A.2).

4 / Gaussian Ensembles

4.1. The Partition Function†

Consider a gas of N point charges with positions $x_1, x_2, \ldots, x_N$ free to move on the infinite straight line $-\infty < x < \infty$. Suppose that the potential energy of the gas is given by

$$W = \frac{1}{2}\sum_i x_i^2 - \sum_{i<j} \ln | x_i - x_j |. \tag{4.1}$$

The first term in W represents a harmonic potential which attracts each charge independently toward the point $x = 0$; the second term represents an electrostatic repulsion between each pair of charges. The logarithmic function comes in if we assume the universe to be two-dimensional. Let this charged gas be in thermodynamical equilibrium at a temperature T, so that the probability density of the positions of the N charges is given by

$$P(x_1, \ldots, x_N) = C \exp\left(\frac{-W}{kT}\right), \tag{4.2}$$

where k is the Boltzmann constant. We immediately recognize that (4.2) is identical to (3.40), provided β is related to the temperature by

$$\beta = (kT)^{-1}. \tag{4.3}$$

This system of point charges in thermodynamical equilibrium is called the Coulomb gas model, corresponding to the Gaussian ensembles.

Following Dyson [1–3], we can define various expressions that relate to our energy-level series in complete analogy with the classical notions of entropy, specific heat, and the like. These expressions,

† Mehta and Dyson [1].

7 / Brownian Motion Model†

7.1. Stationary Ensembles

In Chapter 4 we exploited the idea that the probability $P(x_1,..., x_N)$, (3.40) for the eigenvalues of a random matrix to lie in unit intervals around the points $x_1,..., x_N$,

$$P(x_1,..., x_N) = C_{N\beta}e^{-\beta W}, \tag{7.1}$$

$$W = \frac{1}{2}\sum_1^N x_j^2 - \sum_{i<j} \ln | x_i - y_j |, \tag{7.2}$$

is identical with the probability density of the positions of N unit charges free to move on the infinite straight line $-\infty < x < \infty$ under the influence of forces derived from the potential energy (7.2), according to the laws of classical mechanics, in a state of thermodynamical equilibrium at a temperature given by

$$kT = \beta^{-1}. \tag{7.3}$$

This system of point charges in thermodynamical equilibrium is called the stationary Coulomb gas model or simply the Coulomb gas model, which corresponds to the Gaussian ensembles.

7.2. Nonstationary Ensembles

In this chapter we present an idea of Dyson, generalizing the notion of a matrix ensemble in such a way that the Coulomb gas model acquires meaning not only as a static model in timeless thermodynamical equilibrium but as a dynamical system that may be in an arbitrary nonequilibrium state changing with time. The word "time" in this

† Dyson [4].

chapter always refers to a fictitious time which is a property of the mathematical model and has nothing to do with real physical time.

When we try to interpret Coulomb gas as a dynamical system, we naturally consider it first as an ordinary conservative system in which the charges move as Newtonian particles and exchange energy with one another only through the electric forces arising from the potential (7.2). We then have to give meaning to the velocity of each particle and to regulate the behavior of the random matrix H in such a way that the eigenvalues have the normal Newtonian property of inertia. No reasonable way of doing this has yet been found. Perhaps there is no such way.

After considerable and fruitless efforts to develop a Newtonian theory of ensembles, Dyson [4] discovered that the correct procedure is quite different and much simpler. The x_j should be interpreted as positions of particles in Brownian motion [Chandrasekhar, 1; Uhlenbeck and Ornstein, 1; Wang and Uhlenbeck, 1]. This means that the particles have no well-defined velocities nor do they possess inertia. Instead, they feel frictional forces resisting their motion. The gas is not a conservative system, for it is constantly exchanging energy with its surroundings through these frictional forces. The potential (7.2) still operates on the particles in the following way. The particle at x_j experiences an external electric force

$$E(x_j) = -\frac{\partial W}{\partial x_j} = -x_j + \sum_{\substack{i \\ (i \neq j)}} \frac{1}{x_j - x_i} \tag{7.4}$$

in addition to the local frictional force and the constantly fluctuating force giving rise to the Brownian motion.

