On the use of non-canonical quantum statistics

2000 J. Phys. A: Math. Gen. 33 7067

(http://iopscience.iop.org/0305-4470/33/40/305)

J. Phys. A: Math. Gen. **33** (2000) 7067–7091. Printed in the UK PII: S0305-4470(00)09427-0

On the use of non-canonical quantum statistics

Jani Lukkarinen

Helsinki Institute of Physics, PO Box 9, 00014 University of Helsinki, Finland

E-mail: jani.lukkarinen@helsinki.fi

Received 4 November 1999, in final form 26 May 2000

Abstract. We develop a method using a coarse graining of the energy fluctuations of an equilibrium quantum system which produces simple parametrizations for the behaviour of the system. As an application, we use these methods to gain more understanding on the recently developed Tsallis statistics and we show how energy reparametrizations can be used to produce other similar generalizations of the Boltzmann–Gibbs statistics. We conclude with a brief discussion on the role of entropy and the maximum entropy principle in thermodynamics and in quantum statistics.

1. Introduction

The usefulness of the canonical ensemble in statistical mechanics is remarkable. The standard explanation for this success relies on taking the thermodynamical limit which corresponds to increasing the volume of the system to infinity while keeping all the relevant intensive quantities, i.e. densities, fixed and finite. From this point of view, the canonical ensemble should not have much utility for small systems consisting of only a few particles. This is not completely true and canonical expectation values can also be used for small systems [1], although not very accurately in the direct manner used in thermodynamics.

We shall try to generalize the ideas leading to standard thermodynamics so that these methods can be applied in the analysis of small systems. We shall first concentrate on finding practical approximations which capture the large-scale behaviour of an equilibrium system. For this purpose, we present in sections 3 and 4 a coarse-graining procedure which justifies a modified Gaussian ensemble and we derive a positive saddle-point (PSP) approximation which can be applied for further simplifying the Gaussian results.

As an application of these methods, in section 5 we present how the PSP approximation leads to a generalization of thermodynamics for essentially isolated systems. As a second application, we derive the recently proposed non-extensive Tsallis statistics [2] as a PSP approximation. We conclude with a discussion of the implications for Tsallis thermodynamics and the maximum entropy principle in the last two sections, 8 and 9.

2. Preliminaries in quantum statistics

The standard approach to quantum statistics [3] is defined by using a density matrix $\widehat{\rho}$, which is a non-negative, Hermitian, trace-class operator normalized to one and which gives the

0305-4470/00/407067+25$30.00

expectation value of an observable $\widehat{A}$ by the formula $\langle\widehat{A}\rangle = \mathrm{Tr}(\widehat{A}\widehat{\rho})$. In some complete eigenbasis $|\psi_i\rangle$, the density matrix can thus be expanded as

$$\widehat{\rho} = \sum_i p_i |\psi_i\rangle\langle\psi_i| \tag{1}$$

where the eigenvalues p_i are non-negative, $p_i \geqslant 0$, and they satisfy the normalization condition $\sum_i p_i = 1$.

Suppose next that the system has a discrete energy spectrum with finite degeneracies, which in quantum mechanics is achieved for every sufficiently binding potential. An equilibrium, i.e. time-independent, ensemble is then given by a density matrix which has time-independent eigenvalues and which satisfies the Liouville–von Neumann equation $[\widehat{H}, \widehat{\rho}] = 0$. In this case, the eigenvectors ψ_i can be chosen so that they are also energy eigenvectors; let E_i denote the eigenvalue corresponding to ψ_i.

Assume also that energy is the only relevant conserved quantity in this equilibrium system. By this we mean that the probability of finding any energy eigenstate depends only on energy and not on any other conserved quantum numbers, i.e. we require that $p_i = p_j$ for every pair of indices with $E_i = E_j$. Then we can find a function f such that $\widehat{\rho} = f(\widehat{H})$, which is just another notation for demanding that $p_i = f(E_i)$ in (1). It is also obvious that in this case f can be chosen as a smooth, positive function such that $\int \mathrm{d}E\, f(E) < \infty$. Then we can define a smooth probability density F by the formula $F(E) = f(E)/\int f$, using which the density operator becomes

$$\widehat{\rho} = \frac{F(\widehat{H})}{\mathrm{Tr}\, F(\widehat{H})}. \tag{2}$$

We shall call any smooth probability density F which satisfies (2) a *fluctuation spectrum* of the system. If we require that the energy spectrum is bounded from below, we can normalize the Hamiltonian so that the lowest eigenvalue is strictly positive and then choose a fluctuation spectrum with a support on the positive real axis.

Finally, let us define a few mathematical tools needed in the following discussion. We shall denote the Gaussian probability density with mean 0 and standard deviation λ by G_λ, i.e.

$$G_\lambda(x) = \frac{1}{\sqrt{2\pi\lambda^2}} \exp\left(-\frac{x^2}{2\lambda^2}\right).$$

For a pointwise application of inverse Fourier transforms, we shall also need the class $\mathcal{S}$ of the Schwarz test functions, also called rapidly decreasing functions. They are infinitely many times differentiable, i.e. smooth functions which have the property that the function and any of its derivatives vanish faster than any power at infinity—the distribution G_λ is a prime example of a rapidly decreasing function. For a more precise definition see, for example, chapter 6 of the book on functional analysis by Rudin [4].

3. An overview of the results for standard ensembles

The purpose of this section is to give a brief review of the main results in the special case when they lead to the usual canonical and Gaussian ensembles. We have chosen this peculiar order of presentation, so that there would be a concrete application which can be kept in mind when going through the more abstract results in the following section. It also serves as an introduction to the methods we apply in the following section. We do not imply that these results are obvious—we shall go through their derivations in detail in section 4.

Let F be a fluctuation spectrum of an equilibrium system for which energy is the only relevant conserved quantity and let $\widehat{A}$ be a positive observable. Also assume that:

A1. The system is canonical: its spectrum is bounded from below and $\operatorname{Tr} \mathrm{e}^{-\beta \widehat{H}} < \infty$ for all $\beta > 0$.

A2. The energy fluctuations decay faster than exponentially at high energies: $\mathrm{e}^{\beta E} F(E)$ is a rapidly decreasing function for all real β.

A3. The observable is canonical: $\operatorname{Tr}(\widehat{A}\mathrm{e}^{-\beta \widehat{H}}) < \infty$ for all $\beta > 0$.

We shall call a fluctuation spectrum satisfying condition A2 precanonical and we shall use the name precanonical system for those systems which are canonical and have precanonical fluctuations. Note also that by A1 we can choose F so that its support is bounded from below and thus A2 is a restriction for the behaviour of the fluctuations only near $E \to +\infty$.

The assumptions A1–A3 allow for all $\beta > 0$ the integral representation

$$\operatorname{Tr}\left(\widehat{A}F(\widehat{H})\right) = \int_{\beta-\mathrm{i}\infty}^{\beta+\mathrm{i}\infty} \frac{\mathrm{d}w}{2\pi\mathrm{i}} \bar{F}(w) \operatorname{Tr}\left(\widehat{A}\mathrm{e}^{-w\widehat{H}}\right) \tag{3}$$

where $\bar{F}$ is the Laplace transform of F and the integrand in the above equation is an analytic function in the half-plane $\operatorname{Re} w > 0$. We can use saddle-point methods to find a better integration contour and approximations to the integral. In fact, usually there exists a unique positive saddle point β_A which will be the best choice for β in (3) as the integration contour will then go through this saddle point via the path of steepest descent.

The saddle-point approximation around β_A will then yield 'the best possible canonical approximation' of the trace. This saddle-point approximation can be expressed in terms of $\bar{g}_A(\beta) = \langle \widehat{A}\widehat{H}\rangle/\langle\widehat{A}\rangle$ and $\sigma_A^2(\beta) = \langle \widehat{A}(\widehat{H} - \bar{g}_A)^2\rangle/\langle\widehat{A}\rangle$, where all expectation values are taken in the canonical ensemble at inverse temperature β, and in terms of the following two characteristics of the fluctuation spectrum:

$$a(\beta) = \frac{\int \mathrm{d}x\, F(x)\, \mathrm{e}^{\beta x} x}{\int \mathrm{d}x\, F(x)\, \mathrm{e}^{\beta x}} \qquad \text{and} \qquad b^2(\beta) = \frac{\int \mathrm{d}x\, F(x)\, \mathrm{e}^{\beta x}[x - a(\beta)]^2}{\int \mathrm{d}x\, F(x)\, \mathrm{e}^{\beta x}}. \tag{4}$$

The saddle-point equation is then

$$a(\beta) = \bar{g}_A(\beta)$$

and there is unique a positive solution, β_A, to this equation provided

$$a(0) < \lim_{\beta \to 0^+} \bar{g}_A(\beta). \tag{5}$$

When this is the case, the saddle-point approximation around β_A yields

$$\operatorname{Tr}\left(\widehat{A}F(\widehat{H})\right) \approx \left[2\pi(b^2 + \sigma_A^2)\right]^{-\frac{1}{2}} \mathrm{e}^{a\beta_A + \frac{1}{2}b^2\beta_A^2} \operatorname{Tr}\left(\widehat{A}\mathrm{e}^{-\beta_A \widehat{H}}\right) \tag{6}$$

where all functions are to be evaluated at $\beta = \beta_A$.

Suppose now that the positive saddle-point approximation is good for the partition function, i.e. for $\widehat{A} = \widehat{1}$, and let β_0 be the corresponding saddle-point value. Assume also that $\widehat{A}$ is an observable which does not alter the canonical distribution much, $\bar{g}_A(\beta) = \langle \widehat{A}\widehat{H}\rangle_\beta/\langle\widehat{A}\rangle_\beta \approx \langle\widehat{H}\rangle_\beta$ and $\sigma_A^2 \approx \langle\widehat{H}^2\rangle - \langle\widehat{H}\rangle^2$. Then we can also use β_0 for the saddle-point approximation of $\operatorname{Tr}[\widehat{A}F(\widehat{H})]$ and arrive at the usual canonical formula

$$\frac{\operatorname{Tr}\left[\widehat{A}F(\widehat{H})\right]}{\operatorname{Tr} F(\widehat{H})} \approx \frac{\operatorname{Tr}\left[\widehat{A}\mathrm{e}^{-\beta_0\widehat{H}}\right]}{\operatorname{Tr}\mathrm{e}^{-\beta_0\widehat{H}}}.$$

For large systems, we are typically only interested in 'total' observables such as the total energy and total particle number. These observables depend only on the behaviour of the system at the scale of its total energy and, therefore, we would expect to need only information

about the behaviour of the system coarse grained up to a scale Λ, which is less than the total energy but larger than the microscopic energy scales.

For precanonical systems, it is possible to define a Gaussian approximation which will yield accurate results for most large-scale† canonical observables: if we smooth the fluctuation spectrum by taking a convolution with a Gaussian distribution with variance Λ, then after defining $L = a(0)$ and $\lambda^2 = b^2(0)$, we obtain in the limit $\Lambda \to \infty$,

$$\mathrm{Tr}\left[\widehat{A}\,(F * G_\Lambda)(\widehat{H})\right] = \mathrm{Tr}\left[\widehat{A}\,(G_\lambda * G_\Lambda)(L - \widehat{H})\right]\left(1 + \mathrm{O}(\Lambda^{-3})\right). \tag{7}$$

This choice of parameters L and λ^2 is natural since they are the expectation value and variance of the probability distribution F, as can be seen from (4). They are, in fact, also the best possible, since in theorem 1 we shall prove that any other choice of L and λ^2 guarantees only a slower convergence in $1/\Lambda$.

The precise conditions for an application of the theorem are discussed in section 4.2 and one set of sufficient conditions is given by:

(a) $\lim_{\beta\to 0^+} \bar{g}_A(\beta) = \infty$,
(b) $\beta\bar{g}_A(\beta)$ and $\beta\sigma_A^2(\beta)/\bar{g}_A(\beta)$ stay bounded in the limit $\beta \to 0^+$.

It is relatively easy to check that these conditions hold, for example, for all energy moments ($\widehat{A} = \widehat{H}^k$) when the system consists of N particles in a harmonic potential. In addition, by (5), condition (a) also guarantees the existence of a positive saddle-point solution and thus of a canonical approximation.

