

*Per K. Jakobsen, V.V. Lychagin*

## MAXIMUM ENTROPY WAVE FUNCTIONS

ABSTRACT. In this paper we use the classical Maximum Entropy principle to define maximum entropy wave functions. These are wave functions that maximize the entropy among all wave functions satisfying a finite set of constraints in the form of expectation values. This leads to a nonlinear equation for the wave function that reduces to the usual stationary Schrödinger equation if the energy is the only constraint and the value of the constraint is an eigenvalue. We discuss the extension of the thermodynamical formalism to this case and apply our general formalism to several simple quantum systems, the two-level atom, the particle in a box, the free particle and the Harmonic Oscillator and compare with the results obtained by applying the usual von Neumann quantum statistical method to the same systems.

### CONTENTS

1. Introduction	30
2. The maximum entropy wavefunction	31
3. The thermodynamical formalism	35
4. The two-level atom	37
5. The particle in a box	43
6. The free particle	52
7. The Harmonic Oscillator	53
References	55

---

*2000 Mathematical Subject Classification.* 82B99,81Q99,81P15,94A17.

*Key words and phrases.* Maximum entropy principle, Quantum mechanics, Wavefunctions, Probability theory, Density matrix.

## 1. INTRODUCTION

Probability theory is at its most useful and efficient if it is recognized that it is a theory of inference; the central problem is to represent data in terms of probability distributions and then use these distributions to predict the results of measurements yet to be performed. The probability distributions that we use do not describe any property of the world, only a certain state of information about the world and no frequency interpretation is associated with the distribution. This point of view on probability theory was in fact the point of view of the founding fathers of probability theory and Laplace, in particular, used it to great effect in his work in physics. After the death of Laplace the inference point of view went out of fashion and was not reinstated until last century by the work of Jeffrey[3], Cox[1],[2] and Jaynes[4]. In any application of probability theory as inference the main tools are Bayes' theorem and the Maximum Entropy principle (MaxEnt). The MaxEnt principle has a long and convoluted history that is a reflection of the equally long and convoluted history of probability and statistics. In rudimentary form it was present already in the work of Bernoulli and Laplace in the form of the Principle of Insufficient Reason. Jaynes has shown how efficient the inference point of view is when it is applied to physics in general and statistical mechanics in particular[5]. The canonical equilibrium ensembles and the whole thermodynamical formalism of equilibrium statistical mechanics follows directly from the MaxEnt principle. The only physics input needed is a description of phase space and of the observables. Today hardly anyone would object to the use of Bayes' theorem and the MaxEnt principle in physics and elsewhere, their success in practical terms are undisputed.

In quantum mechanics the fundamental quantity is not a probability density but a complex wave function whose norm square is interpreted as a probability density. This much is universally accepted. The problems starts when we ask if this probability should be taken to be subjective, representing a state of knowledge of an observer, or as in some sense real and irreducible. This unresolved question is as old as quantum mechanics itself and all the giants of physics from the last century has addressed this question in one way or another. It is not our intention to address this somewhat philosophical question, we have a more practical problem in mind.

What are the consequences of treating the problem of assigning a wave function to a physical systems as a problem in statistical inference? It is certainly a natural question to ask in the light of the probabilistic nature

of quantum mechanics and the undisputed practical success of probability theory as statistical inference in applications to classical physics.

## 2. THE MAXIMUM ENTROPY WAVEFUNCTION

When we apply the MaxEnt principle to classical physics the fundamental quantity is a probability density and this is the quantity we vary when we seek the extremal of the entropy subject to the given constraints. To be precise let us assume that a phase space  $\Omega$  and an observable, for example the energy  $H$ , is given. Then the the MaxEnt principle consists of choosing the probability density  $\rho$  that maximize the relative entropy

$$S(\rho) = - \int \rho \ln \frac{\rho}{\rho_0} dx$$

subject to the constraints

$$\int \rho dx = 1$$

$$\int H \rho dx = E$$

Here the probability density  $\rho_0$  is the prior and summarize what is know before any additional constraints are applied. The solution to this constrained optimization problem is the canonical distribution

$$\rho = \frac{\rho_0}{Z} e^{-\beta H}$$

where  $Z(\beta)$  is the partition function determined through normalization

$$Z(\beta) = \int \rho_0 e^{-\beta H} dx$$

If we define  $h = \ln(Z)$  we find the identities

$$E = -\partial_\beta h$$

$$S = h + \beta E$$

$$dS = -\beta dE$$

This constitute the thermodynamical formalism for this simple case of one observable and no external parameters. If more observables and external parameters are included we get the thermodynamical formalism for equilibrium statistical mechanics if  $S$  is identified with the physical entropy (Boltzman's constant set equal to 1). Note however that the thermodynamical formalism applies in any probabilistic setting where your data consists of a prior distribution and constraints in the form of expectations of observables. This point has been stressed by Jaynes. In

classical physics and in classical statistics in general the observables are assumed to be functions defined on some statespace. In all such cases we are lead to the thermodynamical formalism as described above. Expressed in terms of operators this means that all observables are assumed to commute. The two main features that distinguish quantum physics from classical physics and statistics are that the state is now represented by a complex density (the wave function) and that observables in general do not commute.

In order to apply the maximum entropy principle we first need to specify a prior. In the case of classical statistics the prior consists of two pieces of information. The first is a description of the space of events and the second is to assign a prior probability distribution to the events. In classical probability and statistics only the second of these is usually thought of as prior information. In application to quantum mechanics, where the algebra of observables is noncommutative, the choice of space of events is far from self-evident, there are several possibilities, one for each complete set of commuting observables. Each such set will give a space of events that is identified with the spectrum of the commutative subalgebra generated by the complete set of commuting observables. After such a set has been selected the specification of the prior is completed by giving a probability distribution  $\rho$  on the space of events. There is, as usual, no general method for converting prior information into a prior distribution, but no general method exists in the classical setting either. The same particular methods that have been developed for classical statistics can of course be used here also, but there is an additional method that can be used in the quantum context. In an ideal experiment the system is prepared in a eigenstate for some complete set of observables. The probability distribution corresponding to this eigenstate is the right prior to use in such cases.

Assuming that a prior has been selected, the relative entropy is given by the functional

$$S(\varphi, \bar{\varphi}) = - \int \varphi \bar{\varphi} \ln\left(\frac{\varphi \bar{\varphi}}{\rho}\right) dx,$$

where  $\varphi(x)$  now is a complex valued function on the space of events and  $\rho$  is the prior as described above. Note that the entropy is considered to be a function of the wave function and its complex conjugate. This choice reflects the fact that the wavefunction, not the corresponding probability density, is the fundamental variable in quantum mechanics.

