THE

PHYSICAL REVIEW

A journal of experimental and theoretical physics established by E. L. Nichols in 1893

SECOND SERIES, VOL. 133, No. 5A — 2 MARCH 1964

Simple Variational Bound to the Entropy

G. H. DERRICK

Applied Mathematics Department, The University of New South Wales, Kensington, N. S. W., Australia

(Received 21 October 1963)

The following variational principle is obtained for the entropy $S(E)$ of a system with energy E: $S(E)\geqq -k\ln(\mathrm{Trace}U^2)$ for all non-negative Hermitian density matrices U with Trace $U=1$, Trace $HU=E$; H is the Hamiltonian and k is Boltzmann's constant. The equality sign is realized with this principle for the density matrix of the microcanonical ensemble, as well as for a wide class of similar ensembles (in the limit of large volume).

I. INTRODUCTION

THE object of this note is to give a simple variational bound to the entropy $S(E)$ at energy E:

$$S(E)\geqq -k\ln(\mathrm{Trace}\ U^2) \tag{1}$$

for all Hermitian density matrices U (with no negative eigenvalues) for which Trace $U=1$ and Trace $HU=E$; H is the Hamiltonian and k is Boltzmann's constant. The principle (1) has the advantage that U^2 is in general much easier to evaluate than $U\ln U$ which appears in the conventional bound given by von Neumann[1]:

$$S(E)\geqq -k\ \mathrm{Trace}\ U\ln U. \tag{2}$$

The optimum density matrix U_0 for which the equality sign in (1) is realized is

$$U_0=(\beta/e^2)e^{-S(E)/k}\int_{\lambda<E+2/\beta}(E+2/\beta-\lambda)d\hat{E}(\lambda)\,, \tag{3}$$

where $k\beta=dS(E)/dE$ is the reciprocal temperature and $H=\int\lambda d\hat{E}(\lambda)$ is the spectral resolution of the Hamiltonian[1]; if H has a discrete spectrum of eigenvalues E_i and eigenstates $|i\rangle$, then $\hat{E}(\lambda)=\sum_{E_i\leqq\lambda}|i\rangle\langle i|$. The precise value of the normalization constant in (3) depends on the definition adopted for the entropy; the definition used here is

$$e^{S(E)/k}=\mathrm{Trace}\ \hat{E}(E) \tag{4}$$
$$=\text{number of eigenvalues of } H\leqq E.$$

[1] J. von Neumann, *Mathematische Grundlagen der Quantenmechanik* (Springer-Verlag, Berlin/Vienna, 1932); [English transl.: by R. T. Beyer (Princeton University Press, Princeton, 1955)].

It should be noted that while U_0 gives the optimum bound in (1), a wide class of density matrices U actually gives the equality sign apart from terms of relative order $1/N$, where N is the number of particles. For example, the microcanonical ensemble matrix

$$U_m=e^{-S(E)/k}[\hat{E}(E)-\hat{E}(E_1)] \tag{5}$$

with $E_1<E$ gives the correct entropy $S(E)$ when substituted in (1).

PROOF OF (1).

To prove (1) we start from (2) and write $U=V/\mathrm{Trace}\ V$ where V is non-negative, Hermitian, and satisfies Trace $(H-E)V=0$, but is not required to be normalized to unit trace. This gives

$$S(E)\geqq k\ln(\mathrm{Trace}\ V)-k(\mathrm{Trace}\ V\ln V)/\mathrm{Trace}\ V. \tag{6}$$

Now if x is a non-negative real number we have from elementary algebra $-x\ln x\geqq x-x^2$. Since all the eigenvalues of V are non-negative by assumption, it follows that $-\mathrm{Trace}\ V\ln V\geqq\mathrm{Trace}\ (V-V^2)$; hence

$$S(E)\geqq k\ln(\mathrm{Trace}\ V)+k(\mathrm{Trace}[V-V^2])/\mathrm{Trace}\ V. \tag{7}$$

Let us now replace V by λV where λ is a real number, and maximize the right-hand side of (7) with respect to λ for a given V. This yields, for the optimum λ,

$$\lambda=(\mathrm{Trace}\ V)/\mathrm{Trace}\ V^2\quad\text{or}\quad\mathrm{Trace}(\lambda V-\lambda^2V^2)=0.$$

Hence the density matrix V_0 which optimizes the bound in (7) must satisfy $\mathrm{Trace}(V_0-V_0^2)=0$. Thus the principle $S(E)\geqq k\ln(\mathrm{Trace}\ V_1)$, where V_1 is further restricted by Trace $(V_1-V_1^2)=0$, will yield the same optimum bound as (7). If we now write $V_1=U/\text{Trace } U^2$ where Trace $U=1$, which automatically satisfiies the condition Trace $(V_1-V_1^2)=0$, then we obtain the principle (1).