If these two conditions also hold for $\widehat{A} = \widehat{1}$, we can then see that the Gaussian ensemble yields an accurate approximation to the original ensemble,

$$\frac{\mathrm{Tr}\left[\widehat{A}F(\widehat{H})\right]}{\mathrm{Tr}\,F(\widehat{H})} \approx \frac{\mathrm{Tr}\left[\widehat{A}G_\lambda(L - \widehat{H})\right]}{\mathrm{Tr}\,G_\lambda(L - \widehat{H})}$$

at least for all sufficiently large-scale observables $\widehat{A}$.

4. The positive saddle-point approximation

In the preceding section we anticipated the derivation of the canonical ensemble from a saddle-point approximation under very general assumptions. Is there any need for non-canonical ensembles? In principle, the canonical approximation can fail in several different ways as the following examples illustrate.

(a) The density of states increases faster than exponentially or is not discrete and thus the canonical ensemble is ill-defined.
(b) F decays only, for example, polynomially in energy. Then the above arguments relying on analyticity will not hold and the Gaussian approximation might not be applicable.
(c) The addition of the observable $\widehat{A}$ alters the canonical energy distribution too radically and either we cannot use $\beta_A = \beta_0$ or, even worse, $\mathrm{Tr}(\widehat{A}\mathrm{e}^{-\beta\widehat{H}}) = \infty$.
(d) The canonical ensemble gives too rough an approximation for the real statistics. This will always eventually happen when observables resolving ever smaller energy scales are inspected, unless the original fluctuation spectrum is *exactly* canonical.

† These can be defined as observables whose expectation value changes by less than some preset limit when the fluctuation spectrum is coarse grained.

Let us now first concentrate on solving the 'exotic' part of the problem, the exponential increase of density of states (cases (a) and (c)) and the possibility of non-exponentially decaying fluctuations (case (b)). A solution in both cases is to reparametrize the energy so that in the new variables the asymptotic high-energy behaviour is sufficiently regular.

Let $g(E)$ be the function which performs this reparametrization, i.e. let g is be an orientation-preserving diffeomorphism from the real line onto itself. Since we do not want to restrict the speed of the growth of the density of states in any way, we do not assume anything more about the diffeomorphism at this stage, only that it regularizes the high-energy behaviour of both the density of states and the fluctuation spectrum: we assume that in the reparametrized variables the density of states (with the effect of the observable included) does not increase exponentially and that the fluctuation spectrum decreases at least exponentially.

Expressed mathematically, we assume that there exist real numbers β_- and β_+ such that $\beta_- \leqslant 0 < \beta_+$, and:

B1. For all $\beta > \beta_-$, $\widehat{A}$ is a positive g-bounded observable: $\mathrm{Tr}\big(\widehat{A}\mathrm{e}^{-\beta g(\widehat{H})}\big) < \infty$.

B2. The reparametrized fluctuations decrease at least exponentially at high energies: $\mathrm{e}^{\beta x}F(g^{-1}(x))$ is a rapidly decreasing function for all $\beta < \beta_+$.

We also assume that these parameters have been chosen in the best possible manner,

$$\beta_- = \inf\left\{\beta \;\middle|\; \mathrm{Tr}\big(\widehat{A}\mathrm{e}^{-\beta g(\widehat{H})}\big) < \infty\right\} \tag{8}$$

$$\beta_+ = \sup\left\{\beta \;\middle|\; \mathrm{e}^{\beta x}F(g^{-1}(x)) \in \mathcal{S}\right\} \tag{9}$$

and thus either of them might be infinite. Note also that β_- depends on the choice of the observable $\widehat{A}$.

It looks like we lost some generality at this stage. However, by adding a suitable part of the energy dependence to the observable, i.e. by replacing $\widehat{A}$ by $\widehat{A}\Phi(\widehat{H})$, where $\Phi(E)$ is a suitable function, it is usually possible to find such a reparametrization g that β_- and β_+ satisfy the requirement $\beta_- \leqslant 0 < \beta_+$. For example, if the density of states increases like $\mathrm{e}^{\beta'\widehat{H}}$ choosing $\Phi(E) = \mathrm{e}^{-\beta' E}$ would still allow $g = \mathrm{id}$.

Suppose now that we would like to approximate the behaviour of the system by using the reparametrized Gaussian ensemble, $\widehat{\rho} = G_\lambda(L - g(\widehat{H}))$, or the reparametrized canonical ensemble, $\widehat{\rho} = \mathrm{e}^{-\beta g(\widehat{H})}$. It would be natural to assume that the best approximation of this kind follows if we choose the parameters L and λ as the reparametrized expectation value and variance of F and β from the saddle-point approximation of the trace in the partition function. We will now concentrate on deriving conditions under which this intuition is valid and on finding estimates for the errors induced by the approximations.

4.1. Canonical approximation of a trace

It follows easily from assumption B2 that the function

$$\bar{F}(w) = \int_{-\infty}^{\infty} \mathrm{d}x\, \mathrm{e}^{xw}F(g^{-1}(x)) \tag{10}$$

is well defined and analytic in the region $\mathrm{Re}\, w < \beta_+$. Since g was a diffeomorphism and we can typically choose F with a support on the positive real axis, we can usually also define this 'g-transform' of F by

$$\bar{F}(w) = \int_0^{\infty} \mathrm{d}E\, g'(E)\, \mathrm{e}^{g(E)w}F(E).$$

When $g(E) = E$ this is just the Laplace transform and if $g(E) = \ln E$ it coincides with the Mellin transform.

As the inverse Fourier transformation can be taken pointwise for rapidly decreasing functions, we have pointwise for all $\beta < \beta_+$,

$$F(E) = \int \frac{\mathrm{d}p}{2\pi} \mathrm{e}^{\mathrm{i}pg(E)-\beta g(E)} \bar{F}(\beta - \mathrm{i}p) = \int_{\beta-\mathrm{i}\infty}^{\beta+\mathrm{i}\infty} \frac{\mathrm{d}w}{2\pi\mathrm{i}} \bar{F}(w)\, \mathrm{e}^{-wg(E)}. \tag{11}$$

By applying the representation (11) we then obtain for any observable satisfying B1 an integral representation valid for $\beta_- < \beta < \beta_+$,

$$\mathrm{Tr}\big(\widehat{A} F(\widehat{H})\big) = \int_{\beta-\mathrm{i}\infty}^{\beta+\mathrm{i}\infty} \frac{\mathrm{d}w}{2\pi\mathrm{i}} \bar{F}(w)\, \mathrm{Tr}\big(\widehat{A}\mathrm{e}^{-wg(\widehat{H})}\big). \tag{12}$$

The boundedness of the trace also implies that the function $\mathrm{Tr}\big(\widehat{A}\mathrm{e}^{-wg(\widehat{H})}\big)$ is analytic in the half-plane $\mathrm{Re}\, w > \beta_-$ and, therefore, the integrand in (12) is an analytic function in the strip $\beta_- < \mathrm{Re}\, w < \beta_+$ and we proceed to consider its evaluation by the method of steepest descent.

For writing down the saddle-point approximation it will be useful to define

$$\begin{aligned} a(w) &= \frac{\int \mathrm{d}x\, F(g^{-1}(x))\, \mathrm{e}^{wx} x}{\int \mathrm{d}x\, F(g^{-1}(x))\, \mathrm{e}^{wx}} \qquad \text{and} \qquad b^2(w) = \frac{\int \mathrm{d}x\, F(g^{-1}(x))\, \mathrm{e}^{wx} [x - a(w)]^2}{\int \mathrm{d}x\, F(g^{-1}(x))\, \mathrm{e}^{wx}} \\ \bar{g}(w) &= \frac{\mathrm{Tr}\left(\widehat{A}\mathrm{e}^{-wg(\widehat{H})} g(\widehat{H})\right)}{\mathrm{Tr}\left(\widehat{A}\mathrm{e}^{-wg(\widehat{H})}\right)} \qquad \text{and} \qquad \sigma^2(w) = \frac{\mathrm{Tr}\left(\widehat{A}\mathrm{e}^{-wg(\widehat{H})} \big[g(\widehat{H}) - \bar{g}(w)\big]^2\right)}{\mathrm{Tr}\left(\widehat{A}\mathrm{e}^{-wg(\widehat{H})}\right)} \end{aligned} \tag{13}$$

which are well defined over the whole strip apart from the countable set of zeros of the denominators.

In terms of these quantities, the saddle-point equation becomes

$$a(w) - \bar{g}(w) = 0 \tag{14}$$

and the second derivative of the logarithm of the integrand is given by

$$b^2(w) + \sigma^2(w). \tag{15}$$

For real values of w, both b^2 and σ^2 are strictly positive and thus the second derivative becomes a strictly positive quantity on the real part of the strip. Therefore, the first derivative defining the saddle-point equation is a strictly increasing function on the interval $\beta_- < w < \beta_+$ and there exists a real saddle point on this interval if and only if

$$a(\beta_-) < \bar{g}(\beta_-) \qquad \text{and} \qquad a(\beta_+) > \bar{g}(\beta_+). \tag{16}$$

If the real saddle point exists, it is obviously unique and choosing β in (12) equal to this saddle-point value will lead to an integration contour which goes through the real saddle point via the steepest-descent path.

Assume then that the real saddle point exists, denote it by β_A, and parametrize the integration variable in (12) as $w = \beta_A + \mathrm{i}\alpha$. If the integrand, as a function of α, is strongly concentrated in such a neighbourhood of the origin that the quadratic approximation of the logarithm of the integrand is admissible, then the saddle-point approximation yields

$$\mathrm{Tr}\big(\widehat{A} F_\Lambda(\widehat{H})\big) \approx \frac{1}{\sqrt{2\pi(b^2(\beta_A) + \sigma^2(\beta_A))}} \bar{F}(\beta_A)\, \mathrm{Tr}\big(\widehat{A}\mathrm{e}^{-\beta_A g(\widehat{H})}\big).$$

Note that this can happen, since the maximum of the absolute value of the integrand is at $\alpha = 0$, although it cannot be guaranteed without further assumptions. For a derivation of bounds for the accuracy of the saddle-point approximation, we shall need to make a detour via a Gaussian approximation.

4.2. Gaussian approximation of a trace

Suppose we want to coarse grain the fluctuation spectrum as in the previous section and in that way deduce what kind of Gaussian approximation would be most accurate for large-scale observables. The use of a Gaussian convolution is not justified now since we do not know whether the resulting function will give a trace-class operator. Moreover, the transformed function might no longer satisfy the decay condition B2 and we could not use the above integral representation for it.

There is, however, a natural generalization of the usual coarse-graining procedure which will give functions satisfying condition B2: we shall perform the coarse graining by a convolution with a Gaussian distribution *in the reparametrized space*,

$$F_\Lambda(E) = \int \mathrm{d}y\, F(g^{-1}(y))G_\Lambda(g(E)-y)$$
$$= \int \mathrm{d}x\, F(x)g'(x)\frac{1}{\sqrt{2\pi\Lambda^2}}\exp\left[-\frac{1}{2\Lambda^2}(g(E)-g(x))^2\right].$$

It is obvious that this transformation satisfies the following semigroup property:

$$(F_\Lambda)_{\Lambda'} = F_{\sqrt{\Lambda^2+\Lambda'^2}}$$

and that the g-transform changes simply by multiplication,

$$\bar{F}_\Lambda(w) = \mathrm{e}^{\frac{1}{2}\Lambda^2 w^2}\bar{F}(w).$$

As the Gaussian distribution is a rapidly decreasing function, its convolution leaves the class $\mathcal{S}$ invariant. Therefore, if $\mathrm{e}^{\beta x}F(g^{-1}(x))$ is rapidly decreasing, so is $\mathrm{e}^{\beta x}F_\Lambda(g^{-1}(x))$, since clearly

$$\mathrm{e}^{\beta x}F_\Lambda(g^{-1}(x)) = \mathrm{e}^{\frac{1}{2}\beta^2\Lambda^2}\int \mathrm{d}y\, \mathrm{e}^{\beta y}F(g^{-1}(y))G_\Lambda(x-\beta\Lambda^2-y).$$

This proves that if F satisfies condition B2 for some β_+ then so does F_Λ.