Let us first assume that a single observable in the form of an operator  $H$  is given. The case of several operators is treated in the same manner and the expressions for this case will be written down later. The maximum entropy principle now consists of finding the wave function  $\varphi$  such that the entropy is maximal subject to the constraints

$$\begin{aligned}\langle \varphi, \varphi \rangle &= 1 \\ \langle \varphi, H\varphi \rangle &= E\end{aligned}$$

We solve the variational problem using Lagrange multipliers as in the classical case and we vary  $\varphi$  and  $\bar{\varphi}$  independently, as is usual in variational principles involving complex fields. We get the following two equations

$$\begin{aligned}\beta H\varphi + \varphi \ln \left( \frac{\varphi \bar{\varphi}}{\rho} \right) + h\varphi &= 0 \\ \beta H^* \varphi + \varphi \ln \left( \frac{\varphi \bar{\varphi}}{\rho} \right) + h\varphi &= 0,\end{aligned}$$

where  $H^*$  is the adjoint of  $H$ . The equations are consistent only if the observable is a selfadjoint operator. Thus the variational problem can only have a solution if the observable is a selfadjoint operator. If the observable is selfadjoint we are left with the single fundamental equation

$$\beta H\varphi + \varphi \ln \left( \frac{\varphi \bar{\varphi}}{\rho} \right) + h\varphi = 0$$

The Lagrange multipliers are  $\beta$  and  $\lambda$  and we have defined  $h = 1 + \lambda$ . If the observable  $H$  commute with the prior, meaning that it is an element in the algebra of observables generated by the complete set of commuting observables used to specify the space of events, then it is a diagonal operator and  $\varphi$  can be canceled and the equation simplifies to

$$\beta H + \ln \left( \frac{\varphi \bar{\varphi}}{\rho} \right) + h = 0$$

The solution of this equation is the usual canonical ensemble from classical statistics.

$$\varphi \bar{\varphi} = \frac{\rho}{Z} e^{-\beta H}$$

In general however the operator does not commute with the prior and in order to find the maximum entropy wave function we must solve the fundamental equation subject to the two constraints

$$\begin{aligned}\langle \varphi, \varphi \rangle &= 1 \\ \langle \varphi, H\varphi \rangle &= E\end{aligned}$$

Define  $T = \frac{1}{\beta}$  as in classical statistical mechanics. The fundamental equation can then be written as

$$H\varphi + T\varphi \ln \left( \frac{\varphi\bar{\varphi}}{\rho} \right) = \gamma\varphi,$$

where we have defined  $\gamma = -h/\beta$  and  $T$  and  $\gamma$  must be determined by the constraints. Note that if the expectation  $E$  is an eigenvalue of  $H$  then we find that  $T = 0$  and  $\gamma = E$  and the maximum entropy wave function satisfy the stationary Schrödinger equation.

$$H\varphi = E\varphi$$

In classical mechanics the parameter  $T$  is interpreted as temperature and if we use the same language here, we observe that energy eigenstates corresponds to maximum entropy wavefunctions in the zero temperature limit. In this limit the prior  $\rho$ , has no effect. As the temperature  $T$  increase the influence of the prior increases and in the high temperature limit the prior dominates. In fact when  $T \rightarrow \infty$  we get  $\varphi\bar{\varphi} = \rho$ .

Our fundamental equation is a stationary Schrödinger equation with a logarithmic nonlinearity. Nonlinear corrections to the time dependent Schrödinger equations has been proposed on several occasions[8],[7],[6] and the logarithmic nonlinearity has been argued for as especially natural[6] for several mathematical and physical reasons. Most of these corrections has however not survived for long; it is very difficult and maybe impossible to destroy the linearity of the dynamic Schrödinger equation and at the same time retain its unquestionable success in explaining a huge range of atomic and molecular phenomena.

We say this in order to stress that we are not proposing a new universal evolution equation for the wavefunction. Time plays no role in our application of the MaxEnt principle. Our equation is not a dynamic equation and there is nothing universal about the parameter  $T$ , it is just the dual of the energy in the maximum entropy sense. Its value is fixed by the measured mean of the energy. If other observables, instead of or in addition to the energy are measured, the MaxEnt principle produce a similar equation and now there is obviously no connection to the classical Schrödinger equation.

Let  $A_1, A_2, \dots, A_n$  be  $n$  observables represented by selfadjoint operators. Then the maximum entropy wave function for this situation is a solution to

$$\sum_i^n \lambda_i A_i \varphi + \varphi \ln \left( \frac{\varphi\bar{\varphi}}{\rho} \right) + h\varphi = 0$$

subject to the constraints

$$\begin{aligned}\langle \varphi, \varphi \rangle &= 1 \\ \langle A_i \rangle &= a_i, \quad i = 1, \dots, n\end{aligned}$$

The constraints are used to solve for the Lagrange multipliers  $h$  and  $\lambda_1, \dots, \lambda_n$ . A solution to this constrained variational problem is the wave function that best, in the maximum entropy sense, represents a situation where the means of  $n$ , possibly noncommuting observables, are known. It is possible for the problem to have no solution. This is similar to the situation in the classical case. For example if you try to use the observable  $a(x) = x$  when the phase space is the real line you will find that the solution can not be normalized. This is not a defect of the method, it merely indicates that too little information is given for a probability assignment to be possible. There can be more than one solution. In classical thermodynamics this situation is well known and is taken to mean that there are several coexisting thermodynamic phases present in the system. Which one of these cases are realized can only be determined by actually solving the constrained variational problem.

### 3. THE THERMODYNAMICAL FORMALISM

In the classical setting the thermodynamical formalism plays an important role. As long as we only ask for means of observables that are functions of the observables used as constraints in the MaxEnt principle, we never have to know what the probability density is. The mean of all such observables are found by computing partial derivatives of the partition function with respect to the corresponding Lagrange multipliers. If we for example have a single constraint  $a(x)$  then we have

$$\begin{aligned}\langle a \rangle &= -\partial_\lambda \ln Z(\lambda) \\ \langle a^2 \rangle - \langle a \rangle^2 &= -\partial_\lambda \partial_\lambda \ln Z(\lambda),\end{aligned}$$

where  $Z(\lambda)$  is the partition function for the situation of one constraint  $a(x)$ . We will now see how much of the thermodynamical formalism extends to the quantum mechanical case. Since different observables in the thermodynamical formalism are essentially represented by partial derivatives with respect to corresponding Lagrange multipliers we would not expect the formalism to carry over to the quantum case. This is because partial derivatives commutes whereas the observables typically don't. We will see that a small part of the formalism actually does carry over.

Let us for simplicity consider the case of one selfadjoint observable  $A$ . Then the wave function  $\varphi$  is a solution to the MaxEnt principle if it solves the following equation

$$\lambda A\varphi + \varphi \ln \left( \frac{\varphi \bar{\varphi}}{\rho} \right) + h\varphi = 0$$

Define a new function  $\theta$  by

$$\varphi = \frac{1}{\sqrt{Z}}\theta,$$

where  $h = \ln Z$ . Then we find that  $\theta$  satisfy the simplified equation

$$\lambda A\theta + \theta \ln \frac{\theta \bar{\theta}}{\rho} = 0$$

Using the normalization condition for  $\varphi$  we find that

$$Z = \int \theta \bar{\theta} dx$$

The quantity  $Z$  is clearly the analog of the partition function for the classical case. Using the definition of the expectation of an operator in quantum mechanics we get

$$a = \langle A \rangle = \frac{1}{Z} \int \theta \bar{A} \theta dx$$

After some simple manipulation of the equation for  $\theta$  we find the identity

$$\theta \bar{A} \theta = \lambda \bar{\partial}_\lambda \theta A \theta - \lambda \theta \bar{A} \partial_\lambda \theta - \partial_\lambda (\theta \bar{\theta})$$