The equation of motion of the Brownian particle at x_j may be written as

$$\frac{d^2x_j}{dt^2} = -f\frac{dx_j}{dt} + E(x_j) + A(t), \tag{7.5}$$

where f is the friction coefficient and $A(t)$ is a rapidly fluctuating force. For $A(t)$ we postulate the usual properties [Uhlenbeck and Ornstein, 1]

$$\langle A(t_1)\, A(t_2) \cdots A(t_{2n+1})\rangle = 0, \tag{7.6}$$

$$\langle A(t_1)\, A(t_2) \cdots A(t_{2n})\rangle = \sum_{\text{pairs}} \langle A(t_i)\, A(t_j)\rangle \langle A(t_k)\, A(t_l)\rangle \cdots, \tag{7.7}$$

and

$$\langle A(t_1)\, A(t_2)\rangle = \frac{2kT}{f}\,\delta(t_1 - t_2), \tag{7.8}$$

where the summation in (7.7) extends over all distinct ways in which the $2n$ indices can be divided into n pairs.

There is nothing new in the integration of the Langevin equation (7.5). After long enough time for the effect of the initial velocity to become negligible, let x_1 , x_2 ,..., x_N be the positions of the particles at time t. At a later time $t + \delta t$ let these positions be changed to $x_1 + \delta x_1$, $x_2 + \delta x_2$,..., $x_N + \delta x_N$. The δx_j , $j = 1, 2,..., N$, will in general be different for every member of the ensemble. They are random variables. Using (7.6), (7.7), and (7.8) we find that to the first order in the small quantities

$$f\langle \delta x_j\rangle = E(x_j)\,\delta t, \tag{7.9}$$

$$f\langle (\delta x_j)^2\rangle = 2kT\,\delta t, \tag{7.10}$$

and all other ensemble averages, for example, $\langle \delta x_j\, \delta x_l\rangle$, $\langle (\delta x_j)^2\, \delta x_l\rangle$, $\langle (\delta x_j)^3\rangle$, are of a higher order in δt.

An alternative description of Brownian motion is obtained by deriving the Fokker–Planck or Smoluchowski equation. Let $P(x_1$, x_2 ,..., x_N ; $t)$ be the time-dependent joint probability density that the particles will be at the positions x_j at time t. Assuming that the future evolution of the system is completely determined by its present state, with no reference to its past (that is, the process is a Markov process), we obtain

$$P(x_1 ,..., x_N\, ;\, t + \delta t) = \int\cdots\int P(x_1 - \delta x_1 ,..., x_N - \delta x_N\, ;\, t)$$

$$\times\, \psi(x_1 - \delta x_1 ,..., x_N - \delta x_N\, ;\, \delta x_1 ,..., \delta x_N\, ;\, \delta t)\, d(\delta x_1)\cdots d(\delta x_N), \tag{7.11}$$

where ψ under the integral sign is the probability that the positions of the particles will change from $x_1 - \delta x_1$,..., $x_N - \delta x_N$ to x_1 ,..., x_N in a time interval δt. Expanding both sides of (7.11) in a power series of δx_j , δt, using (7.9) and (7.10), and going to the limit $\delta t \to 0$, we get [Uhlenbeck and Ornstein, 1]

$$f\frac{\partial P}{\partial t} = \sum_{j=1}^{N}\left\{kT\frac{\partial^2 P}{\partial x_j} - \frac{\partial}{\partial x_j}[E(x_j)\, P]\right\}. \tag{7.12}$$

Equation 7.12 describes the development of the Coulomb gas with time. If we start from an arbitrary initial probability density P at time $t = t_0$, a unique solution of (7.12) will exist for all $t \geqslant t_0$. Any solution of this sort we call a *time-dependent Coulomb gas model.*