By the above results, we now have an integral representation

$$\mathrm{Tr}\big(\widehat{A}F_\Lambda(\widehat{H})\big) = \int_{\beta-\mathrm{i}\infty}^{\beta+\mathrm{i}\infty}\frac{\mathrm{d}w}{2\pi\mathrm{i}}\mathrm{e}^{\frac{1}{2}\Lambda^2w^2}\bar{F}(w)\,\mathrm{Tr}\big(\widehat{A}\mathrm{e}^{-wg(\widehat{H})}\big) \tag{17}$$

which is valid for all $\beta_- < \beta < \beta_+$ and $\Lambda \geqslant 0$. We would now like to make a quadratic approximation of $\bar{F}(w)$ which would then lead to a Gaussian trace. In fact, this is often possible in the limit of large Λ, but since the discussion becomes a bit technical at this point, we leave the proof of this to appendix A and only state the results here.

Define the functions a, b^2, $\bar{g}$ and σ^2 by equations (13). Then $a(0)$ and $b^2(0)$ are the expectation value and the variance of the reparametrized fluctuation spectrum and we proceed to show that these values yield the best possible approximation of a Gaussian form. If $\beta_- = 0$, we also need to assume that the following three conditions are satisfied:

G1. $a(0) < \bar{g}(0)$.

G2. There exists a constant $c \geqslant 0$ for which

$$\beta^{1+2c}\bar{g}(\beta) \text{ stays bounded in the limit } \beta \to 0^+. \tag{18}$$

G3. $\beta\sigma^2(\beta)/\bar{g}(\beta)$ stays bounded in the limit $\beta \to 0^+$.

With this notation, the following theorem, whose proof we have included in appendix A, holds:

Theorem 1. *Let $L = a(0)$ and $\lambda^2 = b^2(0)$ and assume that B1 and B2 are satisfied for some $\beta_- \leqslant 0 < \beta_+$. If $\beta_- = 0$ and G1–G3 hold, then in the large-scale limit $\Lambda \to \infty$ for all $c \geqslant 0$ satisfying (18),*

$$\mathrm{Tr}\big(\widehat{A} F_\Lambda(\widehat{H})\big) = \bar{F}(0)\,\mathrm{Tr}\big[\widehat{A} G_{\sqrt{\lambda^2+\Lambda^2}}(L - g(\widehat{H}))\big]\big(1 + \mathrm{O}(\Lambda^{-3/(1+c)})\big) \tag{19}$$

and if $\beta_- < 0$, then (19) is true for $c = 0$.

In addition, for any other choice of L or λ^2 only a slower convergence in this limit can be guaranteed.

Since $G_{\sqrt{\lambda^2+\Lambda^2}} = (G_\lambda)_\Lambda$, we can thus conclude that, when the quality of the approximation is measured by expectation values of large-scale observables, the best approximation of the Gaussian form for $F(\widehat{H})$ is given by

$$F(\widehat{H}) \approx \bar{F}(0) G_\lambda(L - g(\widehat{H})). \tag{20}$$

Therefore, we have proved that having either $\beta_- < 0$ or conditions G1–G3 satisfied will lead to the intuitive approximation mentioned earlier in section 4.1.

The standard situation we examined in section 3 implies that $\beta_- = 0$, $\beta_+ = \infty$ and $g(x) = x$. Thus we then need to satisfy the latter three conditions. In fact, it is straightforward to see that conditions (a) and (b) given in section 3 do imply that G1 is satisfied for any $a(0)$, that we can choose $c = 0$ in condition G2, and that G3 holds. Theorem 1 will then justify equation (7).

4.3. Accuracy of the canonical approximation

The canonical approximation was derived in section 4.1 from a real saddle-point approximation of the trace. Estimating the accuracy of this approximation is, however, quite difficult. In this section we shall present estimates for the accuracy of the saddle-point approximation in the special case when the fluctuation spectrum is Gaussian. By the results of the previous section, this can then also be used for more general fluctuation spectra, if the observable is of sufficiently large scale and the conditions for the use of the Gaussian approximation are valid.

Thus assume that the fluctuation spectrum is $F(\widehat{H}) = G_\lambda(L - \widehat{H})$, when $\beta_+ = \infty$, and let β be some value greater than β_-. Later we shall see that the best bounds are obtained if β is the solution to the saddle-point equation

$$\beta\lambda^2 + L - \bar{g}(\beta) = 0. \tag{21}$$

However, since we shall also need estimates for how far from the saddle-point value β can be chosen before the saddle-point approximation loses its accuracy, we shall not fix the value of β yet.

The bounds will be expressed in terms of the ratio $R = \lambda/\sigma(\beta)$. Since the derivations of the bounds are again quite technical, we postpone them to appendix B and only present the results here. The real saddle-point approximation to the Gaussian trace is shown there to yield

$$\mathrm{Tr}\big[\widehat{A} G_\lambda(L - g(\widehat{H}))\big] \approx \mathrm{e}^{\beta L + \frac{1}{2}\beta^2\lambda^2} G_{\sqrt{\lambda^2+\sigma^2(\beta)}}(L + \beta\lambda^2 - \bar{g}(\beta))\,\mathrm{Tr}\big[\widehat{A}\mathrm{e}^{-\beta g(\widehat{H})}\big] \tag{22}$$

and we call the right-hand side 'saddle-point approximation' and denote it by 'PSPA' even when β does not satisfy the saddle-point equation (21).

The first bound for the accuracy of this approximation is obtained by using Jensen's inequality very analogously to what was done in [1]. This shows that for any β the ratio of the Gaussian trace and its PSP approximation is bounded by

$$\exp\left[\frac{\big(L + \beta\lambda^2 - \bar{g}(\beta)\big)^2}{2(\lambda^2 + \sigma^2(\beta))}\right]\left(1 + \frac{1}{R^2}\right)^{\frac{1}{2}} \geqslant \frac{\mathrm{Tr}\big[\widehat{A} G_\lambda(L - g(\widehat{H}))\big]}{\mathrm{PSPA}} \geqslant \left(1 + \frac{1}{R^2}\right)^{\frac{1}{2}} \exp\left[-\frac{1}{2R^2}\right]. \tag{23}$$

These inequalities prove that, if β is close to the saddle-point value and $R \gg 1$, the relative error from the PSPA becomes negligible.

The second bound for the relative error is given for Δ_{PSPA} in

$$\frac{\mathrm{Tr}\big[\widehat{A}G_\lambda(L - g(\widehat{H}))\big]}{\mathrm{PSPA}} = 1 + \Delta_{\mathrm{PSPA}}$$

and it depends on the behaviour of $\sigma^2(\beta + \mathrm{i}\alpha)$ near $\alpha = 0$. Since $\sigma^2(\beta) \neq 0$, we can for all $0 < r < 1$ find a $\rho_r > 0$ such that

$$\left|\frac{\sigma^2(\beta + \mathrm{i}\alpha)}{\sigma^2(\beta)} - 1\right| \leqslant r \qquad \text{for all} \quad -\rho_r \leqslant \alpha \leqslant \rho_r. \tag{24}$$

The bound is then expressed in terms of r and ρ_r as

$$\exp\left[\frac{\big(L + \beta\lambda^2 - \bar{g}(\beta)\big)^2}{2(\lambda^2 + \sigma^2(\beta))}\right]|\Delta_{\mathrm{PSPA}}| \leqslant \frac{3}{2}\frac{r}{1+R^2}\left(1 - \frac{r}{1+R^2}\right)^{-\frac{3}{2}} + \mathrm{e}^{-\frac{1}{2}\rho_r^2\lambda^2}\left[1 + \left(1 + \frac{1}{R^2}\right)^{\frac{1}{2}}\right]. \tag{25}$$

Clearly, the bound is informative only if λ is so large that $\lambda\rho_r \geqslant 1$.

If β is so close to the saddle-point value that $\big(L + \beta\lambda^2 - \bar{g}(\beta)\big)^2 \ll \sigma^2(\beta) + \lambda^2$ and if r can be chosen so that $|\Delta_{\mathrm{PSPA}}|$ is small, then the PSP approximation of a Gaussian trace is very accurate. By applying the first bound (23), we can see that this will happen after a coarse graining with a large enough Λ, if β can be chosen so that $\Lambda \gg \sigma(\beta)$. Unfortunately, unlike was erroneously claimed in section 4 in [1], the last condition cannot be satisfied in a typical thermodynamical limit and the more complicated second bound has to be applied in these cases.

4.4. Expectation values

So far we have inspected the approximation of one trace only. The statistical expectation values are ratios of two traces and as some of the terms cancel in the ratio, we can write down simpler results for the expectation values. To avoid confusion, we shall use the subscript A to denote the quantities related to the trace in the numerator (i.e. depending on the observable) and the subscript 0 for the quantities related to the denominator, which is independent of the observable.

We also take into account the possibility of having the behaviour of part of the original fluctuation spectrum predetermined. If $\Phi(x)$ describes the known behaviour, we use a new fluctuation spectrum determined by $f(x)$ which satisfies $F(x) = f(x)\Phi(x)$ and the observable $\widehat{\Phi}\widehat{A} = \Phi(\widehat{H})\widehat{A}$ instead of the original fluctuation spectrum $F(x)$ and observable $\widehat{A}$. Such a separation is sometimes practical as we shall see in section 7, or even necessary as we noted at the beginning of section 4.

We shall assume in the following that the Gaussian approximation is valid for the observable $\widehat{A}$, when for parameters $L = a(0)$ and $\lambda^2 = b^2(0)$,

$$\langle\widehat{A}\rangle \approx \frac{\mathrm{Tr}\left[\widehat{\Phi}\widehat{A}G_\lambda\left(L - g(\widehat{H})\right)\right]}{\mathrm{Tr}\left[\widehat{\Phi}G_\lambda\left(L - g(\widehat{H})\right)\right]}. \tag{26}$$

We shall then apply the PSP approximation to the Gaussian traces. The saddle points β_0 and β_A are determined by the equations

$$\beta_0\lambda^2 + L - \bar{g}_0(\beta_0) = 0 \qquad \text{and} \qquad \beta_A\lambda^2 + L - \bar{g}_A(\beta_A) = 0 \tag{27}$$

and, in general, they are not equivalent. We estimated in the preceding section the accuracy of the PSP approximation for the Gaussian traces and the bounds derived there can be used for determining whether these approximations are valid. When this is the case, we find from (22) and (27)

$$\langle\widehat{A}\rangle \approx \left(\frac{\lambda^2+\sigma_0^2(\beta_0)}{\lambda^2+\sigma_A^2(\beta_A)}\right)^{\frac{1}{2}} \exp\left[\tfrac{1}{2}(\beta_A-\beta_0)^2\lambda^2\right] \frac{\mathrm{Tr}\left[\widehat{\Phi}\widehat{A}\mathrm{e}^{\beta_A(\bar{g}_0-g(\widehat{H}))}\right]}{\mathrm{Tr}\left[\widehat{\Phi}\mathrm{e}^{\beta_0(\bar{g}_0-g(\widehat{H}))}\right]} \tag{28}$$

where $\bar{g}_0 = \bar{g}_0(\beta_0) = \mathrm{Tr}\left[\widehat{\Phi}g(\widehat{H})\,\mathrm{e}^{-\beta_0 g(\widehat{H})}\right]/\mathrm{Tr}\left[\widehat{\Phi}\mathrm{e}^{-\beta_0 g(\widehat{H})}\right]$.