If we integrate this identity and use the self adjointness of the operator  $A$  we find

$$a = \langle A \rangle = -\partial_\lambda h$$

Using similar manipulation of the equation for  $\theta$  we find that the entropy evaluated at the maximum entropy wave function is given in terms of the Lagrange multipliers as

$$S = h - \lambda \partial_\lambda h$$

These two last formulas are exactly identical to the formulas we get in the classical case of one observable. If we have  $n$  selfadjoint observables the same calculation goes through and we find

$$\begin{aligned} h &= \ln Z(\lambda_1, \dots, \lambda_n) \\ a_i &= \langle A_i \rangle = -\partial_{\lambda_i} h \\ S &= h - \sum_i^n \lambda_i \partial_{\lambda_i} h \end{aligned}$$

just as in the classical case. It is however not true that we can compute expectations of polynomials in the observables  $A_1, \dots, A_n$  in terms of partial derivatives of the partition function. As explained in the beginning of this section no such simple rules should be expected because of the noncommutativity of quantum observables. In fact it is not even possible to compute the expectation of  $A^2$  in terms of the partition function in the situation when there is only one constraint given by the operator  $A$ . We do not at this point know if some quantum version of the thermodynamical formalism exists at all or in terms of what quantities it should be formulated.

In the rest of the paper we will apply the general formalism to several well known quantum models. These are the two-level atom, the particle in a box, the free particle and the harmonic oscillator. In all cases we will give a detailed description of the maximum entropy wave function, entropies etc and also compare our theory to what we get when we use the standard quantum statistical formalism as formulated by Von Neumann[9].

#### 4. THE TWO-LEVEL ATOM

The two-level atom is a quantum system with a two-dimensional state space whose energy operator is given by

$$H = H_0 + H_1 = \begin{bmatrix} E_1 & 0 \\ 0 & E_2 \end{bmatrix} + \varepsilon \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix},$$

where  $E_1$  and  $E_2$  are the energies of the ground state and the excited state and where  $\varepsilon$  is the coupling between the two levels. The Von Neumann quantum statistical formalism[9] describe the system in terms of a density operator that for the two-level system is a positive 2 by 2 matrix with trace equal to one. If the expectation of the energy is the only known quantity then the following density operator is used

$$\rho = \frac{1}{Z} e^{-H/T},$$

where  $Z(T) = \text{Tr}(e^{-H/T})$  is the partition function. It is well known that this operator is also determined as the solution of a constrained extremal problem. In fact for this case one defines the entropy as

$$S(\rho) = -\text{Tr}(\rho \ln \rho)$$

and finds the extremum of  $S$  under the constraint  $\langle H \rangle = E$ . The formula for  $S$  is inspired by the corresponding classical case and is explained in

any textbook on quantum mechanics. We will have something to say about this later.

For the two-level system the partition function is

$$Z(T) = 2e^{-\frac{E_1+E_2}{2T}} \text{Cosh} \left( \frac{\sqrt{4\varepsilon^2 + (E_1 - E_2)^2}}{2T} \right)$$

The energy as a function of  $T$  is given by

$$E = \langle H \rangle = \frac{T^2}{Z} \partial_T Z$$

If  $E_+$  and  $E_-$  are the eigenvalues of the energy operator  $H$  we evidently have

$$\begin{aligned} E &\approx \frac{E_1 + E_2}{2} & T &\longrightarrow \pm\infty \\ E &\approx E_- & T &\gtrsim 0 \\ E &\approx E_+ & T &\lesssim 0 \end{aligned}$$

Figure 1 show the energy as a function of the temperature for the two-level atom as predicted by the von Neumann method.

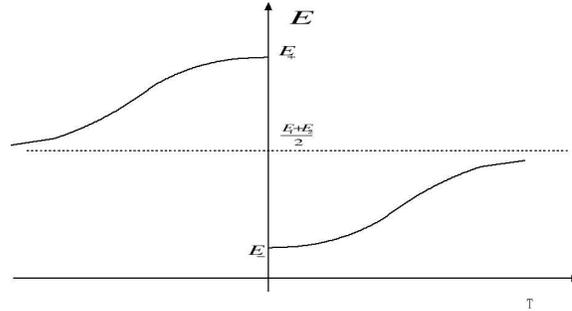


FIGURE 1

Negative temperatures thus corresponds to a inverted system. We will now find the maximum entropy wave function for this system and compare with what we got using the Von Neumann method.

The first step in our method is to select a space of events, or in other words a phasespace. There is of course no unique way of doing this, it all depends on what your prior information consists of. Let us assume that we know that the system is equally likely to be found in the upper and lower energy eigenstate of the unperturbed Hamiltonian  $H_0$ . This means that our space of events  $\Omega$  consists of two points corresponding to

the two elements in the eigenbasis for  $H_0$  and the prior distribution can be taken to be

$$\rho = \left(\frac{1}{2}, \frac{1}{2}\right)$$

The maximum entropy wave function will in this case be a solution to the following equation.

$$H\varphi + T\varphi \ln \varphi \bar{\varphi} = h\varphi,$$

where  $T$  and  $h$  are the Lagrange multipliers. Note that we have redefined the Lagrange multipliers in order to absorb the effect of the prior  $\rho$ , this can not be done if the prior information give a bias to one or the other of the two eigenstates for  $H_0$ . Define the function  $\theta = (w_1, w_2)$  through  $\varphi = \frac{1}{\sqrt{Z}}\theta$ . The normalization condition is satisfied if we choose  $Z = \int \theta \bar{\theta} dx = w_1 \bar{w}_1 + w_2 \bar{w}_2$  and the choice  $h = \ln Z$  gives the following simplified equation for the function  $\theta$ .

$$H\theta + T\theta \ln \theta \bar{\theta} = 0$$

We can without loss of generality cancel a common phase for the function  $\theta$  and write

$$\theta = (a, be^{iv}),$$

where  $a, b$  and  $v$  are real numbers. We now must solve the following nonlinear matrix equation

$$\begin{bmatrix} E_1 & \varepsilon \\ \varepsilon & E_2 \end{bmatrix} \begin{bmatrix} a \\ be^{iv} \end{bmatrix} + T \begin{bmatrix} a \ln a^2 \\ b \ln b^2 \end{bmatrix} = 0$$

By taking real and imaginary parts we find that

$$\begin{aligned} E_1 a + s\varepsilon b + T a \ln a^2 &= 0 \\ E_2 b + s\varepsilon a + T b \ln b^2 &= 0 \end{aligned}$$

where  $s = \pm 1$ . The corresponding wave function is  $\varphi = \frac{1}{\sqrt{Z}}(a, sb)$  and  $s = \pm 1$  corresponds to a phase difference of 0 or  $\pi$  between the two components of the maximum entropy wave function. Note that we can now without loss of generality assume that the coupling  $\varepsilon$  is a positive

number. By manipulating the equation we find that

$$\begin{aligned} a &= \text{Exp}\left(-\frac{E_1 + s\varepsilon x}{2T}\right) \\ b &= x \text{Exp}\left(-\frac{E_1 + s\varepsilon x}{2T}\right) \\ Z &= (1 + x^2) \text{Exp}\left(-\frac{E_1 + s\varepsilon x}{2T}\right) \\ \varphi &= \frac{1}{\sqrt{1 + x^2}}(1, sx) \end{aligned}$$

where  $x$  is a solution to the equation

$$\frac{E_2 - E_1}{2} + \frac{s\varepsilon}{2}\left(\frac{1}{x} - x\right) + T \ln x = 0$$

This transcendental equation can not be solved exactly but a combination of numerical and asymptotic methods give the complete picture of the solutions as a function of the parameter  $T$  as illustrated in figure 2.