At first sight it might appear unlikely that the equality sign could ever be realized in (1), since the equality $-x\ln x=x-x^2$ holds only for $x=0$ and $x=1$. However, with the microcanonical ensemble matrix U_m given by (5), all the eigenvalues of $U_m/\text{Trace } U_m^2$ are in fact equal to 0 or 1, so that U_m does indeed give the correct entropy. [On the other hand, the canonical ensemble matrix, $U=e^{-\beta H}/\text{Trace } e^{-\beta H}$, which gives the correct entropy in (2), fails to do so when substituted into (1).]

THE OPTIMUM DENSITY MATRIX U_0

Let us now prove the result (3) for the density matrix U_0 which gives the true maximum for the right-hand side of (1). [U_0 actually gives an entropy greater by $k(2-\ln 2)$ than does the microcanonical ensemble matrix U_m!] First, we observe that the optimum U must commute with the Hamiltonian H. For, if not, write $U=U_1+U_2$ with

$$\begin{aligned}(U_1)_{ik}&=U_{ik} && i=k\\ &=0 && i\neq k,\end{aligned}$$

where the matrix elements refer to a representation in which H is diagonal. Then U_1 commutes with H,

$$\begin{aligned}&\text{Trace } U_1=\text{Trace } U=1,\\ &\text{Trace } HU_1=\text{Trace } HU=E,\\ &\text{Trace } U_1U_2=0,\\ &\text{Trace } U^2=\text{Trace } U_1^2+\text{Trace } U_2^2>\text{Trace } U_1^2,\\ &-k\ln(\text{Trace } U^2)<-k\ln(\text{Trace } U_1^2).\end{aligned}$$

Hence U_1, which commutes with H, gives a better bound on the entropy than does U.

Assuming then that U commutes with H, choose a representation in which both U and H are diagonal with eigenvalues U_i and E_i, respectively. We have

$$S(E)\geqq -k\ln\textstyle\sum_i U_i^2 \quad \text{subject to} \quad U_i\geqq 0,\ \sum U_i=1,\ \sum E_iU_i=E.$$

It is now a straightforward matter to maximize with respect to U_i, using the method of Lagrange multipliers. This yields for the eigenvalues of the optimum density matrix

$$\begin{aligned}U_i&=C(E+2/\beta-E_i) && E_i<E+2/\beta,\\ &=0 && E_i\geqq E+2/\beta.\end{aligned}$$

The two constants C and β are Lagrange multipliers which we must determine by the conditions Trace $U=1$ and Trace $HU=E$. We show that $k\beta=dS(E)/dE$ and $C=(\beta/e^2)e^{-S(E)/k}$.

$$\begin{aligned}\text{Trace } U&=C\textstyle\sum_{E_i<E+2/\beta}(E+2/\beta-E_i)\\ &=C\int_{\lambda<E+2/\beta}(E+2/\beta-\lambda)d[e^{S(\lambda)/k}]\\ &=C\left[\left\{(E+2/\beta-\lambda)+\frac{k}{S'(\lambda)}-\frac{k^2}{S'(\lambda)}\frac{d}{d\lambda}\frac{1}{S'(\lambda)}\cdots\right\}e^{S(\lambda)/k}\right]_{\lambda=E+2/\beta}\end{aligned}$$

on repeatedly integrating by parts. The lower limit for λ is quite irrelevant because of the dominance of the term $e^{S(\lambda)/k}$ at the upper limit, the entropy being proportional to the volume of the system. Now each term in the above expansion is of order $1/N$ compared to the previous term. Hence, retaining only the leading term for large N

$$1=\text{Trace}U=\frac{Ck}{S'(\lambda_m)}e^{S(\lambda_m)/k}+\text{relative order }\left(\frac{1}{N}\right),$$

where $\lambda_m=E+2/\beta$. Similarly we find

$$0=\text{Trace }(H-E)U=\frac{2Ck}{S'(\lambda_m)}\left[\frac{1}{\beta}-\frac{k}{S'(\lambda_m)}\right]e^{S(\lambda_m)/k}+\text{relative order }\left(\frac{1}{N}\right),$$

whence

$$k\beta=S'(\lambda_m)=S'(E)+\text{ order }(1/N)$$

and

$$C=\frac{S'(\lambda_m)}{k}e^{-S(\lambda_m)/k}=\frac{\beta}{e^2}e^{-S(E)/k}+\text{relative order }\left(\frac{1}{N}\right).$$

Thus we obtain for the eigenvalues of the optimum density matrix U_0

$$\begin{aligned}U_i&=(\beta/e^2)(E+2/\beta-E_i) && E_i<E+2/\beta\\ &=0 && E_i\geqq E+2/\beta.\end{aligned}$$

This completes the proof of (3).