This result can be simplified for those observables, for which it is possible to use $\beta_A = \beta_0$. By the discussion in the previous section, this is possible at least when $(\bar{g}_0(\beta_0) - \bar{g}_A(\beta_0))^2 \ll \lambda^2 + \sigma_A^2(\beta_0)$, i.e. the addition of the observable does not change the value of $\bar{g}$ significantly. If also $\left|\sigma_A^2(\beta_0) - \sigma_0^2(\beta_0)\right| \ll \lambda^2 + \sigma_0^2(\beta_0)$, the expectation value can be approximated by the usual canonical formula,

$$\langle\widehat{A}\rangle \approx \frac{\mathrm{Tr}\left[\widehat{\Phi}\widehat{A}\mathrm{e}^{-\beta_0 g(\widehat{H})}\right]}{\mathrm{Tr}\left[\widehat{\Phi}\mathrm{e}^{-\beta_0 g(\widehat{H})}\right]}. \tag{29}$$

5. Essentially microcanonical systems and PSP entropy

Most applications of statistical methods to physical systems consider a large number of particles constrained into a fixed region of space, either by having them in a container (gases and liquids) with (potential) walls which prevent the particles from escaping, or by an attractive interaction between the particles (solids). In both cases, the interactions with the environment happen via the boundary of the container and it is usually plausible to assume that in an equilibrium the total energy fluctuations are negligible.

This motivates the use of the microcanonical ensemble, where energy fluctuations are neglected entirely and the density operator is proportional to $\delta(\widehat{H} - E)$. However, in real systems, there are interactions with the environment, and although the energy fluctuations are small, they are not non-existent. One of the motivations for our inspection of the Gaussian ensembles was to develop methods for examination of the effect of these fluctuations.

We want now to inspect how and when the fluctuations *can* be neglected. First, we shall assume that the fluctuation spectrum has a compact support, i.e. it is zero outside some finite interval, and we shall denote its mean and variance by

$$E = \int \mathrm{d}x\, F(x)x \qquad \text{and} \qquad \varepsilon^2 = \int \mathrm{d}x\, F(x)(x-E)^2. \tag{30}$$

We shall also assume that the reparametrization g is essentially linear on the support of F, i.e. that we can use the approximation $g(y) \approx g(E) + (y-E)g'(E)$ when computing $\bar{F}$. This leads to the approximations

$$\bar{F}(0) \approx g'(E) \qquad a(0) \approx g(E) \qquad \text{and} \qquad b^2(0) \approx g'(E)^2\varepsilon^2$$

which become exact in the limit $\varepsilon \to 0$. Note also that since the support of F is compact, we now have $\beta_+ = \infty$ and the parameter range for β is thus given by $\beta > \beta_-$.

Thus we can define the parameters of the generalized Gaussian ensemble by $L = g(E)$ and $\lambda^2 = g'(E)^2\varepsilon^2$, and the saddle equation yielding the best canonical approximation becomes

$$\beta\lambda^2 + L - \bar{g}(\beta) = 0.$$

We would like to put $\lambda \to 0$ in this equation, and consider only solutions to the simpler equation

$$g(E) = \bar{g}(\beta). \tag{31}$$

We have seen that this kind of change to the saddle-point value is possible provided the new value satisfies $(L + \beta\lambda^2 - \bar{g}(\beta))^2 \ll \lambda^2 + \sigma^2(\beta)$, and thus the solution of (31) is a good approximation if, for example, $\beta^2\lambda^2 \ll \beta\sigma(\beta)$. We now assume that ε is so small that this approximation can be made and we shall call such systems essentially microcanonical.

A straightforward application of the dominated convergence theorem reveals that $\lim_{\beta\to\infty} \bar{g}(\beta) = g(E_0)$, where E_0 is the lowest energy for which the expectation value of the observable $\widehat{A}$ is non-vanishing. For example, if $\widehat{A} = \widehat{1}$ then E_0 is the ground state energy. Since $\bar{g}$ is continuous and monotonically decreasing, we can then deduce that a necessary and sufficient condition for equation (31) to have solutions $\beta > \beta_-$ is given by $g(E_0) < L < \bar{g}(\beta_-)$ or, if expressed in terms of energy,

$$E_0 < E < g^{-1}(\bar{g}(\beta_-)).$$

The microcanonical entropy S_{mc} is defined as the logarithm of the density of states. One possible quantitative definition for the density of states at energy E at the scale ε, would be given by the number of states in the energy interval $[E - \frac{1}{2}\varepsilon, E + \frac{1}{2}\varepsilon]$ divided by the length of the interval, ε. Then the microcanonical entropy $S_{\mathrm{mc}}(E, \varepsilon)$ would be equal to $\ln \operatorname{Tr} F(\widehat{H})$, where F is the fluctuation spectrum proportional to the characteristic function of the interval $[E - \frac{1}{2}\varepsilon, E + \frac{1}{2}\varepsilon]$. We shall generalize this a bit and inspect the approximation of $\ln \operatorname{Tr} F(\widehat{H})$ for a general fluctuation spectrum F. The Gaussian (20) and saddle-point approximations (22) yield the estimates

$$\ln \operatorname{Tr} F(\widehat{H}) \approx \ln \frac{\bar{F}(0)}{\sqrt{2\pi\lambda^2}} + \ln \operatorname{Tr} \exp\left(-\frac{(L - g(\widehat{H}))^2}{2\lambda^2}\right) \tag{32}$$

$$\approx -\frac{1}{2}\ln(2\pi) - \frac{1}{2}\ln\frac{\lambda^2 + \sigma^2}{\bar{F}(0)^2} + \beta\bar{g}(\beta) + \ln \operatorname{Tr} \mathrm{e}^{-\beta g(\widehat{H})} \tag{33}$$

where in the latter formula we have used β defined by equation (31) for the choice $\widehat{A} = \widehat{1}$. We have also not included the possible prefactors $\widehat{\Phi}$ which where introduced in section 4.4—these would unnecessarily complicate the formulae and their inclusion only amounts to the replacements $\operatorname{Tr} \to \operatorname{Tr} \widehat{\Phi}$.

In (33), the first two terms are logarithmic corrections depending on the scale of the fluctuations and they are usually dominated by the remaining terms. For reasons which will become apparent soon, we shall call the remaining terms the *PSP entropy,*

$$S_{\mathrm{psp}}(\beta; \widehat{H}) = \beta\frac{\operatorname{Tr}\left(g(\widehat{H})\,\mathrm{e}^{-\beta g(\widehat{H})}\right)}{\operatorname{Tr}\mathrm{e}^{-\beta g(\widehat{H})}} + \ln \operatorname{Tr}\mathrm{e}^{-\beta g(\widehat{H})}. \tag{34}$$

Comparing this with the definition of the canonical Gibbs entropy, S_{can}, then shows that $S_{\mathrm{psp}}(\beta; \widehat{H}) = S_{\mathrm{can}}(\beta; g(\widehat{H}))$, i.e. this is the same as the Gibbs entropy if the Hamiltonian $\widehat{H}$ is replaced by the reparametrized energy operator $g(\widehat{H})$.

One consequence of this identification is that the maximum entropy principle also holds for S_{psp} in the following form:

Theorem 2 (Maximum PSP entropy). *Suppose $g(\widehat{H})$ is a self-adjoint operator and E is a real parameter such that equation (31) has a solution $\beta > \beta_-$. Then there is a unique positive operator $\widehat{\rho}$ which maximizes the Gibbs entropy functional*

$$S[\widehat{\rho}] = \operatorname{Tr}(-\widehat{\rho}\ln\widehat{\rho})$$

under the restrictions

$$\operatorname{Tr}\widehat{\rho} = 1 \qquad \text{and} \qquad g(E) = \operatorname{Tr}(\widehat{\rho} g(\widehat{H}))$$

and this $\widehat{\rho}$ *has the canonical form*

$$\widehat{\rho} = \frac{\mathrm{e}^{-\beta g(\widehat{H})}}{\operatorname{Tr}\mathrm{e}^{-\beta g(\widehat{H})}}.$$

This theorem follows by an application of the standard result for the canonical ensemble when the self-adjoint operator $\widehat{H}$ is replaced by $g(\widehat{H})$; for a proof of the standard case, see, for example, section 4.2.2 of [3].

Define $\beta(E)$ as the function which maps an energy in the allowed interval $E_0 < E < g^{-1}(\bar{g}(\beta_-))$ to the solution of equation (31). An application of the implicit function theorem then shows that β is differentiable and that

$$\beta'(E) = -\frac{1}{g'(E)}\sigma^2(\beta(E)).$$

Since by equation (34), $\mathrm{d}S_{\mathrm{psp}}(\beta)/\mathrm{d}\beta = -\beta\sigma^2(\beta)$, we can conclude that

$$\frac{\mathrm{d}S_{\mathrm{psp}}(E)}{\mathrm{d}E} = \beta g'(E) \tag{35}$$

where we abuse the notation in the usual way and denote $S_{\mathrm{psp}}(E) = S_{\mathrm{psp}}(\beta(E))$. One immediate conclusion from this equation is that if $\beta_- < 0$, then the inverse $E(S)$ is unique only if the two regions, $E < g^{-1}(\bar{g}(0))$ when $\beta > 0$ and $E > g^{-1}(\bar{g}(0))$ when $\beta < 0$, are treated separately.

5.1. Choosing effective Hamiltonians by the maximum entropy principle

We shall now show that it is possible to use the maximization of S_{psp} for choosing parameters in effective Hamiltonians. This method is a simple extension of the standard one—for an application of the standard method see, for example, chapter 9.3 of the book by Balian [3].

The easiest way of applying the maximum entropy principle is through an effective potential U, which in the standard case $g = \mathrm{id}$ is called free energy and denoted by F. Define U by

$$U(T; \widehat{H}) = -T \ln \operatorname{Tr}\mathrm{e}^{-\frac{1}{T}g(\widehat{H})} \tag{36}$$

for all $T \in \mathbb{R}$ for which the trace converges. A differentiation of this equation then shows that

$$\frac{\mathrm{d}}{\mathrm{d}T}U(T; \widehat{H}) \equiv U'(T; \widehat{H}) = -S_{\mathrm{psp}}\left(\frac{1}{T}; \widehat{H}\right)$$

and thus we obtain the following corollary to theorem 2:

Corollary 3. *Let* $\widehat{H}$ *be the Hamiltonian, let* g *be an energy reparametrization and let* E *be a parameter such that equation (31) has a solution* $\beta(E) > \beta_-$ *and define* $T = 1/\beta(E)$.

If $\widehat{h}$ *is a self-adjoint operator for which*

$$g(E) = \frac{\operatorname{Tr}\left(g(\widehat{H})\,\mathrm{e}^{-\frac{1}{T}g(\widehat{h})}\right)}{\operatorname{Tr}\mathrm{e}^{-\frac{1}{T}g(\widehat{h})}} \tag{37}$$

then

$$U'(T; \widehat{h}) \geqslant U'(T; \widehat{H}) \tag{38}$$

and the minimum is only attained for $\widehat{h} = \widehat{H}$.

Note that in (37) the expectation value is of the *original* Hamiltonian.

The theorem is usually used in the following manner. Suppose that the true Hamiltonian $\widehat{H}$ is too complicated for the computation of the trace in $U(T; \widehat{H})$, but there is a physically plausible effective Hamiltonian $\widehat{H}_{\text{eff}}$ depending on some parameters $\{x_i\}$ for which the function $U(T, \{x_i\}; \widehat{H}_{\text{eff}})$ and its T-derivative $U'(T, \{x_i\}; \widehat{H}_{\text{eff}})$ can be computed. Also assume that it is possible to compute and invert equation (37) for one of the variables. By inserting the solution to $U'(T, \{x_i\}; \widehat{H}_{\text{eff}})$ and minimizing the result with respect to the remaining variables, we find the best possible—in the sense of entropy—parameters for the effective Hamiltonian.

As a curiosity, let us point out that $U(T)$ can also be obtained from a Legendre transform of the reparametrized energy. When $\beta_- = 0$, the inverse of $S_{\text{psp}}(E)$ is unique and if we denote it by $E_{\text{psp}}(S)$, it is easy to check that $U(T)$ the definition

$$U(T) = \inf_S \{g(E_{\text{psp}}(S)) - TS\}$$

is equivalent to (36). If $\beta_- < 0$, the inverse is double valued, and the Legendre transformation of the first branch will define $U(T)$ for $T > 0$ and the transformation of the other branch for $T < 0$.

6. The axioms of Tsallis statistics

In the following sections we shall discuss a non-standard application of the PSP approximation: we shall inspect the connection between the so-called Tsallis statistics and a PSP approximation with a certain non-trivial energy reparametrization. For this comparison, we shall present the main properties of the Tsallis statistics in this section.