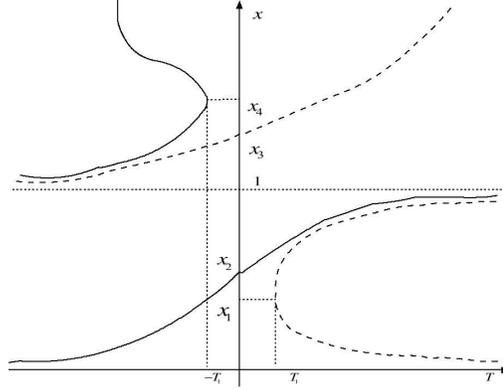


FIGURE 2.

where

$$\begin{aligned} x_1 &= -\frac{T_1}{\varepsilon} + \sqrt{\left(\frac{T_1}{\varepsilon}\right)^2 - 1} \\ x_2 &= -\left(\frac{E_2 - E_1}{2\varepsilon}\right) + \sqrt{\left(\frac{E_2 - E_1}{2\varepsilon}\right)^2 + 1} \\ x_3 &= \left(\frac{E_2 - E_1}{2\varepsilon}\right) + \sqrt{\left(\frac{E_2 - E_1}{2\varepsilon}\right)^2 + 1} \\ x_4 &= \frac{T_1}{\varepsilon} + \sqrt{\left(\frac{T_1}{\varepsilon}\right)^2 - 1} \end{aligned}$$

and where  $T_1$  is determined by the condition  $x_1 x_4 = 1$ . The dotted curve in the figure corresponds to  $s = 1$  and the solid curve to the case  $s = -1$ .

All the other quantities in the theory can now be computed from  $x$ . Figure 3 gives the energy as a function of  $T$  for the weak interaction case. This means that the interaction energy  $\varepsilon$  is less than half the energy difference between the two levels. This is of course the usual situation. A similar figure exists for the strong interaction case, but we will not discuss it here.

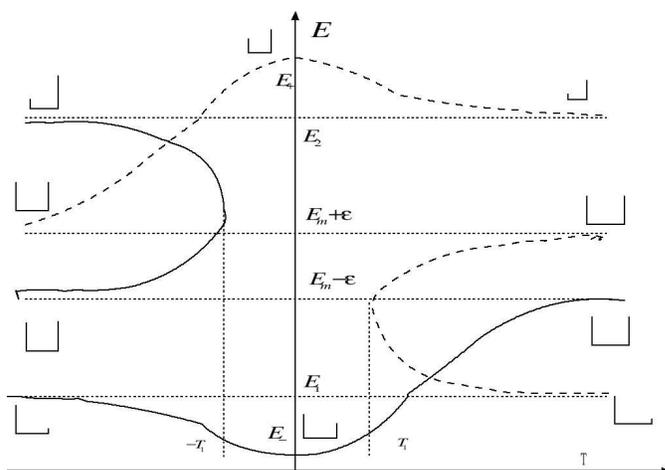


FIGURE 3.

For each value of the parameter  $T$  there are up to four different values of the energy corresponding to equally many different wave functions. The small drawings indicate the probabilities  $p_1$  and  $p_2$  of finding the system in the two eigenstates of  $H_0$  for the wave function corresponding to that point in the diagram. The lowest energy eigenstate is to the left in each drawing. The dotted curves and the solid curves corresponds to the two cases  $s = 1$  and  $s = -1$  as in the previous figure. Since the entropy is  $S = p_1 \ln p_1 + p_2 \ln p_2$  each eigenstate has zero entropy and the uniform case  $p_1 = p_2 = \frac{1}{2}$  has maximum entropy. Using this it is easy to read off how the entropy vary along each branch in the figure. If we for example consider the lower solid curve, then the entropy start off at zero for  $T = -\infty$ , increase monotonically and reach its maximum at  $T = \infty$ .

In order to find the maximum entropy wave function corresponding to a given value of the energy  $E = \langle H \rangle$  we draw a horizontal line and mark the intersection of this line with the curves in the figure. There is always more than one intersection but only one of these corresponds to a wave function of maximal entropy. Thus for each value of  $E$  there is a unique

$T$  and a wave function depending on  $T$  that is a maximum entropy wave function for the given value of the energy.

Using the figure and numerical calculations we find that

- (1) For  $E_- < E < E_m - \varepsilon$ , points on the lower solid curve for  $T > 0$  corresponds to the maximum entropy wave function. Thus for these energies this curve gives the energy,  $E(T)$ , as a function of  $T$  and  $T$  is positive and run from zero to infinity. If we denote the corresponding curve found by using the density matrix by  $E_D(T)$  we have

$$\begin{aligned} E(T) &\approx E_D(T) \text{ for } T \gtrsim 0 \\ E(T) &\lesssim \frac{E_1 + E_2}{2} - \varepsilon \quad T \longrightarrow \infty \\ E_D(T) &\lesssim \frac{E_1 + E_2}{2} \quad T \longrightarrow \infty \end{aligned}$$

For  $T$  close to zero the two curves are indistinguishable but for large  $T$  they asymptotes to different energies. The difference is determined by the size of the coupling between the two levels. If this is small, then they are close also for large  $T$ .

- (2) For  $E_m + \varepsilon < E < E_+$ , points on the upper dotted curve for  $T < 0$  corresponds to the maximum entropy wave function. We now find

$$\begin{aligned} E(T) &\approx E_D(T) \text{ for } T \lesssim 0 \\ E(T) &\lesssim \frac{E_1 + E_2}{2} + \varepsilon \quad T \longrightarrow -\infty \\ E_D(T) &\lesssim \frac{E_1 + E_2}{2} \quad T \longrightarrow -\infty \end{aligned}$$

As in the previous case our method and the density matrix method gives results that are indistinguishable for  $T$  close to zero and that differ by an amount given by the interaction energy when  $T$  goes to negative infinity.

- (3) When the energy cross the value  $E_m + \varepsilon$ , points on the solid upper curve corresponds to the maximum entropy wave function. The parameter  $T$  start out large negative and then increase as we increase the energy past  $E_m + \varepsilon$ . At a critical energy between  $E_m - \varepsilon$  and  $E_m + \varepsilon$  the point representing the maximum entropy wave function jumps to the lower dotted curve and the corresponding  $T$  jumps from a negative to a positive value. As  $E$  increase from the critical energy towards  $E_m + \varepsilon$  the point representing the

maximum entropy wave function follows the lower dotted curve and the parameter  $T$  increase towards positive infinity.

The behavior described under point 3 above has no counterpart in the density matrix approach ,but for  $\varepsilon$  small it all occurs in a very narrow range of energies.