Equation 7.12 implies in turn (7.9) and (7.10). To see this we multiply both sides of (7.12) by x_j and integrate over all x_i. Making the usual assumptions that $P(x_1, ..., x_N ; t)$, as well as its derivatives, vanish quite fast on the boundary, we get on partial integration

$$f \frac{d}{dt} \langle x_j \rangle = \langle E(x_j) \rangle, \tag{7.13}$$

where

$$\langle F \rangle = \int F(x_1, ..., x_N)\, P(x_1, ..., x_N ; t)\, dx_1 \cdots dx_N$$

is the ensemble average of F. Starting at the positions $x_1, ..., x_N$ and executing the motion for a small time interval δt, we find that (7.13) is the same as (7.9). Similarly, by multiplying by x_j^2 and integrating (7.12) we have

$$f \frac{d}{dt} \langle x_j^2 \rangle = 2kT + 2\langle x_j\, E(x_j) \rangle,$$

which together with $\langle (\delta x_j)^2 \rangle = \langle x_j^2 \rangle - \langle x_j \rangle^2$ yields (7.10).

Thus the descriptions of the motion by (7.9) and (7.10), and by (7.12) are equivalent. Also there exists a unique solution to (7.12) which is independent of time, and this time independent solution is given by (7.1) and (7.2).

A Brownian motion model can also be constructed for the matrix H, of which x_j are the eigenvalues. The independent real parameters $H_{ij}^{(\lambda)}$; $1 \leqslant i < j \leqslant N$, $0 \leqslant \lambda \leqslant \beta - 1$, which determine all the matrix elements of H, are $p = \frac{1}{2}N(N+1) + \frac{1}{2}N(N-1)(\beta - 1)$ in number. Let us denote them by H_μ, where μ is a single index that runs from 1 to p and replaces the three indices i, j, and λ. Suppose that the parameters H_μ have the values $H_1, H_2, ..., H_p$ at time t and $H_1 + \delta H_1, ..., H_p + \delta H_p$ at a later time $t + \delta t$. Brownian motion of H is defined by requiring that each δH_μ be a random variable with the ensemble averages

$$f \langle \delta H_\mu \rangle = -H_\mu\, \delta t, \tag{7.14}$$

$$f \langle (\delta H_\mu)^2 \rangle = g_\mu kT\, \delta t, \tag{7.15}$$

where

$$g_\mu = g_{ij}^{(\lambda)} = 1 + \delta_{ij} = \begin{cases} 2, & \text{if } i = j, \\ 1, & \text{if } i \neq j. \end{cases} \tag{7.16}$$

All other averages are of a higher order in δt. This is a Brownian motion of the simplest type, the various components H_μ being completely uncoupled and each being subject to a fixed simple harmonic force. The Smoluchowski equation which corresponds to (7.14) and (7.15) is

$$f\frac{\partial P}{\partial t} = \sum_\mu \left[\tfrac{1}{2} g_\mu kT \frac{\partial^2 P}{\partial H_\mu^{\,2}} + \frac{\partial}{\partial H}(H_\mu P)\right], \tag{7.17}$$

where $P(H_1, \ldots, H_p ; t)$ is the time-dependent joint probability density of H_μ. The solution to (7.17) which corresponds to a given initial condition $H = H'$ at $t = 0$, is known explicitly [Uhlenbeck and Ornstein, 1].

$$P(H, t) = C(1 - q^2)^{-(1/2)p} \exp\left[-\frac{\operatorname{tr}(H - qH')^2}{2kT(1 - q^2)}\right], \tag{7.18}$$

$$q = \exp\left[\frac{-t}{f}\right]. \tag{7.19}$$

The solution shows that the Brownian process is invariant under symmetry preserving unitary transformations of the matrix H; in fact, the awkward-looking factor g_μ in (7.15) is put in to ensure this invariance. When $t \to \infty$, $q \to 0$, and the probability density (7.18) tends to the stationary form,

$$P(H_1, \ldots, H_p) = (\text{constant}) \exp\left(\frac{-1}{2kT} \operatorname{tr} H^2\right), \tag{7.20}$$

which is the unique time-independent solution of (7.17). Note that with the relation (7.3) between β and the temperature kT (7.20) is essentially the same as (2.72).

