Maximum Entropy Principle for Physical Systems

John Wyatt
4/21/04

In thermodynamic equilibrium, the probability distribution will be the solution to the optimization problem:

$$\text{Maximize} \quad S = -k_B \sum_{k=1}^{n} p_k \ln p_k \tag{1}$$

subject to the constraints

$$\sum_{i=1}^{n} p_i E_i = E \tag{2}$$

$$\sum_{i=1}^{n} p_i = 1 \tag{3}$$

$$p_i \geq 0, \quad 1 \leq i \leq n \tag{4}$$

Lagrange Multipliers for Physical Systems

We ignore the constraint $p_i \geq 0$ and will find that it automatically is satisfied for this problem.

In order to make the Lagrange multipliers agree with those in the lecture notes, we maximize instead

$$\phi(\boldsymbol{p}) = -\sum_{k=1}^{n} p_k \ln p_k \tag{5}$$

Lagrange multipliers tell us the maximum will be found at the point (or, in general, among the several points) where

$$\nabla\phi(\boldsymbol{p}) = \alpha'\nabla\left(\sum_{i=1}^{n} p_i\right) + \beta\nabla\left(\sum_{i=1}^{n} p_i E_i\right), \tag{6}$$

i.e.,

$$\frac{\partial\phi}{\partial p_j} = \alpha'\frac{\partial}{\partial p_j}\left(\sum_{i=1}^{n} p_i\right) + \beta\frac{\partial}{\partial p_j}\left(\sum_{i=1}^{n} p_i E_i\right). \tag{7}$$

Evaluating the derivatives, for each value of *j*,

$$\frac{\partial\phi}{\partial p_j} = \frac{\partial}{\partial p_j}\left(-\sum_{k=1}^{n} p_k \ln p_k\right) = -\left(\ln p_j + 1\right) \tag{8}$$

$$\frac{\partial}{\partial p_j}\sum_{i=1}^{n} p_i = 1 \tag{9}$$

$$\frac{\partial}{\partial p_j}\sum_{i=1}^{n} p_i E_i = E_j \tag{10}$$

Substituting (8), (9) and (10) into (7) gives, for each *j*,

$$-\left(\ln p_j + 1\right) = \alpha' + \beta E_j \tag{11}$$

$$\ln p_j = -\left(\alpha' + 1\right) - \beta e_j = -\alpha - \beta E_j \tag{12}$$

$$p_j = e^{-\alpha} e^{-\beta E_j} \tag{13}$$

This gives the general form of the maximum entropy solution. The remaining goal is to choose α and β so that both constraints (2) and (3) are satisfied.

Satisfying constraint (3) that the probabilities sum to 1 lets us eliminate α:

$$\sum_{j=1}^{n} p_j = \sum_{j=1}^{n} e^{-\alpha} e^{-\beta E_j} = e^{-\alpha} \sum_{j=1}^{n} e^{-\beta E_j} = 1, \tag{14}$$

so

$$e^{-\alpha} = \frac{1}{\sum_{j=1}^{n} e^{-\beta E_j}}, \tag{15}$$

i.e., the general solution has the form

$$\boxed{p_j{}^* = \frac{e^{-\beta E_j}}{\sum_{j=1}^{n} e^{-\beta E_j}}, \quad 1 \le j \le n}, \tag{16}$$

where $p_j{}^*$ indicates the value of p_j that maximizes the entropy. The Lagrange multiplier method works and the form of the solution is still valid if there are infinitely many energy states, $n \to \infty$. Since the energies are all positive, we see that with infinitely many energies we must have

$\beta > 0$ so the probabilities can be normalized. In that case the probability a given state is occupied shrinks exponentially with its energy. But with a finite number of energies it is also possible to have $\beta < 0$ and for the probability a state is occupied to *grow* exponentially with its energy. The phenomenon, called *population inversion*, underlies the operation of all lasers.

A General Feature of Lagrange Multipliers

A general geometric feature of the solution to Lagrange multiplier problems will help us interpret β. Lets begin with a simple example:

$$\text{Maximize } \phi(x,y) \tag{17}$$

subject to

$$g(x,y) = ax + by = G \tag{18}$$

Lagrange multipliers tells us that the optimal solution will be found at a point where, for some λ,

$$\nabla\phi(x,y) = \lambda\nabla g(x,y), \tag{19}$$

i.e.,

$$\frac{\partial\phi}{\partial x}(x,y) = \lambda\frac{\partial g}{\partial x}(x,y) = \lambda a \tag{20}$$

$$\frac{\partial\phi}{\partial y}(x,y) = \lambda\frac{\partial g}{\partial y}(x,y) = \lambda b, \tag{21}$$