## 5. THE PARTICLE IN A BOX

We will in this section find the maximum entropy wave function for a particle in a box and compare with what we get using the usual approach to quantum statistics. For this problem we will let our phase space be the real line between 0 and 1. We will consider a single observable given by the operator

$$H = -\partial_{xx}.$$

This operator represents the energy for a particle in a box if we use units where Planck's constant is 1 etc. In order to set up and solve the equation for the maximum entropy wave function we must assign a prior wavefunction. We will in this example assume that our prior information is that the particle can be anywhere in the box with uniform probability. A prior distribution that represents this information is chosen to be

$$\rho = 1 \quad 0 \leq x \leq 1.$$

The maximum entropy wave function is now a solution to the following nonlinear boundary value problem

$$\begin{aligned} -\partial_{xx}\varphi + T\varphi \ln \varphi \bar{\varphi} &= h\varphi, \\ \varphi(0) = \varphi(1) &= 0 \end{aligned}$$

The Lagrange multipliers  $h$  and  $T$  are determined by the constraints

$$\begin{aligned} E &= \langle \varphi, H\varphi \rangle \\ 1 &= \langle \varphi, \varphi \rangle \end{aligned}$$

Note that since the entropy is given by the formula

$$S = - \int \varphi \bar{\varphi} \ln \varphi \bar{\varphi} dx$$

it is clear that if we can find a real wave function that satisfy the constrained maximum problem then no complex solution can have higher entropy and the real solution is a maximum entropy wavefunction. We will in the following assume that  $\varphi$  is real.

If we multiply the differential equation by  $\varphi_x$  and integrate we find the following equation

$$-(\varphi_x)^2 + T(\varphi^2 \ln \varphi^2 - \varphi^2) - h\varphi^2 + c = 0,$$

where  $c$  is a constant of integration. We can simplify this equation by introducing a new function  $\theta$  through the identity

$$\varphi(x) = \text{Exp}\left(\frac{T+h}{2T}\right)\theta\left(\sqrt{|T|x}\right).$$

Then the function  $\theta$  is found to satisfy the following simplified boundary value problem

$$\begin{aligned}(\theta_y)^2 &= a + s\theta^2 \ln \theta^2 \\ \theta(0) &= \theta(\sqrt{|T|}) = 0\end{aligned}$$

where  $s = \text{sgn}(T)$  and  $a$  is a new constant. The solutions of the boundary value problem is quite different depending on whether  $s$  is 1 or  $-1$ . We will first discuss the case  $s = -1$  corresponding to negative values for the Lagrange multiplier  $T$ .

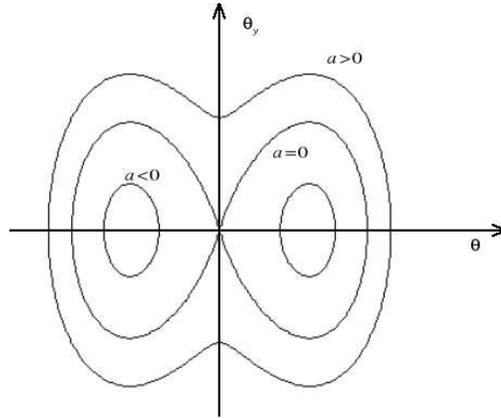


FIGURE 4

Figure 4 is a drawing of the integral curves of the differential equation when  $s = -1$ . Note that the integral curves for  $a < 0$  does not cross the  $\theta_y$  axis. This means that the nonlinear boundary problem does not have a solution for these values of the constant. The value  $a = 0$  corresponds to the separatrix and  $a > 0$  corresponds to integral curves outside the separatrix. All these curves cross the  $\theta_y$  axis. The ground state of the nonlinear boundary value problem corresponds to the situation where the integral curve starts on the  $\theta_y$  axis with  $\theta(0) = 0$  and have completed half a period when it cross the  $\theta$  axis. In order to find which value of  $a$

that realize this situation for a given value of  $T < 0$  we must solve the following integral equation

$$\int_0^{\bar{\theta}} \frac{d\theta}{\sqrt{a - \theta^2 \ln \theta^2}} = \frac{1}{2} \sqrt{|T|},$$

where  $\bar{\theta}$  is a solution of the equation

$$a - \bar{\theta}^2 \ln \bar{\theta}^2 = 0$$

Using the Lambert W function  $\bar{\theta}$  can be expressed in closed form as

$$\bar{\theta} = \sqrt{\frac{a}{W(a)}}$$

For a given  $T < 0$ , solution of the nonlinear integral equation gives us  $a(T)$  and the function  $\theta$  corresponding to this value of  $a$  is found by integrating the differential equation. Using these,  $\varphi$  and all quantities of interest can be found

$$\begin{aligned} \varphi(x) &= \left( \frac{1}{\sqrt{|T|}} \int_0^{\sqrt{|T|}} \theta(y) dy \right)^{-\frac{1}{2}} \theta(\sqrt{|T|x}) \\ E(T) &= \int_0^1 (\varphi_x)^2 dx \\ S &= \int_0^1 \varphi^2 \ln \varphi^2 dx \end{aligned}$$

In general the integral equation must be solved numerically but the separatrix corresponding to  $a = 0$  can easily be found explicitly and is a Gaussian. It has infinite period and thus can not be a solution of our boundary value problem but we expect that as  $T \rightarrow -\infty$  we will have  $a \rightarrow 0$  and the solution should look more and more like a Gaussian. Our numerical calculations verifies that this is the case. In figure 5 the solution and the Gaussian are compared for  $T = -10$  and  $T = -20$ . The Gaussian is the dotted curve. We see that the solution for large negative  $T$  is converging to the Gaussian.

If we use this observation and substitute the separatrix solution into the expression for  $E(T)$ , now taking the integral from  $-\infty$  to  $\infty$ , we get the following expression for the value of  $E(T)$  for large negative values

$$E(T) \sim \frac{1}{2} |T|.$$

For  $T \lesssim 0$  the nonlinear eigensolutions converge to the quantum mechanical ground state for the particle in a box and the energy approaches the groundstate energy  $E_1 = \pi^2$  from above.

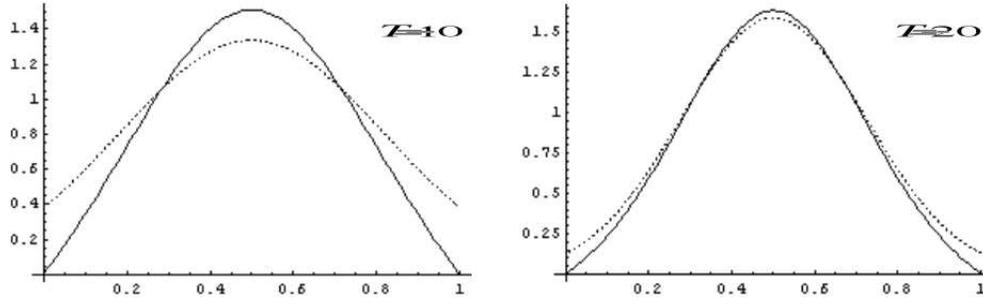


FIGURE 5

So far we have discussed the ground state of the nonlinear eigenvalue problem. In general for each integer  $p$  there is a solution that starts at the  $\theta_y$  axis with  $\theta(0) = 0$  and that complete  $\frac{1}{2p}$  of a period at the first intersection with the  $\theta$  axis. For this solution the corresponding  $a(T)$  and solution  $\theta$  is determined by

$$\int_0^{\bar{\theta}} \frac{d\theta}{\sqrt{a - \theta^2 \ln \theta^2}} = \frac{1}{2p} \sqrt{|T|},$$

where  $\bar{\theta}$  is as in the formula for the ground state. We can absorb this equation into the equation determining the ground state if we let  $T \rightarrow \frac{T}{p^2}$ . For large negative  $T$  the nonlinear eigensolution approach more and more closely to a sequence of Gaussian where the number of Gaussian is equal to  $p$ . Figure 6 compare the sequence of Gaussian and the nonlinear eigensolution for  $p = 2$ . The dotted line is the Gaussian and the full line is the nonlinear eigensolution.

