Advanced statistical mechanics
(WS 12/13, FU Berlin)

Problem sheet 6

Due date: December 5, 2012

Problem 19: Lagrange multipliers (3+3)

Consider the the function f given by $f(x, y) = x^3 y$.

a) Determine all extrema of this function on the unit circle, i.e., under the constraint that $x^2 + y^2 = 1$, by use of the Lagrange multiplier technique.

b) Determine for each extremal point whether it is a local/global minimum/maximum or a saddle point.

Problem 20: Jaynes' principle of maximal entropy (2+1+1+2+2+2+1+1+4*=12+4*)

Let Γ be a finite sample space, i.e., Γ has finite cardinality $|\Gamma| = n$ and let $M^{(j)} : \Gamma \to \mathbb{R}$ be independent random variables for $j = 1, \dots, m$. We call the $M^{(j)}$ *observables* and assume that they are fixed. The goal is to find a probability vector $p \in [0, 1]^n$ which yields the values $K^{(j)} \in \mathbb{R}$ for their expectation values, i.e.,

$$\langle M^{(j)} \rangle = K^{(j)} \tag{1}$$

and under these constraints maximizes the Shannon entropy

$$I(p) = -\sum_{\gamma=1}^{n} p_\gamma \ln(p_\gamma). \tag{2}$$

This can be done by use of the Lagrange multiplier technique.

In the end this will allow us to write the maximal value of the Shannon entropy as a function of the $K^{(j)}$ and derive *Gibbs' fundamental equation* from Jaynes' principle of entropy maximization.

a) Denoting the Lagrange multiplier for the constraint $\sum_{\gamma=1}^{n} p_\gamma = 1$ by $(\psi - 1)$ and those of (1) by λ_j show that an extremal point of the Shannon entropy I is given by

$$p_\gamma = \mathrm{e}^{-\psi}\, \mathrm{e}^{-\lambda \cdot M_\gamma}, \tag{3}$$

where $\lambda \cdot M_\gamma = \sum_{j=1}^{m} \lambda_j M_\gamma^{(j)}$.
Hint: In order to obtain the given signs, you should define the Lagrange function as follows:

$$L = I - \lambda \cdot (\langle M \rangle - K) - (\psi - 1)\Big(\sum_{\gamma=1}^{n} p_\gamma - 1\Big) \tag{4}$$

b) Prove that the Shannon entropy is a strictly concave function by showing that its second derivatives on $]0, 1[^n$ are strictly negative.

c) Conclude that the extremal point found in Problem 20a is a maximum of the Shannon entropy.

d) Prove that ψ is a function of λ only by showing that

$$\psi(\lambda) = \ln \sum_{\gamma=1}^{n} \mathrm{e}^{-\lambda \cdot M_\gamma}. \tag{5}$$

(Of cours, λ is itself a function of K.)

e) Show that the value of the maximal Shannon entropy is

$$S(K) \coloneqq \psi(\lambda(K)) + \sum_{j=1}^{m} \lambda_j(K)\, K^{(j)}, \tag{6}$$

which we just call *entropy* from now on.
Hint: Remember Eq. (1).

f) Prove the so-called *Gibbs' fundamental equation*

$$\mathrm{d}S = \sum_{j=1}^{m} \lambda_j \,\mathrm{d}K^{(j)}. \tag{7}$$

Remark: In particular, this means that $\frac{\partial S}{\partial K^{(j)}} = \lambda_j$.
Hint: Show first that $\frac{\partial \psi}{\partial \lambda_j} = -K^{(j)}$, e.g., using Eqs. (5) and (3).

One can show that ψ is a smooth and strictly convex function of λ.

g) Using this, show that $\psi = -S^*$, i.e., that ψ is the negative Legendre transform of S and vice versa. *Hint:* Remember the definition of the Legendre transform from Problem 5.

Now consider the special case where $M^{(1)} = H$ is the Hamiltonian of the system, and $M^{(2)} = \hat{V}$ is the volume, i.e., $K^{(1)} = U$ and $K^{(2)} = V$.

h) Conclude that $\lambda_1 = 1/T$ and $\lambda_2 = p/T$.

In the case where the state space is $\Gamma = \mathbb{R}^{6N}$, with N being the particle number the probability measure on Γ is given by a probability density function $\rho : \Gamma \to \mathbb{R}$ and the Shannon information by

$$I(\rho) = -\int_\Gamma \rho(\gamma) \ln(\rho(\gamma))\; \mathrm{d}^{6N}\gamma. \tag{8}$$

i) *Bonus question:* Using calculus of variations and Jaynes' principle derive the results equivalent to Eqs. (3) and (5) for the case $\Gamma = \mathbb{R}^{6N}$, i.e.

$$\rho(\gamma) = \mathrm{e}^{-\psi}\, \mathrm{e}^{-\lambda\cdot M_\gamma} \tag{9}$$

and

$$\psi(\lambda) = \ln \int_\Gamma \mathrm{e}^{-\lambda\cdot M_\gamma}\, \mathrm{d}^{6N}\gamma. \tag{10}$$

Remark: This means that everything works exactly the same way as in the discrete case. The probability vector p can be replaced by the probability density function ρ and the sums over γ become integrals.