i.e.,

$$\nabla\phi(x,y)=\lambda\begin{pmatrix}a\\b\end{pmatrix} \tag{22}$$

Example

$$\text{Maximize}\quad \phi(x,y)=x^2+y^2 \tag{23}$$

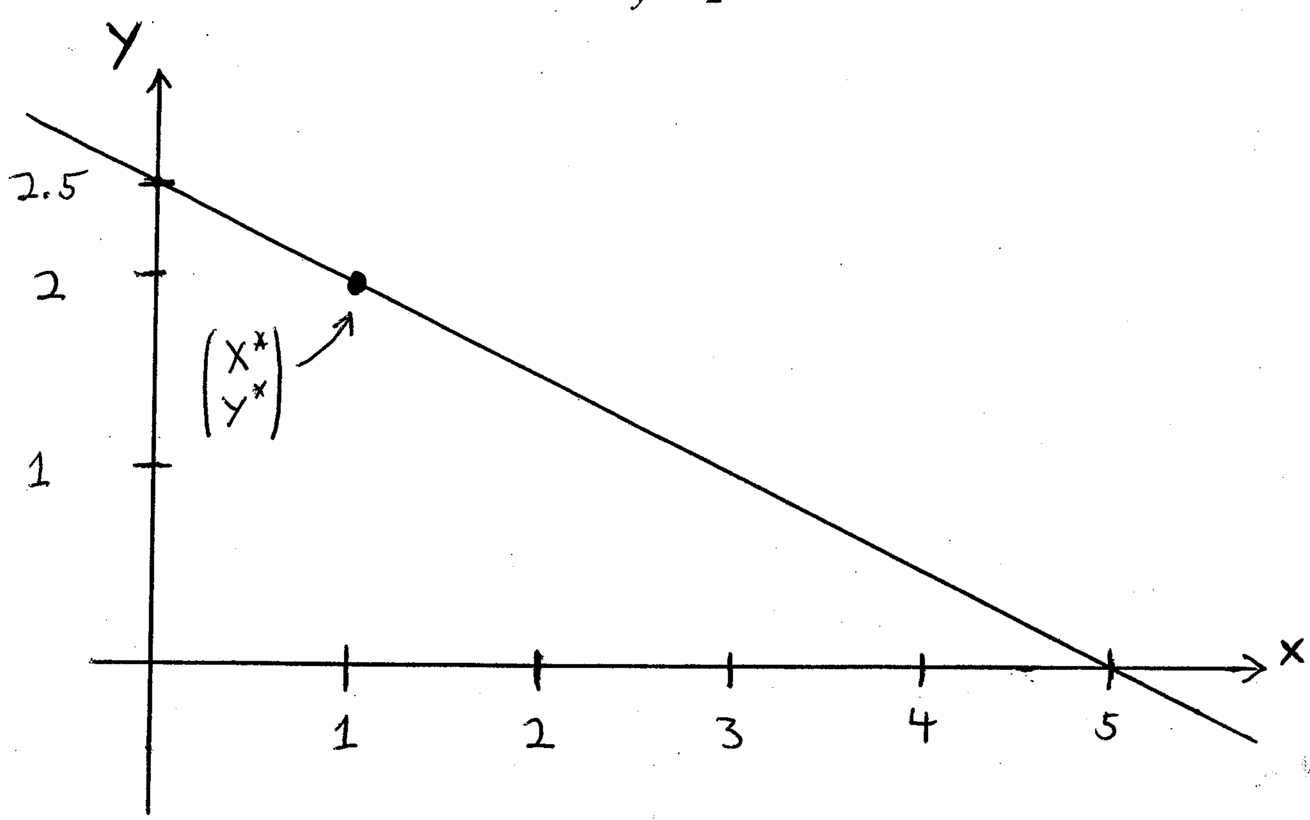
$$\text{subject to}\quad g(x,y)=x+2y=5 \tag{24}$$

$$\nabla\phi=\begin{pmatrix}2x\\2y\end{pmatrix}=\lambda\nabla g=\lambda\begin{pmatrix}1\\2\end{pmatrix} \tag{25}$$

$$\begin{aligned}2x&=\lambda\\2y&=2\lambda,\end{aligned}$$

i.e.,

$$\begin{aligned}y&=2x\\5=x+2y=x+4x&=5x\\x^*&=1\\y^*&=2\end{aligned} \tag{26}$$



Sensitivity to Constraints

Suppose we alter the optimum solution

$$\begin{pmatrix} x^* \\ y^* \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \end{pmatrix} \tag{27}$$