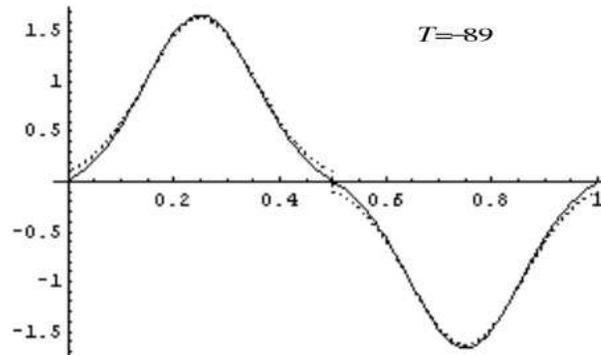


FIGURE 6

As for the ground state this correspondence between Gaussian and eigensolutions gives us a asymptotic estimate for the energy as a function

of  $T$  for large negative  $T$  that turns out to be exactly the same as for the groundstate. For  $T \lesssim 0$  the  $p$ 'th nonlinear eigenstate converge to the  $p$ 'th excited state for the particle in a box  $E_p = p^2\pi^2$ . This completes our description of the eigensolutions for the case of negative  $T$ .

We now turn to the case when  $T > 0$ . Figure 7 show the phase portrait for this case.

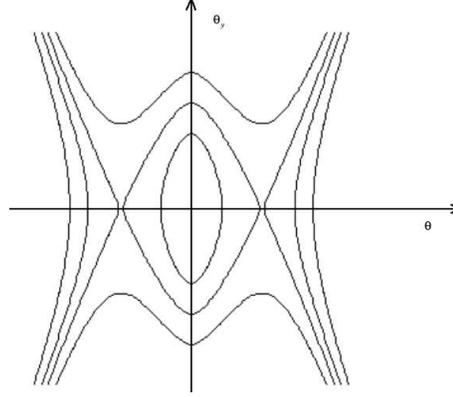


FIGURE 7

Only for  $a$  in the range  $0 < a < e^{-1}$  will there be closed integral curves that intersect the  $\theta_y$  axis. For each  $a$  in this range the integral curves actually consists of three disjoint components but only one of them is closed. Since only closed integral curves can correspond to solutions that satisfy the boundary conditions, only this range of  $a$ 's is relevant. The closed components of the integral curves are created as a point at the origin for  $a = 0$ , expands for increasing  $a$  and reach an separatrix of infinite period for  $a = e^{-1}$ . In order to find the value of  $a$  and nonlinear eigensolution  $\theta$  corresponding to a given value of  $T$  we must solve the following equation

$$\int_0^{\bar{\theta}} \frac{d\theta}{\sqrt{a + \theta^2 \ln \theta^2}} = \frac{1}{2p} \sqrt{T},$$

where now

$$\bar{\theta} = \sqrt{\frac{-a}{W_{-1}(-a)}},$$

where  $W_{-1}$  is a branch of the Lambert function and where the integer  $p$  enumerate the nonlinear eigensolutions as previously. For this case the equation for the separatrix can not be solved in closed form. However if we expand the equation close to the point  $\theta = e^{-1}$  writing  $\theta = e^{-1} + \varepsilon$  we

get the following equation for  $\varepsilon$

$$(\varepsilon_y)^2 = 2\varepsilon^2$$

and this equation has solutions  $\varepsilon \sim e^{-\sqrt{2}y}$ . We postulate that this formula gives the shape of the eigensolutions over the whole interval, also close to the endpoints. Since large  $T$  corresponds to  $a$  close to  $e^{-1}$  we thus postulate that the following formula gives a approximation to the eigensolutions for large  $T$

$$\varphi(x) = \alpha(T) \begin{cases} 1 - e^{-\sqrt{2T}x} & 0 \leq x \leq \frac{1}{2} \\ 1 - e^{-\sqrt{2T}(1-x)} & \frac{1}{2} \leq x \leq 1 \end{cases}$$

The function  $\alpha(T)$  is determined by normalization. For large  $T$  we find that  $\alpha(T) \sim 1$ . Numerical calculations show that the formula is a very good approximation even for  $T$  as small as 5.

Similar formulas applies for  $p > 1$ . As  $T$  become large the nonlinear eigenstate corresponding to  $p = 1$  becomes flat over a larger and larger part of the interval and vary very quickly near  $x = 0$  and  $x = 1$  in order to satisfy the boundary conditions. The nonlinear eigenstates corresponding to higher values of  $p$  consists of essentially constant pieces joined together by very short intervals where the functions vary very quickly. In figure 8 we compare the approximate and exact solutions for  $p = 1$ . The approximate formula is the dotted curve.

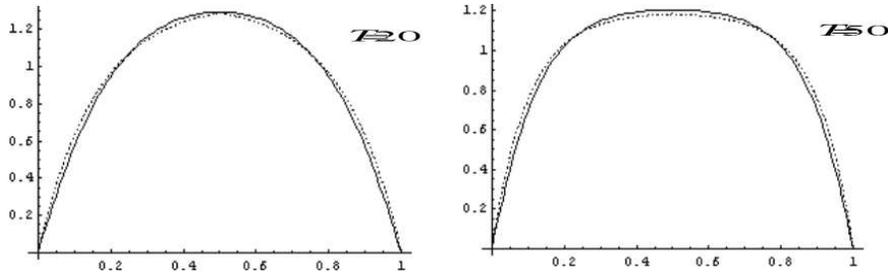


FIGURE 8.

Using these formulas we can now compute expressions for the energy and entropy for large values of  $T$ .

$$E_p(T) \sim p\sqrt{2T}$$

$$S_p(T) \sim -\frac{ap}{\sqrt{2T}},$$

where  $a$  is a positive numerical constant. These formulas show that for fixed large  $T$  the nonlinear eigenstate corresponding to  $p = 1$  has the

highest entropy. Numerical calculations show that this is generally true. Figure 9 show the energy  $E_1(T)$ , as a function of  $T$  for the ground state  $p = 1$ . The small pictures show the shape of the nonlinear eigenstate at the corresponding point in the diagram.