We are now in a position to state the main result of this chapter.

Theorem 7.1. *When the matrix H executes a Brownian motion according to the simple harmonic law* (7.14), (7.15), *starting from any initial conditions whatever, its eigenvalues $x_1, x_2, \ldots, x_N$ execute a Brownian motion that obeys the equations of motion* (7.9), (7.10), *and* (7.12) *of the time-dependent Coulomb gas.*

To prove the theorem we need only show that (7.9) and (7.10) follow from (7.14) and (7.15). Suppose, then, that (7.14) and (7.15) hold. We have seen that the process described by (7.14) and (7.15) is independent of the representation of H. Therefore we may choose the representation so that H is diagonal at time t. The instantaneous values of H_μ at time t are then

$$H_{jj}^{(0)} = x_j , \qquad j = 1, 2,..., N, \tag{7.21}$$

and all other components are zero. At a later time $t + \delta t$ the matrix $H + \delta H$ is no longer diagonal and its eigenvalues $x_j + \delta x_j$ must be calculated by perturbation theory. We have to the second order in δH

$$\delta x_j = \delta H_{jj}^{(0)} + \sum_{\substack{i \\ (i \neq j)}} \sum_{\lambda=0}^{\beta-1} \frac{(\delta H_{ij}^{(\lambda)})^2}{x_j - x_i} . \tag{7.22}$$

Higher terms in the perturbation series will not contribute to the first order in δt. When we take the ensemble average on each side of (7.22) and use (7.14), (7.21), (7.15), (7.3), and (7.4), the result is (7.9). When we take the ensemble average of $(\delta x_j)^2$, only the first term on the right side of (7.22) contributes to the order δt, and this term gives (7.10) by virtue of (7.15) and (7.16). The theorem is thus proved.

When the limit $t \to \infty$ is taken, Theorem 7.1 reduces to Theorem 3.2. This new proof of Theorem 3.2 is in some respects more illuminating. It shows how the repulsive Coulomb potential (7.2), pushing apart each pair of eigenvalues, arises directly from the perturbation formula (7.22). It has long been known that perturbations generally split levels that are degenerate in an unperturbed system. We now see that this splitting effect of perturbations is quantitatively identical with the repulsive force of the Coulomb gas model.

Theorem 7.1 is a much stronger statement than Theorem 3.2. It shows that the electric force (7.4), acting on the eigenvalues x_j, has a concrete meaning for any matrix H whatever, not only for an ensemble of matrices in stationary thermal equilibrium. The force $E(x_j)$ is precisely proportional to the mean rate of drift of x_j which occurs when the matrix H is subjected to a random perturbation.

7.3. Some Ensemble Averages

We now describe a general property of the time-dependent Coulomb gas model which may be used to calculate a few ensemble averages. Dyson observed that if $G = G(x_1, \ldots, x_N)$ is any function of the positions of the charges, not depending explicitly on time, then the time variation of $\langle G \rangle$, the ensemble average of G, is governed by the equation

$$f\frac{d}{dt}\langle G\rangle = -\sum_j \left\langle \frac{\partial W}{\partial x_j}\frac{\partial G}{\partial x_j}\right\rangle + kT\sum_j \left\langle \frac{\partial^2 G}{\partial x_j^2}\right\rangle. \tag{7.23}$$

This equation is obtained by multiplying (7.12) throughout by G and partial integrations; W is given by (7.2).