In 1988, Tsallis proposed [2] a possible generalization of the principles leading to Boltzmann statistics. Motivated by the usefulness of entropy functionals different from the Gibbs one in the analysis of fractal phenomena, he proposed a two-parameter generalization of the Gibbs entropy and expectation values of observables and then derived a generalization of the Boltzmann statistics using the maximum entropy principle. In the following years, the new results where applied to many interesting non-extensive phenomena, such as stellar polytropes [5], anomalous diffusion [6, 7] and electron plasma columns [8, 9]. The results from a maximization of a Tsallis entropy functional for the electron plasma column were especially promising, since this gave results more consistent with experimental observations than a maximization of a similar Gibbs entropy.

In these applications the principles were set up in the form of axioms, and the following are cited from [8]:

Axiom 1 (Escort probabilities). *The system is described by W microscopic state probabilities $p_i \geqslant 0$, which can be used to define a density matrix $\widehat{\rho}$. These 'escort probabilities' satisfy $\sum_{i=1}^{W} p_i = \operatorname{Tr} \widehat{\rho} = 1$.*

Axiom 2 (q-entropy). *The entropy of the system is defined in terms of two real parameters k and q,*

$$S_q[p_i] = k \frac{\sum_{i=1}^{W} p_i^q - 1}{1 - q} = k \frac{\operatorname{Tr} \widehat{\rho}^q - 1}{1 - q}. \tag{39}$$

Axiom 3 (q-expectation value). *The expectation value of an observable $\widehat{A}$ which has an expectation value a_i in the state number i is given by the formula*

$$\langle \widehat{A} \rangle_q = \sum_{i=1}^{W} p_i^q a_i = \operatorname{Tr}(\widehat{A} \widehat{\rho}^q).$$

With these axioms a maximization of S_q, while holding q and the q-expectation value of the Hamiltonian fixed, will yield [10] the canonical Tsallis ensemble

$$p_i = \frac{1}{Z_q} \max(0, 1 - (1-q)\beta_q \varepsilon_i)^{1/(1-q)} \tag{40}$$

where Z_q enforces the normalization $\sum p_i = 1$ and β_q results from using a Lagrange multiplier technique in the maximization. It is evident from this formula that in the limit $q \to 1$ the result reduces to the standard canonical ensemble.

These axioms are, however, slightly controversial. As has been noted before [11], in this axiomatic version we do not have a proper normalization of the expectation values, since typically $\langle 1 \rangle_q \neq 1$ if $q \neq 1$. Here we propose that the easiest solution, the so-called third choice presented in [12], is the correct solution. In section 7 we present some possible conditions under which the canonical Tsallis ensemble would offer a better description of the system than the canonical Boltzmann ensemble. However, for this we find it necessary for axiom 3 to be replaced by

Axiom 3′ (q-expectation value). *The expectation value of an observable $\widehat{A}$ which has an expectation value a_i in the state number i is given by the formula*

$$\langle \widehat{A} \rangle_q = \frac{\sum_{i=1}^W p_i^q a_i}{\sum_{i=1}^W p_i^q} = \frac{\operatorname{Tr}\left[\widehat{A}\widehat{\rho}^q\right]}{\operatorname{Tr}\widehat{\rho}^q}.$$

We shall discuss the use and meaning of these axioms in more detail in section 8, but first we need to show how the canonical Tsallis ensemble can be obtained from a PSP approximation.

7. Tsallis statistics from a PSP approximation

For the definition of the reparametrization which leads to the Tsallis statistics we need a new energy parameter, E_m. This parameter denotes the maximum energy the system can have and we shall take this into account explicitly by using a prefactor $\Phi(E) = \theta(E_m - E)$ in the traces. Physically, E_m could represent an energy beyond which the system would evaporate or, if we are using effective Hamiltonians, beyond which the effective description is no longer valid. For now, let E_m simply be a parameter which bounds the allowed energy range from above.

Naturally, it is necessary to have E_m greater than the ground state energy E_0 for this kind of restriction to make any sense. To simplify the following discussion we now assume that the Hamiltonian has been normalized so that the ground state energy is equal to zero; this can always be accomplished by replacing $\widehat{H}$ by $\widehat{H} - E_0$. Then we have $E_m > 0$ and we can define the reparametrization by $g(E) = -\ln(1 - E/E_m)$. Applying this to equation (29) shows that the canonical approximation is then given by

$$\langle \widehat{A} \rangle \approx \frac{\operatorname{Tr}\left[\widehat{A}\Phi(\widehat{H})(1 - \widehat{H}/E_m)^\beta\right]}{\operatorname{Tr}\left[\Phi(\widehat{H})(1 - \widehat{H}/E_m)^\beta\right]}. \tag{41}$$

If we now define q and β_q by

$$q = \frac{\beta}{\beta + 1} \qquad \text{and} \qquad \beta_q = \frac{1}{(1-q)E_m} \tag{42}$$

then the right-hand side of (41) becomes equal to the canonical Tsallis expectation values as defined by (40) and axiom 3′. This alone would justify the expression for the Tsallis expectation values as an approximation to the original ensemble. It also gives a connection, via the saddle-point equation, between the parameters of the Tsallis ensemble, q and β_q, and

the characteristics of the original ensemble, E_m, L and λ^2. It does not, however, tell us when this approximation would be better than the usual canonical one. We shall return to this in the following section.

In order to get some feeling of how the saddle-point approximation behaves in this case, we shall now assume that the density of states increases polynomially and in the computation of the traces we shall apply the approximation

$$\operatorname{Tr} f(\widehat{H}) \simeq c \int_0^\infty \mathrm{d}x\, x^{n-1} f(\omega x) \tag{43}$$

where ω defines the relevant energy scale for the increase of the density of states and c is a constant independent of f. This approximation is not quite as arbitrary as it looks, since it can be derived, for example, from the high-energy behaviour of a system of n harmonic oscillators. For energy observables $\widehat{A} = \widehat{H}^k$ equation (43) yields

$$\operatorname{Tr}\left[\widehat{\Phi}\widehat{H}^k \mathrm{e}^{-\beta g(\widehat{H})}\right] \simeq c(E_m/\omega)^n E_m^k \int_0^1 \mathrm{d}y\, y^{n+k-1}(1-y)^\beta \propto E_m^k \frac{\Gamma(n+k)\Gamma(\beta+1)}{\Gamma(n+k+\beta+1)}. \tag{44}$$

Clearly, for all $k \geqslant 0$ we now have $\beta_- = -1$ and thus we need not worry about the applicability of the Gaussian approximation for large-scale observables.

We could now examine the accuracy of the canonical approximation to the Gaussian traces by using the bounds derived in section 4.3. Going through these computations, however, is quite tedious and probably not very interesting, so we shall give a few numerical examples instead.

By the approximation (44), the canonical expectation values are now

$$\langle \widehat{H}^k \rangle_\beta \simeq E_m^k \prod_{j=0}^{k-1} \frac{1}{1+(\beta+1)/(n+j)} \tag{45}$$

where the parameter β is determined from the saddle-point equation involving the digamma function $\psi(x) = \frac{\mathrm{d}}{\mathrm{d}x} \ln \Gamma(x)$,

$$L + \beta\lambda^2 = \psi(n+\beta+1) - \psi(\beta+1). \tag{46}$$

For instance, the expectation value and the variance of energy are then

$$p \equiv \frac{\langle \widehat{H} \rangle}{E_m} = \frac{1}{1+(\beta+1)/n} \qquad \text{and} \qquad \langle \widehat{H}^2 \rangle - \langle \widehat{H} \rangle^2 = \frac{1-p}{n+p} \langle \widehat{H} \rangle^2 \tag{47}$$

and the percentage of fluctuations of the total energy in the canonical ensemble is, therefore, approximately $\sqrt{(1-p)/n}$. If we had used the Boltzmann ensemble here, the total energy fluctuations would have been given by $1/\sqrt{n}$, and thus they would have been larger than in the Tsallis case by a factor of $\sqrt{1-p}$.

We have given a few examples of what kinds of errors the coarse graining and the Gaussian, PSPA and canonical approximations induce in an initially microcanonical fluctuation spectrum in table 1. Of course, the Gaussian approximation is closest to the microcanonical case when $\lambda \ll 1$ and the closer E/E_m is to 1, the larger values of λ can be used before the energy expectation value changes significantly. The canonical Tsallis approximation has the expected behaviour, but the accuracy of the PSPA result is surprisingly good, especially at reproducing the energy fluctuations of the Gaussian ensemble. We also want to point out that β can have negative values here even though this makes the density operator apparently singular.

Table 1. The effect of coarse graining and the computation of expectation values in the different approximations: the first row gives the expectation value of energy, $\langle\widehat{H}\rangle/E_m$, and the second row gives the percentage of energy fluctuations, $\mathrm{var}(\widehat{H})^{1/2}/\langle\widehat{H}\rangle$. The three expectation values refer to equations (26), (28) and (29), respectively.

Parameters							Expectation values		
n	E/E_m	λ	β	β_A	λ/σ_0	q	G_λ	PSPA	Tsallis
5	0.1	0.05	26.3	29.9	0.65	0.96	0.1503	0.1507	0.1548
				33.3			0.2345	0.2329	0.4049
5	0.5	0.5	1.81	2.05	0.93	0.64	0.6398	0.6422	0.6403
				2.26			0.1912	0.1885	0.2525
5	0.99	0.1	−0.66	−0.63	0.03	−1.90	0.9899	0.9843	0.9355
				−0.61			0.0010	0.0336	0.1043
5	0.99	5.0	−0.09	−0.08	3.84	−0.10	0.8450	0.8450	0.8456
				−0.08			0.1603	0.1603	0.1625
100	0.1	0.05	155.7	156.5	1.00	0.99	0.3891	0.3891	0.3895
				157.3			0.0554	0.0554	0.0780
100	0.5	0.05	71.4	72.0	0.56	0.99	0.5806	0.5806	0.5801
				72.5			0.0316	0.0316	0.0646
100	0.99	0.1	0.47	0.48	0.10	0.32	0.9900	0.9901	0.9855
				0.49			0.0010	0.0008	0.0120

8. Towards Tsallis thermodynamics

Consider the following experiment: take a macroscopic piece of material in thermal equilibrium and separate a small sample from it. Let the sample be contained in some finite volume by a finite potential and wait until it again reaches an equilibrium. Now the sample will have lost energy and possibly particles by evaporation from the surface; in particular, there will be no particles with enough kinetic energy to escape the potential barrier in the sample. If the original temperature was high enough, the average energy density would also decrease and one would expect the system to end up in a state with the average energy per particle near the escape potential and with a sharp drop for higher energies. If the number of particles in the final sample is small enough, the broad high-energy tail indicated by the canonical ensemble will give an observable change to the predicted behaviour of the system. The example in section 7 shows that under these conditions the Tsallis statistics with a suitable choice of E_m can offer a better alternative.

There are also instances when the canonical ensemble is ill-defined, but we can still resort to the Tsallis ensemble with its finite energy cut-off. If the density of states increases faster than exponentially, the canonical ensemble cannot be used, whereas the Tsallis ensemble will always produce finite expectation values. Another, more surprising, example is given by systems with a 'Coulomb-like' spectrum: these systems have a point spectrum only below a certain value of E_c, while the spectrum above E_c is continuous or, what is relevant to our case, does not contain any eigenvalues. For these systems any trace-class operator of the form $F(\widehat{H})$ must have a cut-off at the energy E_c—thus a canonical ensemble will not work, but a Tsallis ensemble with $E_m < E_c$ is well defined.