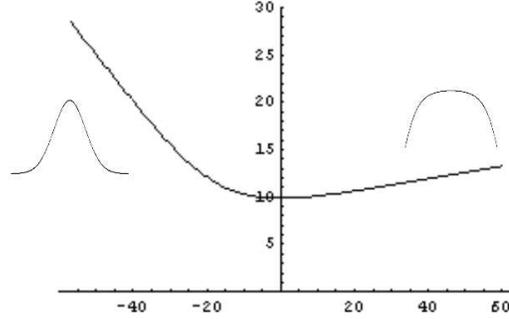


FIGURE 9

For any given value of the energy,  $E$ , above the ground state energy  $\pi^2$ , there are finitely many nonlinear eigenstates that has  $E$  as expectation value for the energy operator but the one on the  $p = 1$  branch for positive  $T$  always have maximal entropy. The maximal entropy wave functions thus always has a general shape as in figure 8. This is actually intuitively obvious, since for a given energy these are the eigensolutions that are least localized. They are exactly as localized as they need to be in order to reproduce the measured mean. For a given energy the point on  $E_1(T)$  for negative  $T$  corresponds to the wave function that has minimum entropy while satisfying the constraints. These are the minimum entropy wave functions. Note that as  $T$  goes to negative infinity the corresponding minimum entropy wave function goes to a delta function located at  $x = \frac{1}{2}$ . In this limit we recover a classical particle with a definitive location.

We will now compare with what we get by using the usual density operator approach to this problem. The ensemble is then described by the operator

$$\rho = \frac{1}{Z(T)} e^{-\frac{H}{T}}$$

and the energy as a function of  $T$  is given by

$$E(T) = Tr(H\rho)$$

If we expand in an eigenbasis for  $H = -\partial_{xx}$  we get the expression

$$E(T) = \frac{\sum_{i=1}^{\infty} n^2 \text{Exp}\left(-\frac{\pi^2 n^2}{T}\right)}{\sum_{i=1}^{\infty} \text{Exp}\left(-\frac{\pi^2 n^2}{T}\right)}$$

These formulas only makes sense for positive values of  $T$ . Note that in this context the parameter  $T$  is interpreted as the physical temperature (Boltzman's constant equal to 1). Figure 10 show the function  $E(T)$ .

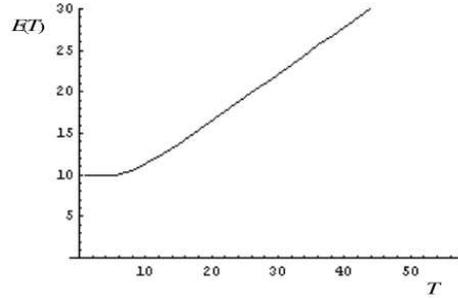


FIGURE 10

Note that the graph of  $E(T)$  for positive  $T$  is quite similar to the graph of  $E_1(T)$  for negative  $T$ . It is in fact easy to show that  $E(T)$  for large positive  $T$  and  $E_1(T)$  for large negative  $T$  both grow at the same rate

$$E(T) \sim \frac{1}{2}T \quad T \gg 1$$

$$E_1(T) \sim \frac{1}{2}T \quad T \ll -1$$

This correspondence actually extends to higher moments also. Figure 11 show the standard deviation

$$\sigma(T) = \sqrt{\langle H^2 \rangle - \langle H \rangle^2}$$

as a function of  $T$  using the minimum entropy wave function and the density operator. In order to compare them more easily, we have flipped the sign on  $T$  for the first of these in order to have both on the same interval. The dotted curve is the standard deviation computed on the basis of the density operator.

They appear to stay close to each other and increase at the same rate. Asymptotics show that they do in fact grow at the same rate and this common rate is

$$\frac{1}{\sqrt{2}}T$$

Thus for the case of a particle in a box we have a close correspondence between the minimum entropy wave function and quantum statistics based on the von Neumann entropy and the canonical ensemble. For the particle in a box this is important because it implies that, from a maximum entropy point of view, using the canonical ensemble is the worst possible

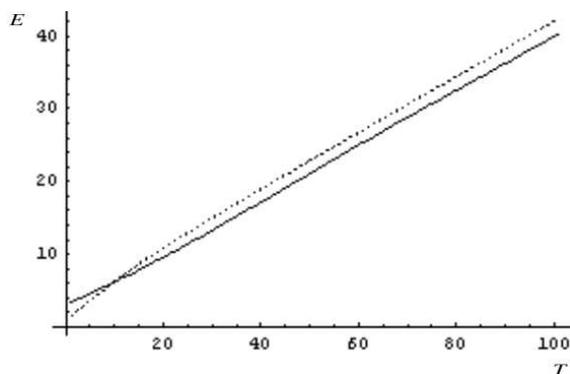


FIGURE 11

thing we can do. Of course we know that the use of the canonical ensemble is very well established in quantum statistics and that it will take a lot to change this practice. Our calculations in this section show however that may be it is time to ask some pointed questions about the canonical ensemble and the von Neumann entropy from which it is derived.

Recall that the von Neumann entropy is given by the formula

$$S = -Tr(\rho \ln \rho)$$

In all applications of the maximum entropy principle in classical probability and statistics the prior plays a pivotal role. Where is the prior in the von Neumann entropy?

The connection between the wave function and probability was first noted by Max Born. The probability  $P(U)$  of finding a particle in a domain  $U$  is given by

$$P(U) = \int_U \varphi \bar{\varphi} dx$$

This formula leads to the interpretation of  $\rho = \varphi \bar{\varphi}$  as a probability density. From this, introducing the density operator and writing down the canonical ensemble is but a short step. Note that in quantum mechanics the relation between the wave function and probability theory appears accidental, something that could have been different. It is interesting in this connection that probability theory has been generalized[10] in a very natural way to operator valued probabilities and in this extended theory the connection between probabilities and a natural generalization of the wave function is not accidental but necessary. Furthermore in this generalized theory probability densities do not exist, only wave functions exists. Thus a notion of entropy, if any exists at all in this generalized context, must be formulated in terms of wave functions, not densities. In

fact our decision to use the entropy  $S(\varphi, \bar{\varphi})$  was in part inspired by this observation.

In our formalism there appears to be a interesting connection between the prior and the potential. If we assume that the prior is specified in physical space then if the energy is our only observable, our fundamental equation for the maximal entropy wave function is

$$H\varphi + T\varphi \ln \frac{\varphi \bar{\varphi}}{\rho} = h\varphi$$

For quantum analogs of classical mechanical systems,  $H = K + V$  where  $V$  is the potential energy and  $K$  is the kinetic energy operator. Note that the fundamental equation can be written as

$$H'\varphi + T\varphi \ln \varphi \bar{\varphi} = h\varphi,$$

where the operator  $H'$  is given by

$$H' = K + V - T \ln \rho$$

Thus the prior appear as an extra term in the classical potential energy. The extra term will have a different effect depending on the sign of the parameter  $T$ . If  $T$  is positive then the extra term will give a repulsive force in the classical domain whereas for negative  $T$  the force will be attractive. Thus for positive  $T$  the term should be expected to have a dispersive effect leading to a broadening of probability distributions whereas in the opposite case it will have a localizing effect on the distributions. This is what we observed for the case investigated in this section.