to some nearby point

$$\begin{pmatrix} x^* + \delta x \\ y^* + \delta y \end{pmatrix}, \tag{28}$$

The nearby point might be a nonoptimal value, or it might be the optimal solution subject to the altered value of the constraint

$$g(x,y) = G + \delta G \tag{29}$$

How do ϕ and g change as we move to the nearby point? By the chain rule for calculus

$$\delta\phi = \frac{\partial\phi}{\partial x}(x^*, y^*)\,\delta x + \frac{\delta\phi}{\partial y}(x^*, y^*)\,\delta y \tag{30}$$

$$= (\nabla\phi)_{(x^*,y^*)} \bullet \begin{pmatrix} \delta x \\ \delta y \end{pmatrix} \tag{31}$$

and

$$\delta g = \frac{\delta g}{\partial x}(x^*, y^*)\delta x + \frac{\partial g}{\partial y}(x^*, y^*)\delta y = (\nabla g)_{(x^*,y^*)} \bullet \begin{pmatrix} \delta x \\ \delta y \end{pmatrix}. \tag{32}$$

But by the principle of Lagrange multipliers,

$$\nabla \phi(x^*, y^*) = \lambda \nabla g(x^*, y^*), \tag{33}$$

and therefore

$$\delta\phi = (\nabla\phi) \bullet \begin{pmatrix} \delta x \\ \delta y \end{pmatrix} = \lambda (\nabla g) \bullet \begin{pmatrix} \delta x \\ \delta y \end{pmatrix} = \lambda \delta g, \tag{34}$$

Lagrange Multiplier Sensitivity Principle

For *any* small perturbation about the optimal solution to a Lagrange multiplier problem with a single constraint

$$\boxed{\frac{\delta\phi}{\delta g} = \lambda} \tag{35}$$

Interpretation of β

Returning to our original problem

$$\text{maximize} \quad \phi(\boldsymbol{p}) = -\sum_{k=1}^{n} p_k \ln p_k \tag{36}$$

subject to

$$g_1(\boldsymbol{p}) = \sum_{i=1}^{n} p_i = 1 \tag{37}$$

$$g_2(\boldsymbol{p}) = \sum_{i=1}^{n} p_i E_i = E \tag{38}$$

at the optimum solution $\boldsymbol{p}^*$ we have

$$\nabla\phi(\boldsymbol{p}^*) = \alpha\nabla g_1(\boldsymbol{p}^*) + \beta\nabla g_2(\boldsymbol{p}^*) \tag{39}$$

and therefore for *any* small perturbation to $(\boldsymbol{p}^*+\delta\boldsymbol{p})$ we have

$$\delta\phi = (\nabla\phi)\cdot(\delta\boldsymbol{p}) = \alpha(\nabla g_1)\cdot(\delta\boldsymbol{p}) + \beta(\nabla g_2)\cdot(\delta\boldsymbol{p}) = \alpha\delta g_1 + \beta\delta g_2 \tag{40}$$

$$\delta\phi = \alpha\delta g_1 + \beta\delta g_2 \tag{41}$$

In particular, for any perturbation $\delta\boldsymbol{p}$ such that $\boldsymbol{p}^*+\delta\boldsymbol{p}$ is a valid probability distribution, i.e.,

$$\sum_{i=1}^{n} (p_i^* + \delta p_i) = 1 \tag{42}$$

$$\delta g_1 = \sum_{i=1}^{n} \delta p_i = 0, \tag{43}$$

we have

$$\delta\phi = \beta\delta g_2 = \beta\delta\left(\sum_{i=1}^{n} p_i E_i\right) = \beta\delta E. \tag{44}$$

$$\beta\sum_{i=1}^{n} E_i \delta p_i = \beta\delta E$$

Recalling that our original goal was to maximize

$$S = k_B\phi = -k_B \sum_{k=1}^{n} p_k \ln\left(p_k\right), \tag{45}$$

for any such perturbation about $\boldsymbol{p}^*$,

$$\delta S = k_B \beta \delta E \tag{46}$$

$$\boxed{\frac{\delta S}{\delta E} = k_B \beta} \tag{47}$$

Looking ahead to a comparison with classical thermodynamics, where temperature plays the role

$$\frac{\partial E}{\partial S} = T, \text{ in degrees Kelvin,} \tag{48}$$

we anticipate that

$$k_B \beta = \frac{1}{T} \tag{49}$$

$$\boxed{\beta = \frac{1}{k_B T}} \tag{50}$$

and therefore

$$\boxed{p_k{}^* = \frac{e^{-E_k/k_BT}}{\sum_{k=1}^{n} e^{-E_k/k_BT}}}. \tag{51}$$