We now wish to examine what happens if the fluctuation spectrum F is essentially microcanonical, i.e. when the total energy fluctuations are small enough that the reparametrization g is essentially linear on the support of F and we can use the saddle-point values from the approximate equation (31). Let E denote the expectation value of the distribution F, when necessarily $E < E_m$, and let $\beta(E)$ be the solution to the approximate

saddle-point equation,

$$\ln(1 - E/E_m) = \frac{\mathrm{Tr}\big[\ln(1 - \widehat{H}/E_m)\theta(E_m - \widehat{H})(1 - \widehat{H}/E_m)^{\beta}\big]}{\mathrm{Tr}\big[\theta(E_m - \widehat{H})(1 - \widehat{H}/E_m)^{\beta}\big]}. \tag{48}$$

The Tsallis parameters q and β_q are then obtained from β and E_m as in the previous section, by equation (42). We shall also adopt the notation

$$\widehat{\omega} = \big(1 - (1-q)\beta_q \widehat{H}\big)^{1/(1-q)}\, \theta(1 - (1-q)\beta_q \widehat{H}) \tag{49}$$

when the escort matrix defined in section 6 is given by $\widehat{\rho} = \widehat{\omega}/\operatorname{Tr}\widehat{\omega}$.

Let us then assume that the canonical approximation is accurate for the energy observable: $E \simeq \operatorname{Tr}(\widehat{H}\widehat{\rho}^{\,q})/\operatorname{Tr}\widehat{\rho}^{\,q}$. In the examples in table 1 this is the case whenever the energy E lies near the cut-off E_m. Using the definition (49), we can then deduce that

$$1 - \frac{E}{E_m} = \frac{\operatorname{Tr}\widehat{\omega}}{\operatorname{Tr}\widehat{\omega}^q}. \tag{50}$$

We showed in section 5 that the PSP entropy is an approximation to the logarithm of the density of states, $\ln W$. Applying the definition of β and of the Tsallis parameters then yields the formula

$$\ln W \approx S_{\mathrm{psp}} = -\frac{q}{1-q}\ln(1 - E/E_m) + \ln \operatorname{Tr}\widehat{\omega}^q = \frac{1}{1-q}\ln\left[\frac{\operatorname{Tr}\widehat{\omega}^q}{(\operatorname{Tr}\widehat{\omega})^q}\right] \tag{51}$$

where in the last equality we have used (50). Thus, for $\widehat{\rho} = \widehat{\omega}/\operatorname{Tr}\widehat{\omega}$ the q-entropy satisfies

$$S_q[\widehat{\rho}] = \frac{1}{1-q}\left(\frac{\operatorname{Tr}\widehat{\omega}^q}{(\operatorname{Tr}\widehat{\omega})^q} - 1\right) \approx \frac{1}{1-q}\big(W^{1-q} - 1\big). \tag{52}$$

This relation between the density of states and the Tsallis entropy was one of the motivations used in the original paper [2] for the definition of S_q.

The other results proven in section 5 apply here as well, in particular, we can conclude that the energy dependence on changes of β while holding E_m fixed satisfies

$$\frac{\partial}{\partial\beta}S_{\mathrm{psp}} = \frac{\beta/E_m}{1 - E/E_m}\frac{\partial}{\partial\beta}E. \tag{53}$$

Interestingly, the Tsallis parameters yield a similar equation if we fix q and let β_q vary. For this, define $\beta_q(E, q)$ as the solution to the equation

$$E = \frac{\operatorname{Tr}(\widehat{H}\widehat{\omega}^q)}{\operatorname{Tr}\widehat{\omega}^q}.$$

Then a straightforward but lengthy computation shows that

$$\frac{\partial}{\partial\beta_q}\frac{\ln(1 + (1-q)S_q)}{1-q} = \frac{\beta_q}{1 - (1-q)\beta_q E}\frac{\partial}{\partial\beta_q}E$$

which by (51) and (52) is essentially the same equation as in (53).

These results clarify the difference between the use of the Tsallis parameters and the PSP parameters. In the first case, the increase of the fluctuation spectrum, defined by q, is a known quantity, while the boundary conditions determine the position of the energy cut-off, defined by β_q. In the PSP approximation, it is more natural to vary the energy of the system while holding the energy cut-off fixed. Otherwise, these two methods are very similar, as both have a simple formula for the change of energy and a maximum entropy principle which can be applied for a determination of effective Hamiltonians, as explained in section 5.

In real experiments performed on a sample of matter in a container, one usually examines the properties of the sample matter by altering some easily controllable physical conditions of the environment. For example, if the temperature of the container is the varying quantity, then a natural parameter for the experiment would be the energy density of the sample matter at the boundary. Unfortunately, unless the state of the system is homogeneous, the energy density on the boundary is not determined trivially from the total energy given by the Hamiltonian. It is, of course, possible that the boundary conditions can be included naturally to the dynamics by using effective Hamiltonians and this would then determine the relations between the parameters of the experiment and the parameters of the PSP approximation.

There has already been a lot of discussion of the physical meaning of the Tsallis parameter q in the literature. For example, for systems exhibiting multifractal behaviour, the value of q could be quite straightforwardly determined as was shown in [13]. The value of q can also be related to the degree of non-extensivity of energy and the speed of growth of certain classical long-range potentials [14].

9. Speculations concerning the maximum entropy principle

As mentioned earlier, the Tsallis ensemble is usually derived by postulating the Tsallis entropy functional (39) and then finding its maximum under the condition that the energy expectation value is fixed. This approach is motivated by the standard derivation of the canonical ensemble by a similar maximization of the Gibbs entropy functional.

The motivation behind these derivations lies in the intuitive association of the maximum likelihood ensemble with the least restricted, maximum disorder, ensemble. Since for systems with infinitely many degrees of freedom the natural choice of associating the maximum likelihood with equiprobability is not possible, it is then necessary to postulate an entropy functional which quantitatively measures the disorder of any density operator. The Gibbs entropy functional $S[\widehat{\rho}] = \mathrm{Tr}(-\widehat{\rho}\ln\widehat{\rho})$ can be derived by requiring the entropy to be additive for independent systems. The additivity of the thermodynamical entropy, on the other hand, is closely related to the extensivity of thermodynamical systems (see, for example, the axiomatic derivation of Lieb and Yngvason [15] which clearly highlights the role of extensivity). Since the Tsallis entropy is only sub- or superadditive [8], by this reasoning it would be useful only for systems which are non-extensive.

In this work we have adopted a different implementation of the maximum disorder principle: by coarse graining of energy we try to find a simple *parametrization* of the underlying ensemble which would have all the same relevant predetermined properties as the original ensemble. The coarse graining will naturally increase the number of energy states which participate in the ensemble and thus we are, in a way, always increasing the microcanonical entropy of the ensemble.

We now claim that in spite of the similarities, the present approach is conceptually easier and more readily applicable to a wider variety of situations. First, the possible fluctuations of the parameters can also be properly taken into account in this formulation and the effect of these fluctuations can be analysed. Secondly, although the energy reparametrization g, which is responsible for the appearance of non-standard ensembles, plays the same role as the different entropy functionals in the maximum entropy derivations, its interpretation is more tangible and, consequently, choosing between different non-standard ensembles should be easier.

We would also like to emphasize the role of entropy differently here. Instead of a fundamental, philosophical role the entropy has been endowed, we would like to emphasize its *practical* importance both as a function encoding the dependence of energy on the temperature

parameter and as an effective potential which enables one to choose parameters in effective descriptions of the system.

As we have shown, our definition of PSP entropy, equation (34), leads to an extension of both of these properties. The relation between the changes of total energy and PSP entropy are given in (35), from which the standard case follows by setting $g = \mathrm{id}$. The maximization of the statistical PSP entropy was proven in theorem 2 and its corollary provides a way of choosing between effective Hamiltonians. The Tsallis axioms also offer another, slightly different, way of applying the maximization of entropy which leads to one of the canonical PSP ensembles.

10. Conclusions

We have presented a generalized coarse-graining procedure which can be used for analysing a wide variety of quantum systems. The main purpose of the coarse graining was to develop simple parametrizations which could be used for analysing the behaviour of the system when its energy is varied. We restricted to the energy only, but the method can be easily extended to the case when there are many relevant conserved quantities.

One important application was to systems for which the Boltzmann ensemble either does not converge, or produces too large total energy fluctuations. We have shown that in this case the Tsallis statistics offers a generalization which is more likely to work, but which still retains some of the useful properties of classical thermodynamics: maximization of an entropy functional and a simple formula for the differential change of energy. The cost in this case is the inclusion of an extra parameter which needs to be chosen in a suitable way which is dependent on the behaviour of the system—the traces involved in the computations are also likely to be prohibitively difficult to compute.

In the final section we have raised some objections to the use of the maximum entropy principle as a fundamental justification of the canonical Boltzmann ensemble. In particular, for purposes of physics teaching, it has the disadvantage that it either leads to the idea that the canonical ensemble is the only sensible statistical description of an arbitrary physical system or that thermodynamics can only be applied in the thermodynamical limit in which all the standard ensembles agree. We consider both of these ideas to be misguidingly restrictive.

The general guidelines for choosing an ensemble could be the following: the number of parameters should be as few as possible and have as direct as possible relation to the physical characteristics of the system. In addition, the traces needed for the evaluation of the ensemble should be computable for the given Hamiltonian and, naturally, the ensemble must reproduce the original ensemble within the required accuracy. Of the coarse-grained ensembles we considered, it should always be possible to find a Gaussian ensemble which is accurate enough. The Gaussian expectation values, however, are usually difficult to evaluate, while the corresponding canonical ensembles are often easier in this respect. Of the canonical ensembles, the Boltzmann ensemble is the simplest but, when it cannot be used, the canonical Tsallis ensemble is a good second trial.

There is, however, one more possibility we have not stressed yet. From table 1 we can see that the PSP approximation of expectation values, equation (28), can give a significantly better approximation to the Gaussian expectation values than the canonical ensemble. On the other hand, for the simple case $g(x) = x$, everything needed in the PSP approximation formula is given by expectation values in the usual canonical ensemble. Using the PSP approximation would thus enable the approximation of Gaussian expectation values without having to compute anything but canonical expectation values—this might be especially useful for small samples of macroscopic matter for which the canonical expectation values are computable, but not accurate.

Acknowledgments

The author warmly thanks C Tsallis for his help and wishes to express gratitude to C Tsallis and I Procaccia for a stimulating discussion. I am also indebted to A Kupiainen for valuable discussions.

Appendix A. Proof of the Gaussian approximation theorem

We shall begin the proof of theorem 1 from equation (17) which was shown in section 4.2 to follow from assumptions B1 and B2. In other words, we inspect the integral representation

$$\mathrm{Tr}\big(\widehat{A}F_\Lambda(\widehat{H})\big) = \int_{\beta-\mathrm{i}\infty}^{\beta+\mathrm{i}\infty} \frac{\mathrm{d}w}{2\pi\mathrm{i}} \mathrm{e}^{\frac{1}{2}\Lambda^2 w^2} \bar{F}(w)\, \mathrm{Tr}\big(\widehat{A}\mathrm{e}^{-wg(\widehat{H})}\big) \tag{A1}$$

which is valid for all $\beta_- < \beta < \beta_+$ and $\Lambda \geqslant 0$ with an integrand analytic in the strip $\beta_- < \mathrm{Re}\, w < \beta_+$. The proof will depend on the properties of the real saddle-point approximation in the large-Λ limit so we shall now first examine the large-Λ dependence of the real saddle-point value.

By differentiation of the logarithm of the integrand we arrive at the saddle-point equation

$$\Lambda^2 w + a(w) - \bar{g}(w) = 0 \tag{A2}$$

where the functions a and g are defined in (13). A differentiation of the left-hand side yields $\Lambda^2 + b^2(w) + \sigma^2(w)$, which is clearly strictly positive for real w. Thus the restriction of the left-hand side to the real axis is strictly increasing and we have a real saddle point β on the interval $\beta_- < \beta < \beta_+$ if and only if

$$\Lambda^2\beta_- + a(\beta_-) < \bar{g}(\beta_-) \qquad \text{and} \qquad \Lambda^2\beta_+ + a(\beta_+) > \bar{g}(\beta_+). \tag{A3}$$

Since $\beta_+ > 0$, the second inequality is satisfied as soon as Λ becomes large enough. If $\beta_- < 0$, then the same holds for the first inequality as well, but if $\beta_- = 0$, then we need to have $a(0) < \bar{g}(0)$ which is equivalent to condition G1 of section 4.2.