As a first example, choose

$$R = \sum_j x_j^2 \tag{7.24}$$

for G so that

$$\frac{\partial W}{\partial x_j}\frac{\partial R}{\partial x_j} = -2\sum_{\substack{i\\(i\neq j)}} \frac{x_j}{x_j - x_i} + 2x_j^2,$$

$$\frac{\partial^2 R}{\partial x_j^2} = 2,$$

and (7.23) becomes

$$\begin{aligned} f\frac{\partial\langle R\rangle}{\partial t} &= -2\langle R\rangle + N(N-1) + 2kTN \\ &= 2(R_\infty - \langle R\rangle), \end{aligned} \tag{7.25}$$

with

$$R_\infty = \tfrac{1}{2}N(N-1) + kTN. \tag{7.26}$$

The solution of (7.25) is

$$\langle R\rangle = R_0 q^2 + R_\infty(1 - q^2), \tag{7.27}$$

where q is given by (7.19) and R_0 is the value of $\langle R\rangle$ at $t = 0$. Equation 7.27 shows that the ensemble average $\langle R\rangle$ approaches its equilibrium value R_∞ with exponential speed as $t \to \infty$.

Next take $G = W$ in (7.23), so that

$$\begin{aligned} \left(\frac{\partial W}{\partial x_j}\right)^2 = &\sum_{\substack{i\\(i\neq j)}} \left[\left(\frac{1}{x_j - x_i}\right)^2 - \frac{2x_j}{x_j - x_i}\right] \\ &+ \sum_{\substack{i,l\\(i,j,l \text{ all}\\ \text{different})}} [(x_j - x_i)(x_j - x_l)]^{-1} + x_j^2 \end{aligned} \tag{7.28}$$

and

$$\frac{\partial^2 W}{\partial x_j^2} = -1 + \sum_{\substack{i \\ (i \neq j)}} (x_j - x_i)^{-2}. \tag{7.29}$$

On performing a summation over j the second term in (7.28) drops out (cf. Appendix A.20), whereas the second term in the first bracket gives $-2N(N-1)$. Substituting in (7.23) and simplifying, we get

$$f\frac{\partial \langle W \rangle}{\partial t} = (kT - 1) \sum_{\substack{i,j \\ (i \neq j)}} \langle (x_j - x_i)^{-2} \rangle + (N^2 - N + NkT) - \sum_j \langle x_j^2 \rangle. \tag{7.30}$$

For the stationary Coulomb gas at temperature kT the left side of (7.30) vanishes and (7.26) may be used on the right. Thus we find a "virial theorem" for the stationary gas:

$$\sum_{\substack{i,j \\ (i \neq j)}} \langle (x_j - x_i)^{-2} \rangle = \frac{N(N-1)}{2(1 - kT)}. \tag{7.31}$$

The probability density of eigenvalues becomes proportional to $| x_i - x_j |^{\beta}$, when two eigenvalues x_i, x_j come close together. The ensemble average of $(x_j - x_i)^{-2}$ is therefore defined only for $\beta > 1$ and (7.30) and (7.31) hold only for $kT < 1$.

An especially interesting case, $\beta = 1$, requires a passage to the limit in (7.30). As $kT \to 1$, we have for any fixed value of Δ

$$\begin{aligned} \lim(kT - 1) \int_{-\Delta}^{\Delta} | y |^{\beta - 2} \, dy &= \lim(kT - 1)(\beta - 1)^{-1} \, 2\Delta^{\beta - 1} \\ &= -2. \end{aligned} \tag{7.32}$$

We obtain the correct limit in (7.30) if we replace

$$(kT - 1)(x_j - x_i)^{-2}$$

with

$$-2(x_j - x_i)^{-1} \, \delta(x_j - x_i), \tag{7.33}$$

which has a well-defined meaning as an ensemble average when $kT = 1$, for the probability density then contains a factor $| x_j - x_i |$; (7.30) thus becomes in the limit

$$f\frac{d \langle W \rangle}{dt} = -2 \sum_{\substack{i,j \\ (i \neq j)}} \langle | x_j - x_i |^{-1} \, \delta(x_j - x_i) \rangle + N^2 - \sum_j \langle x_j^2 \rangle, \qquad kT = 1. \tag{7.34}$$

The corresponding "virial theorem" is

$$\sum_{\substack{i,j \\ (i \neq j)}} \langle (x_j - x_i)^{-1} \, \delta(x_j - x_i) \rangle = \tfrac{1}{4} N(N-1), \qquad kT = 1 \qquad (7.35)$$

for the stationary gas.