## 6. THE FREE PARTICLE

The Hamiltonian for a free particle on the real line is

$$H = -\partial_{xx}$$

The Hamiltonian for a free particle has a continuous spectrum with formal eigenfunctions  $\varphi_k = \frac{1}{\sqrt{2\pi}}e^{ikx}$  where

$$H\varphi_k = k^2\varphi_k$$

The partition function is

$$Z = \int_{-\infty}^{\infty} \langle \varphi_k, e^{-H/T} \varphi_k \rangle dk = \sqrt{\pi T}$$

The energy as a function of  $T > 0$  is

$$E = \langle H \rangle = Tr(H\rho) = \frac{1}{2}T$$

In order to apply our formalism to this case we have to specify a prior distribution  $\rho$  on the line. We will use that the choice  $\rho = 1$  gives a very good correspondence between our theory and the Von Neumann approach. From a strict probabilistic point of view this prior does of course not make sense since it is not normalizable, but our theory still gives sensible answers. In fact the choice of a prior representing total ignorance about a unbounded random variable is a difficult, interesting and unsolved problem in statistics. This problem is evaded in various ways in classical and quantum statistics but it is not solved. In quantum statistics the prior, if it exists at all, is embedded in the definition of the trace and from the elementary theory of traces it is quite evident that this prior is in fact uniform.

Thus we will proceed with  $\rho = 1$  in our theory. The equation we must solve is

$$-\partial_{xx}\varphi + T\varphi \ln \varphi \bar{\varphi} = h\varphi$$

This is the same equation as in the previous section but now we need solutions that are normalizable on the whole real line. From the phase plots for positive and negative values of  $T$  and the discussion there it is evident that solutions normalizable on the real line only exist for negative values of  $T$  and in fact correspond exactly to the separatrix. The maximum entropy wave function is found by integrating the equation for the separatrix as in the previous section.

$$\varphi(x) = \sqrt[4]{\frac{|T|}{\pi}} e^{-\frac{1}{2}|T|x^2}$$

The energy can now easily be found

$$E = \langle H \rangle = \langle \varphi, H\varphi \rangle = \frac{1}{2}|T|$$

This is exactly the same as we found for the von Neumann method if we let  $T \rightarrow -T$ . In fact it is simple to show by direct calculations that all moments  $\langle H^n \rangle$  are the same for the von Neumann method and our maximum entropy wave functions.

## 7. THE HARMONIC OSCILLATOR

The Hamiltonian for the Harmonic oscillator on the line can be written as

$$H = -\frac{1}{2}\partial_{xx} + \frac{1}{2}\varepsilon x^2,$$

where  $\varepsilon > 0$  is a measure of the width of the potential. The von Neumann method gives the following expression for the energy as a function of

$T > 0$

$$E(T) = \frac{\sqrt{\varepsilon}}{2} \coth\left(\frac{\sqrt{\varepsilon}}{2T}\right)$$

The maximum entropy wave function for the Harmonic Oscillator is a solution of

$$-\frac{1}{2}\partial_{xx}\varphi + \frac{1}{2}\varepsilon x^2\varphi + T\varphi \ln \varphi \bar{\varphi} = h\varphi,$$

where we assume a uniform prior with the same caveat as before. As an interesting aside note that for  $T > 0$  this equation is the same as the equation for a free particle with a Gaussian prior

$$\rho = \sqrt{\frac{a}{\pi}} e^{-ax^2}$$

if we use  $\varepsilon = 2Ta$ .

We will in this paper not find the general solution to the above equation but limit ourselves to finding a special closed form solution. It is well known that the logarithmic Schrödinger equation with a harmonic potential has an exact Gaussian solution. This solution is

$$\varphi(x) = \sqrt{\frac{b(T)}{\pi}} e^{-\frac{1}{2}b(T)x^2},$$

where

$$b(T) = \sqrt{T^2 + \varepsilon} - T$$

From this is now a simple matter to show that the energy as a function of  $T$  is

$$E(T) = \frac{1}{2}\sqrt{T^2 + \varepsilon}$$

For  $\varepsilon \rightarrow 0$  the harmonic oscillator potential goes away and the solution approaches the maximum entropy wave function for the free case for  $T < 0$  and the energy becomes  $E(T) = \frac{1}{2}|T|$  as for the free case. For  $T > 0$  the solution approaches a uniform non-normalizable function. The entropy for the Gaussian solution is

$$S(T) = \frac{1}{2} - \frac{1}{2} \ln \frac{b(T)}{\pi}$$

The energy as a function of  $T$  is symmetric in  $T$  so that for a fixed value of  $E$  there are two values of  $T$

$$T = \pm\sqrt{2E - \varepsilon}$$

and thus two possible Gaussian solutions. These are different because  $b(T)$  is not symmetric in  $T$ . The expression for the entropy shows that the one corresponding to positive  $T$  has highest entropy of the two. We do not know if this solution has maximum entropy among all solutions of

the equation but we strongly suspect this because the solution we have found connects to the ground state of the Harmonic Oscillator when  $T \rightarrow 0$  and we have seen that the branch emanating from the ground state is in fact the maximum entropy solution for the particle in a box. For small values of  $T$  our energy function and the one derived using the von Neumann method agrees but for large  $T$  they deviate. Our grow as  $\sim \frac{1}{2}|T|$  whereas the one derived using the von Neumann formula grows as  $\sim T$ .

This means that for a given value of the energy our method predicts a higher value of  $T$  and thereby a larger value of the dispersion than the von Neumann method. One could say that our Harmonic Oscillator is "hotter" than the usual one.

However note that our result depends on an explicit choice of prior information whereas the prior behind the von Neumann is hidden. As noted before the elementary theory of trace indicate strongly that the prior behind the von Neumann method is the uniform distribution on the index space of whatever basis chosen to evaluate the trace. In the current example this means that the von Neumann prior is a uniform prior in the discrete uncountable space of energies for the Harmonic oscillator. If we use this prior, our method produce exactly the same result as the von Neumann method in all cases. The strength of our method is that the choice of prior information is explicit and enforced.

#### REFERENCES

- [1] "Probability, Frequency and Reasonable Expectation", American Journal of Physics, 14, pages 1-13, 1946.
- [2] "The algebra of probable inference", Johns Hopkins University Press, 1961.
- [3] "Theory of Probability", H. Jeffreys, Oxford University Press, 1961.
- [4] "Probability Theory: The Logic of Science", E. T. Jaynes, Cambridge University press, 2003.
- [5] "Information Theory and Statistical Mechanics", E. T. Jaynes, The Physical Review, Vol 106, no.4, pages 620-630, May 15, 1957.
- [6] Bialynicki-Birula and Mycielski, Ann. Phys. (N. Y. ) 100, 62, 1976.
- [7] "Weinberg's nonlinear quantum mechanics and the Einstein-Poldolsky-Rosen Paradox", Joseph Polchinski, Phys. Rev. Lett. 66, no 4, 28 Januar 1991.
- [8] "Precision tests of Quantum Mechanics", S. Weinberg, Phys. Rev. Lett. 62, no. 5, 30 January, 1989.
- [9] "Mathematical Foundation of Quantum Mechanics", J. von Neuman, Princeton University Press, 1955.
- [10] "Operator valued probability theory", Per Jakobsen and Valentin Lychagin, Lobachevskii Journal of Mathematics, 16, pages 17-56, 2004.

UNIVERSITY OF TROMSO, 9020 TROMSO, NORWAY

*E-mail address:* Per.Jakobsen@matnat.uit.no

*E-mail address:* Valentin.Lychagin@matnat.uit.no

Received June 28, 2006