Let us now assume that either of these conditions hold and let $\Lambda_c \geqslant 0$ be the infimum of values for which (A3) is valid. Then for all $\Lambda > \Lambda_c$ there is a unique real saddle point which we denote by $\beta(\Lambda)$. As the left-hand side of (A2) is strictly increasing, we can also deduce that $\beta(\Lambda)$ never changes sign:

(a) if $a(0) < \bar{g}(0)$, then $\beta(\Lambda) > 0$,
(b) if $a(0) = \bar{g}(0)$, then $\beta(\Lambda) = 0$,
(c) if $a(0) > \bar{g}(0)$, then $\beta(\Lambda) < 0$

for all $\Lambda > \Lambda_c$.

An application of the implicit function theorem to the defining equation (A2) proves that $\beta(\Lambda)$ is a smooth function from (Λ_c, ∞) to (β_-, β_+). A differentiation of the defining equation then shows that

$$\beta'(\Lambda) = -\frac{2\beta\Lambda}{\Lambda^2 + b^2(\beta) + \sigma^2(\beta)}$$

from which we can deduce that β is either a monotonically decreasing positive ($a(0) < \bar{g}(0)$), zero ($a(0) = \bar{g}(0)$) or monotonically increasing negative ($a(0) > \bar{g}(0)$) function of Λ. In all three cases, the limit $\beta_\infty = \lim_{\Lambda\to\infty} \beta(\Lambda)$ exists and belongs to the interval (β_-, β_+) if $\beta_- < 0$, and to the interval $[0, \beta_+)$ if $\beta_- = 0$.

We will next show that $\beta_\infty = 0$. Assume, to obtain a contradiction, that $\beta_\infty \neq 0$. By the above results, then β_∞ always belongs to the interval (β_-, β_+) and thus both $\lim_{\Lambda\to\infty} \bar{g}(\beta(\Lambda))$ and $\lim_{\Lambda\to\infty} a(\beta(\Lambda))$ exist and are finite. However, taking the limit $\Lambda \to \infty$ in the defining equation (A2) then shows that we must have $\beta(\Lambda) \to 0$, which is a contradiction.

Thus we have shown that the saddle-point solution goes to zero smoothly and monotonically as the scale Λ is increased. Since we are only interested in the large-scale behaviour, we shall next expand $\ln \bar{F}$ in the neighbourhood of the origin; this is possible since $\bar{F}$ is analytic and $\bar{F}(0) \neq 0$. The expansion yields

$$\bar{F}(w) = \bar{F}(0)\,\mathrm{e}^{Lw+\frac{1}{2}\lambda^2 w^2}(1 + \Delta\bar{F}(w)) \tag{A4}$$

where $L = a(0)$, $b^2 = \lambda^2(0)$ and the correction term $\Delta\bar{F}$ satisfies

$$\lim_{w\to 0} \frac{\Delta\bar{F}(w)}{w^k} < \infty \tag{A5}$$

for $k = 3$.

Let us then consider the expansion (A4) when L and λ^2 are not necessarily equal to the default values $a(0)$ and $b^2(0)$. It is easy to deduce that if $\lambda^2 \neq b^2(0)$, equation (A5) is true only for $k \leqslant 2$ and if we have $L \neq a(0)$, then (A5) holds only for $k \leqslant 1$. However, as is evident from the definition (10), $|\bar{F}(w)| \leqslant \bar{F}(\mathrm{Re}\, w)$ always, and we find from the definition (A4) a bound

$$|\Delta\bar{F}(\beta + \mathrm{i}\alpha)|\,\mathrm{e}^{-\frac{1}{2}\alpha^2\lambda^2} \leqslant 1 + \frac{\bar{F}(\beta)}{\bar{F}(0)\,\mathrm{e}^{\beta L+\frac{1}{2}\beta^2\lambda^2}} \tag{A6}$$

which is valid for all $\beta < \beta_+$ and all real α, L and λ^2.

For the proof of the theorem, we need to consider the difference

$$\begin{aligned}\Delta_G &\equiv \mathrm{Tr}\big(\widehat{A}F_\Lambda(\widehat{H})\big) - \bar{F}(0)\,\mathrm{Tr}\big[\widehat{A}G_{\sqrt{\lambda^2+\Lambda^2}}(L - g(\widehat{H}))\big] \\ &= \int_{\beta-\mathrm{i}\infty}^{\beta+\mathrm{i}\infty} \frac{\mathrm{d}w}{2\pi\,\mathrm{i}}\,\mathrm{Tr}\big(\widehat{A}\mathrm{e}^{-wg(\widehat{H})}\big)\bar{F}(0)\,\mathrm{e}^{Lw+\frac{1}{2}(\lambda^2+\Lambda^2)w^2}\Delta\bar{F}(w)\end{aligned} \tag{A7}$$

where we have used the g-transform of the Gaussian trace, the integral representation (A1), and defined the function $\Delta\bar{F}(w)$ by equation (A4). For each $\Lambda \geqslant 0$, let us parametrize the integration variable in (A7) as $w = \beta + \mathrm{i}\alpha$. To prove the convergence properties stated in the theorem, we need to choose $\beta(\Lambda)$ so that it decays like a negative power of Λ: we assume that $c \geqslant 0$ and that β has been chosen so that

$$\beta^{1+c}\Lambda \text{ stays bounded in the limit } \Lambda \to \infty. \tag{A8}$$

In particular, this means that $\beta \to 0$ when $\Lambda \to \infty$.

The absolute value of the correction term then has an upper bound

$$|\Delta_G| \leqslant \bar{F}(0)\,\mathrm{e}^{\beta L+\frac{1}{2}\beta^2(\lambda^2+\Lambda^2)}\,\mathrm{Tr}\left[\widehat{A}\mathrm{e}^{-\beta g(\widehat{H})}\right]\int_{-\infty}^{\infty}\frac{\mathrm{d}\alpha}{2\pi}\mathrm{e}^{-\frac{1}{2}(\lambda^2+\Lambda^2)\alpha^2}|\Delta\bar{F}(\beta + \mathrm{i}\alpha)| \tag{A9}$$

and we shall inspect the behaviour of the remaining integral next.

Let k be such that (A5) is true; as was mentioned earlier, we can always choose at least any $k \leqslant 1$. Then there are constants $M > 0$ and $m > 0$ such that for all $|w| \leqslant 2m$,

$$|\Delta\bar{F}(w)| \leqslant M|w|^k.$$

Also assume that Λ is so large that $|\beta| < m$. Then for all $|\alpha| < m$,

$$\Lambda^{k/(1+c)}|\Delta\bar{F}(\beta + \mathrm{i}\alpha)| \leqslant M|\beta\Lambda^{1/(1+c)} + \mathrm{i}(\alpha\Lambda)\Lambda^{-c/(1+c)}|^k \leqslant M(|\beta|\Lambda^{1/(1+c)} + |\alpha|\Lambda)^k$$

which implies that

$$\Lambda^{k/(1+c)}\sqrt{2\pi(\lambda^2+\Lambda^2)}\int_{-m}^{m}\frac{\mathrm{d}\alpha}{2\pi}\mathrm{e}^{-\frac{1}{2}(\lambda^2+\Lambda^2)\alpha^2}|\Delta\bar{F}(\beta+\mathrm{i}\alpha)|$$

stays bounded in the limit $\Lambda\to\infty$. In fact, the same is then also true when the integration limits are replaced by $\pm\infty$, since the remaining integral over the values $|\alpha|\geqslant m$ does not contribute at all to the limit $\Lambda\to\infty$; to see this, apply the dominated convergence theorem to the remainder.

By equation (A9) this proves that

$$\frac{|\Delta_G|}{\bar{F}(0)\,\mathrm{Tr}\bigl[\widehat{A}G_{\sqrt{\lambda^2+\Lambda^2}}(L-g(\widehat{H}))\bigr]}\leqslant\frac{\mathrm{e}^{\beta L+\frac{1}{2}\beta^2(\lambda^2+\Lambda^2)}\,\mathrm{Tr}\bigl[\widehat{A}\mathrm{e}^{-\beta g(\widehat{H})}\bigr]}{\mathrm{Tr}\bigl\{\widehat{A}\exp\bigl[-\frac{1}{2(\lambda^2+\Lambda^2)}(L-g(\widehat{H}))^2\bigr]\bigr\}}\mathrm{O}(\Lambda^{-k/(1+c)})\quad\text{(A10)}$$

where k and c are any values allowed by (A5) and (A8), respectively. In the final part of the proof we derive conditions when the remaining multiplicative term stays bounded for large Λ.

First, define an auxiliary variable x by the equation $x=L+\beta(\lambda^2+\Lambda^2)$. Applying this in (A10), reveals that the multiplicative term is the inverse of

$$\frac{\mathrm{Tr}\bigl[\widehat{A}\mathrm{e}^{-\beta g(\widehat{H})}\mathrm{e}^{-\frac{1}{2(\lambda^2+\Lambda^2)}(x-g(\widehat{H}))^2}\bigr]}{\mathrm{Tr}\bigl[\widehat{A}\mathrm{e}^{-\beta g(\widehat{H})}\bigr]}.\quad\text{(A11)}$$

Since (A11) is always less than one, it cannot improve the convergence of the right-hand side in (A10). Thus it is enough to show that it does not spoil the convergence for the optimal choice $L=a(0)$.

Since $\widehat{A}$ is positive, Jensen's inequality can be applied in (A11) and we obtain that the logarithm of (A11) is always greater than or equal to

$$-\frac{1}{2(\lambda^2+\Lambda^2)}\frac{\mathrm{Tr}\bigl[\widehat{A}\mathrm{e}^{-\beta g(\widehat{H})}(x-g(\widehat{H}))^2\bigr]}{\mathrm{Tr}\bigl[\widehat{A}\mathrm{e}^{-\beta g(\widehat{H})}\bigr]}=-\frac{(x-\bar{g}(\beta))^2}{2(\lambda^2+\Lambda^2)}-\frac{\sigma^2(\beta)}{2(\lambda^2+\Lambda^2)}.\quad\text{(A12)}$$

Both of the terms in (A12) are negative and thus they have to be bounded separately for the boundedness of the multiplicative term in (A10).

The first term is small only if β satisfies

$$|L+\beta(\lambda^2+\Lambda^2)-\bar{g}(\beta)|\lesssim\sqrt{\lambda^2+\Lambda^2}.\quad\text{(A13)}$$

If we multiply this by β^{1+2c} and take the limit $\Lambda\to\infty$, it becomes evident that the conditions (A13) and (A8) can be compatible only if $\beta^{1+2c}\bar{g}(\beta)$ stays bounded in the limit $\beta\to 0$, i.e. only if G2 holds with the same value of c.

The converse is also true in the sense that if $c\geqslant 0$ is such that G2 holds, then we can find $\beta(\Lambda)$ which satisfies both (A13) and (A8). Such a β is defined by the equation

$$L+\beta(\lambda^2+\Lambda^2)-\bar{g}(\beta)=0\quad\text{(A14)}$$

as we shall now show. A brief reflection shows that for a Gaussian fluctuation spectrum $F(\widehat{H})=G_{\sqrt{\lambda^2+\Lambda^2}}(L-g(\widehat{H}))$, we would have $a(w)=L+w(\lambda^2+\Lambda^2)$ and thus (A14) is the saddle-point equation for a Gaussian fluctuation spectrum. By the results proven at the beginning of this section, we then know that if $\beta_-<0$ or $L<\bar{g}(0)$, equation (A14) defines for all large enough Λ a smooth function $\beta(\Lambda)$, which monotonically approaches zero as Λ increases. Such $\beta(\Lambda)$ trivially satisfies (A13) and by multiplying (A14) by β^{1+2c} we can see that G2 then also implies that (A8) is true for the same value of c.

Since it is part of the assumptions so far that either $\beta_-<0$ or $a(0)<\bar{g}(0)$, we have thus shown that if G2 holds, then—at least for the optimal choice $L=a(0)$—it is possible to define

$\beta(\Lambda)$ by (A14) and this function satisfies the decay condition (A8). Also note that if $\beta_- < 0$, then $\bar{g}(0) < \infty$ and G2 is trivially satisfied for the best allowed choice $c = 0$.