DISCUSSION

Corresponding to the bound (1) for the entropy we have the related variation principles for the free-energy F and the thermodynamic potential $\Omega=-PV$:

$$F\leqq\text{Trace } HU+(1/\beta)\ln(\text{Trace } U^2),\tag{8}$$

$$\Omega\leqq\text{Trace}(H-\mu N_{\text{op}})U+(1/\beta)\ln(\text{Trace } U^2),\tag{9}$$

where U is non-negative, Hermitian, and has unit trace; in (9) the formalism of second quantization is used, N_{op} is the number operator and μ the chemical potential.

As an application, let us show how the principle (9) can be used to obtain the Husimi equations[2] for the variationally best independent-particle model. As trial density matrix in (9) we take[3]

$$U=(\text{constant})\theta(\lambda-\tilde{H})\,, \qquad (10)$$

where

$$\tilde{H}=\sum_{k,k'}\gamma_{kk'}a_k{}^\dagger a_{k'}\,,$$

a_k is the annihilation operator in an arbitrary representation, λ and $\gamma_{kk'}$ are variation parameters and θ is the step function:

$$\theta(x)=0 \qquad x<0$$
$$\quad =1 \qquad x>0$$

for a real argument x, while the step function of an operator is to be understood in the sense given by von Neumann.[1] The trial matrix (10) may be regarded as that of a microcanonical ensemble for an effective Hamiltonian $\tilde{H}$,[4] and we seek the best such, i.e., that which optimizes the bound (9).

Let us write

$$\Sigma(\lambda)=\ln[\text{Trace}\,\theta(\lambda-\tilde{H})]\,,$$
$$\tilde{\beta}=d\Sigma(\lambda)/d\lambda\,,$$
$$\tilde{U}=\exp(-\tilde{\beta}\tilde{H})/\text{Trace}\exp(-\tilde{\beta}\tilde{H})\,.$$

$\tilde{U}$ may be regarded as the grand canonical ensemble matrix which corresponds to the microcanonical ensemble matrix (10); because of the well-known equivalence of the grand and microcanonical ensembles[5] $\tilde{U}$ will give the same statistical averages of extensive variables as does U, apart from terms of relative order $\ln N/N$ or $1/N$. Thus

$$\text{Trace}(H-\mu N_{\text{op}})U\approx\text{Trace}(H-\mu N_{\text{op}})\tilde{U}$$
$$-\ln(\text{Trace}\,U^2)=\Sigma(\lambda)$$
$$\approx-\text{Trace}\,\tilde{U}\ln\tilde{U}\,,$$

where the neglected terms are irrelevant in the limit of large volume. Hence (9) gives

$$\Omega\leqq\text{Trace}\,(H-\mu N_{\text{op}}+(1/\beta)\ln\tilde{U})\tilde{U}.$$

This is precisely the conventional bound to Ω corresponding to (2), using a grand canonical-type independent-particle matrix $\tilde{U}$.

Thus the trial matrix (10) used to optimize the bound (9) will simply reproduce the Husimi formalism.[6]

Finally, we note that (1) and (2) are special cases of a more general principle, true for any positive number a:

$$S(E)\geqq-(k/a)\ln(\text{Trace}\,U^{a+1})$$

subject to U non-negative, Trace $U=1$, Trace $HU=E$. This principle is readily proved from (2) by using the inequality

$$-x\ln x\geqq(x-x^{a+1})/a,\quad \text{for}\quad a>0,\quad x\geqq0,$$

(1) is obtained by taking $a=1$, and (2) by letting $a\to0$.

[2] K. Husimi, Proc. Phys. Maths. Soc. Japan, 22, 264(1940).

[3] We could equally well take the form $\theta(\lambda-\tilde{H})-\theta(\lambda_1-\tilde{H})$ with $\lambda_1<\lambda$; because of the rapid increase of Trace $\theta(\lambda-\tilde{H})$ with λ, the second term is quite irrelevant.

[4] Strictly, $\tilde{H}$ corresponds to an effective one-particle $H-\mu N_{\text{op}}$, rather than to a Hamiltonian.

[5] The equivalence may be shown in the present case by writing the step function in (10) as an integral transform:

$$\theta(\lambda-\tilde{H})=\frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty}\frac{e^{\beta(\lambda-\tilde{H})}}{\beta}d\beta,\ \text{where}\ c>0.$$

After carrying out the statistical averages using the operator

$$e^{\beta(\lambda-\tilde{H})}\,,$$

we then perform the integration over β using the saddle-point method; only the neighborhood of $\beta=\tilde{\beta}$ contributes significantly.

[6] It should be noted that $\tilde{U}$ would not lead to the Husimi equations if used as a trial density matrix in (9). Indeed, $\tilde{U}$ does not give the correct value of Ω in (9) even in the absence of interaction.