For the boundedness of the second term in (A12), we shall need either $\beta_- < 0$ or assume the last of the three conditions, G3. If $\beta_- < 0$, then $\sigma^2(0) < \infty$ and the second term vanishes in the limit $\Lambda \to \infty$. If $\beta_- = 0$ and $L < \bar{g}(0)$, we can define β by equation (A14) when

$$\frac{\sigma^2(\beta)}{\lambda^2 + \Lambda^2} = \frac{\beta\sigma^2(\beta)}{\bar{g}(\beta) - L}$$

and clearly it is then sufficient that $\beta\sigma^2(\beta)/\bar{g}(\beta)$ stays bounded in the limit $\beta \to 0^+$. Note that this proves the sufficiency of condition G3 for the relevant case $L = a(0)$.

Combining the results we then find from equations (A7) and (A10) that if either $\beta_- < 0$ or G1–G3 hold, then for $L = a(0)$ and $\lambda^2 = b^2(0)$ we have

$$\mathrm{Tr}\big(\widehat{A}F_\Lambda(\widehat{H})\big) = \bar{F}(0)\,\mathrm{Tr}\left[\widehat{A}G_{\sqrt{\lambda^2+\Lambda^2}}(L - g(\widehat{H}))\right]\left(1 + \mathrm{O}(\Lambda^{-k/(1+c)})\right)$$

with $k = 3$. We have also seen that changing λ would allow using only $k = 2$ and changing L only $k = 1$ or nothing at all—if $\beta_- = 0$ and $L > \bar{g}(0)$, there are no solutions to equation (A14) and the logarithmic bound we have in (A12) becomes non-conclusive. This gives a precise meaning for the statement at the end of the theorem and completes the proof.

Appendix B. Derivation of the bounds for the PSP approximation of a Gaussian trace

In this section we first derive the 'saddle-point approximation' (22) for the Gaussian trace and then show how the two bounds for the accuracy of this approximation, equations (23) and (25), can be obtained. As in section 4.3, let L and λ be some parameters which define a Gaussian fluctuation spectrum. Then $\beta_+ = \infty$ and we shall inspect the 'canonical' approximation of the Gaussian trace for any $\beta > \beta_-$. Let us also define $R = \lambda/\sigma(\beta)$.

By equation (12) we can then use the integral representation,

$$\mathrm{Tr}\left[\widehat{A}G_\lambda(L - g(\widehat{H}))\right] = \mathrm{e}^{\beta L + \frac{1}{2}\lambda^2\beta^2}\int_{-\infty}^{\infty}\frac{\mathrm{d}\alpha}{2\pi}\mathrm{e}^{-\frac{1}{2}\lambda^2\alpha^2}\,\mathrm{Tr}\big(\widehat{A}\mathrm{e}^{-\beta g(\widehat{H})}\mathrm{e}^{\mathrm{i}\alpha(L+\beta\lambda^2-g(\widehat{H}))}\big). \tag{B1}$$

The PSP approximation is now derived by expanding $\ln f(\beta + \mathrm{i}\alpha)$ around $\alpha = 0$, where $f(w) = \mathrm{Tr}\big(\widehat{A}\mathrm{e}^{-wg(\widehat{H})}\big)$.

$f(w)$ is an analytic function in the region $\mathrm{Re}\, w > \beta_-$ whose restriction to the line $\mathrm{Im}\, w = 0$ is a strictly positive function. Although the logarithm of f is possibly not well defined in the whole region, the following representation is nevertheless valid for all $\beta > \beta_-$ and $\alpha \in \mathbb{R}$:

$$f(\beta + \mathrm{i}\alpha) = f(\beta)\,\mathrm{e}^{-\mathrm{i}\alpha\bar{g}(\beta)}\exp\left[-\alpha^2\int_0^1 \mathrm{d}t\,(1-t)\sigma^2(\beta + \mathrm{i}\alpha t)\right] \tag{B2}$$

where $-\bar{g}(w)$ and $\sigma^2(w)$, as defined by (13), are the first and second derivative of $\ln f$. If there is a singularity on the integration contour in (B2), the integral has to be evaluated by an infinitesimal deformation of the contour.

Since σ^2 is analytic and $\sigma^2(\beta) > 0$, we can for all $0 < r < 1$ find a $\rho > 0$ such that

$$\left|\frac{\sigma^2(\beta + \mathrm{i}\alpha)}{\sigma^2(\beta)} - 1\right| \leqslant r \qquad \text{for all} \quad -\rho \leqslant \alpha \leqslant \rho. \tag{B3}$$

Then let $|\alpha| \leqslant \rho$ and define z by $z = \frac{1}{2}(r-1)\sigma^2(\beta) + \int_0^1 \mathrm{d}t\,(1-t)\sigma^2(\beta + \mathrm{i}\alpha t)$. By (B3), now $\mathrm{Re}\, z \geqslant 0$ and $|z| \leqslant r\sigma^2(\beta)$. However, since for every complex w with $\mathrm{Re}\, w \geqslant 0$ it is true that $|\mathrm{e}^{-w} - 1| \leqslant |w|$, we also have the inequality

$$\left|\mathrm{e}^{-\alpha^2 z} - \mathrm{e}^{-\frac{1}{2}r\alpha^2\sigma^2(\beta)}\right| \leqslant \left|\mathrm{e}^{-\alpha^2 z} - 1\right| + \left|\mathrm{e}^{-\frac{1}{2}r\alpha^2\sigma^2(\beta)} - 1\right| \leqslant \tfrac{3}{2}r\alpha^2\sigma^2(\beta).$$

A comparison with (B2) then shows that we have proven the following bound for the quadratic approximation of f:

$$\left| \frac{f(\beta+\mathrm{i}\alpha)}{f(\beta)\,\mathrm{e}^{-\mathrm{i}\alpha\bar{g}(\beta)-\frac{1}{2}\alpha^2\sigma^2(\beta)}} - 1 \right| \leqslant \frac{3}{2} r\alpha^2\sigma^2(\beta) \exp\left(\frac{1}{2}r\alpha^2\sigma^2(\beta)\right) \tag{B4}$$

valid for all $|\alpha| \leqslant \rho$. Since always $|f(\beta+\mathrm{i}\alpha)| \leqslant f(\beta)$, it is also obvious that for all real α the left-hand side is bounded by

$$\exp\left(\frac{1}{2}\alpha^2\sigma^2(\beta)\right) + 1. \tag{B5}$$

Applying the quadratic expansion of f to (B1) we obtain

$$\mathrm{Tr}\left[\widehat{A}G_\lambda(L-g(\widehat{H}))\right] = \mathrm{e}^{\beta L+\frac{1}{2}\lambda^2\beta^2}\,\mathrm{Tr}\left[\widehat{A}\mathrm{e}^{-\beta g(\widehat{H})}\right] G_{\sqrt{\lambda^2+\sigma^2(\beta)}}(L+\beta\lambda^2-\bar{g}(\beta))\,(1+\Delta_{\mathrm{PSPA}}) \tag{B6}$$

where Δ_{PSPA} is defined by

$$G_{\sqrt{\lambda^2+\sigma^2(\beta)}}(L+\beta\lambda^2-\bar{g}(\beta))^{-1} \int_{-\infty}^{\infty} \frac{\mathrm{d}\alpha}{2\pi} \mathrm{e}^{-\frac{1}{2}\lambda^2\alpha^2+\mathrm{i}\alpha(L+\beta\lambda^2)} \left[\frac{f(\beta+\mathrm{i}\alpha)}{f(\beta)} - \mathrm{e}^{-\mathrm{i}\alpha\bar{g}(\beta)-\frac{1}{2}\alpha^2\sigma^2(\beta)} \right].$$

By (B4) and (B5) then

$$\begin{aligned} &G_{\sqrt{\lambda^2+\sigma^2}}(L+\beta\lambda^2-\bar{g})|\Delta_{\mathrm{PSPA}}| \\ &\qquad \leqslant \frac{3}{2}r\sigma^2 \int_{-\rho}^{\rho} \frac{\mathrm{d}\alpha}{2\pi}\,\alpha^2 \mathrm{e}^{-\frac{1}{2}\alpha^2(\lambda^2+(1-r)\sigma^2)} + \int_{|\alpha|\geqslant\rho} \frac{\mathrm{d}\alpha}{2\pi} \left(\mathrm{e}^{-\frac{1}{2}\alpha^2\lambda^2} + \mathrm{e}^{-\frac{1}{2}\alpha^2(\lambda^2+\sigma^2)}\right) \\ &\qquad \leqslant \frac{1}{\sqrt{2\pi}} \left[\frac{3}{2}\frac{r}{\sigma} \big(R^2+1-r\big)^{-\frac{3}{2}} + \frac{\mathrm{e}^{-\frac{1}{2}\rho^2\lambda^2}}{\lambda} + \frac{\mathrm{e}^{-\frac{1}{2}\rho^2(\lambda^2+\sigma^2)}}{\sqrt{\lambda^2+\sigma^2}} \right] \end{aligned} \tag{B7}$$

where we have extended the first integral over the whole real line and used

$$\int \mathrm{d}x\, x^2 \exp\left[-\tfrac{1}{2}x^2\right] = \sqrt{2\pi}$$

and approximated the second integral as in

$$\int_{|\alpha|\geqslant\rho} \frac{\mathrm{d}\alpha}{2\pi}\,\mathrm{e}^{-\frac{1}{2}\alpha^2\lambda^2} = 2\int_0^{\infty} \frac{\mathrm{d}y}{2\pi}\,\mathrm{e}^{-\frac{1}{2}(y+\rho)^2\lambda^2} \leqslant \mathrm{e}^{-\frac{1}{2}\rho^2\lambda^2} \int_{-\infty}^{\infty} \frac{\mathrm{d}y}{2\pi}\,\mathrm{e}^{-\frac{1}{2}y^2\lambda^2}.$$

The second bound (25) follows now easily from (B7).

The first bounds, inequalities (23), are a straightforward consequence of the result

$$\exp\left[-\frac{(L+\beta\lambda^2-\bar{g}(\beta))^2}{2(\lambda^2+\Lambda^2)} - \frac{\sigma^2(\beta)}{2(\lambda^2+\Lambda^2)}\right] \leqslant \frac{\mathrm{Tr}\left\{\widehat{A}\exp\left[-\frac{1}{2(\lambda^2+\Lambda^2)}(L-g(\widehat{H}))^2\right]\right\}}{\mathrm{e}^{\beta L+\frac{1}{2}\beta^2(\lambda^2+\Lambda^2)}\,\mathrm{Tr}\left[\widehat{A}\mathrm{e}^{-\beta g(\widehat{H})}\right]} \leqslant 1$$

which was derived by using Jensen's inequality near equation (A11) in appendix A.

References

[1] Lukkarinen J 1999 *J. Phys. A: Math. Gen.* **32** 287
[2] Tsallis C 1988 *J. Stat. Phys.* **52** 479
[3] Balian R 1991 *From Microphysics to Macrophysics* vol I (Berlin: Springer)
[4] Rudin W 1974 *Functional Analysis* (New Delhi: Tata McGraw-Hill)
[5] Plastino A R and Plastino A 1993 *Phys. Lett.* A **174** 384
[6] Tsallis C, Levy S V F, Souza A M C and Maynard R 1995 *Phys. Rev. Lett.* **75** 3589
[7] Zanette D H and Alemany P A 1995 *Phys. Rev. Lett.* **75** 366
[8] Boghosian B M 1996 *Phys. Rev.* E **53** 4754
[9] Huang X-P and Driscoll C F 1994 *Phys. Rev. Lett.* **72** 2187
[10] Guerberoff G R and Raggio G A 1996 *J. Math. Phys.* **37** 1776
[11] Czachor M and Naudts J 1999 *Phys. Rev.* E **59** 2497
[12] Tsallis C, Mendes R S and Plastino A R 1998 *Physica* A **261** 534
[13] Lyra M L and Tsallis C 1998 *Phys. Rev. Lett.* **80** 53
[14] Jund P, Kim S G and Tsallis C 1995 *Phys. Rev.* B **52** 50
[15] Lieb E H and Yngvason J 1999 *Phys. Rep.* **310** 1