MaxEnt and the Boltzmann principle

Wednesday, September 6, 2017
BME/CHE/PHY 558, Physical & Quantitative Biology
Rutgers University: Chemical Thermodynamics
Lecturer: Gábor Balázsi
Finding maxima and minima
Extrema: Where all partial derivatives are $= 0$

At extremum points, $f$ is flat in all directions – it stays constant as $x$ and $y$ change infinitesimally.

This requires that the differential $df = 0$:

$$df (x, y) = \frac{\partial f}{\partial x} dx + \frac{\partial f}{\partial y} dy = 0$$

Since $dx$ and $dy$ are independent, to assure $df = 0$:

$$\frac{\partial f}{\partial x} = 0 \quad \frac{\partial f}{\partial y} = 0$$

More generally:

$$\frac{\partial f}{\partial x_i} = 0 \quad \text{For all } i=1, 2, 3, \ldots$$
Example: Find the maximum!

Setting both partial derivatives = 0 defines the extremum in 2D: \[ \frac{\partial f}{\partial x} = \frac{\partial f}{\partial y} = 0 \]

\[ f(x, y) = ? \text{(Parabola)} \]

\[ \frac{\partial f}{\partial x} = ? \]

\[ \frac{\partial f}{\partial y} = ? \]
Example: Finding a maximum

Setting both partial derivatives = 0 defines the extremum in 2D:

\[
\frac{\partial f}{\partial x} = \frac{\partial f}{\partial y} = 0
\]

\[
f(x, y) = 4 - (x - 2)^2
\]

(a) Not a Maximum

(b) A Maximum
Example: finding the minimum of a paraboloid

\[ f(x, y) = (x - x_0)^2 + (y - y_0)^2 \]

\[ \frac{\partial f}{\partial x} = 2(x - x_0) = 0 \]

\[ \frac{\partial f}{\partial y} = 2(y - y_0) = 0 \]

\[ x^* = x_0 \]

\[ y^* = y_0 \]
Saddles are tricky...

Maximum: \( \frac{\partial f}{\partial x} = \frac{\partial f}{\partial y} = 0 \)

2\textsuperscript{nd} order partials < 0

and \( M > 0 \)

Minimum: \( \frac{\partial f}{\partial x} = \frac{\partial f}{\partial y} = 0 \)

2\textsuperscript{nd} order partials > 0

and \( M > 0 \)

Saddle: \( \frac{\partial f}{\partial x} = \frac{\partial f}{\partial y} = 0 \)

\( M < 0 \)

The Hessian is: \( M = \left. \frac{\partial^2 f}{\partial x^2} \right|_{x_0,y_0} \left. \frac{\partial^2 f}{\partial y^2} \right|_{x_0,y_0} - \left[ \left. \frac{\partial^2 f}{\partial x \partial y} \right|_{x_0,y_0} \right]^2 \)
Optimization with constraints. Lagrange multipliers.

Joseph-Louis Lagrange (1736 – 1813)
Finding an extremum subject to a constraint

At extrema with constraints, \( f \) is flat in a direction that satisfies the constraint: \( g(x, y) = 0 \)

We must have \( df = 0 \) for a specific pair \((dx, dy)\): \[
\frac{df}{dx} dx + \frac{df}{dy} dy = 0
\]

Which pair \((dx, dy)\)? Obtain these from the constraint:

\[
g(x, y) = 0 \quad \Rightarrow \quad \frac{dg}{dx} dx + \frac{dg}{dy} dy = 0 \quad \Rightarrow \quad \frac{dg}{dx} dx = -\frac{dg}{dy} dy
\]

Substitute back into \( df = 0 \). Obtain extremum, satisfying the constraint.

\[
\frac{df}{dx} dx + \frac{df}{dy} \frac{\partial g}{\partial x} dx = 0
\]
Finding the minimum subject to a constraint

Consider the function

\[ f(x, y) = x^2 + y^2 \]

\[ df(x, y) = 2\,dx + 2\,dy = 0 \]

We seek a dependency between \( dx \) and \( dy \) corresponding to the constraint.

Constraint:

\[ g(x, y) = x + y = 6 \]

\[ dg(x, y) = dx + dy = 0 \]

\[ 2\,dx - 2\,y\,dx = 0 \]

Then, for minimum with constraint:

\[ 2x = 6 \Rightarrow x^* = 3, \quad y^* = 3 \]
The method of Lagrange multipliers

\[ df(x, y) = \frac{\partial f}{\partial x} \, dx + \frac{\partial f}{\partial y} \, dy = 0 \]  
Both must be satisfied

\[ dg(x, y) = \frac{\partial g}{\partial x} \, dx + \frac{\partial g}{\partial y} \, dy = 0 \]

\[ \frac{dy}{dx} = - \left( \frac{\frac{\partial f}{\partial x}}{\frac{\partial f}{\partial y}} \right) = - \left( \frac{\frac{\partial g}{\partial x}}{\frac{\partial g}{\partial y}} \right) \]

\[ \frac{\partial f}{\partial x} = \lambda \frac{\partial g}{\partial x} \]  
\[ \frac{\partial f}{\partial y} = \lambda \frac{\partial g}{\partial y} \]  
Simply start with these.  
No need to write:  
\( df=0 \) and \( dg=0 \)
Example: Lagrange multipliers

Find a minimum on paraboloid $f(x,y)$, given the constraint $g(x,y)$

$$f(x,y) = x^2 + y^2$$

$$\frac{\partial f}{\partial x} = \lambda \frac{\partial g}{\partial x} \quad 2x = \lambda$$

$$\frac{\partial f}{\partial y} = \lambda \frac{\partial g}{\partial y} \quad 2y = \lambda$$

$$2x = 6 \Rightarrow x^* = 3, \ y^* = 3$$

$$g(x,y) = x + y = 6$$

$(x^*, y^*) = (3,3)$
Example: Find rectangle with maximum area

Given a piece of rope of length 40 ft, what is the largest rectangular yard we can enclose?

Area: \( A = xy \)

Perimeter: \( P = 2x + 2y = 40 \)

\[
\frac{\partial A}{\partial x} = \lambda \frac{\partial P}{\partial x} \\
\frac{\partial A}{\partial y} = \lambda \frac{\partial P}{\partial y}
\]

\[\begin{align*}
y &= 2\lambda \\
x &= 2\lambda
\end{align*}\]

\[x = y \quad 4x = 40 \Rightarrow x^* = 10, \quad y^* = 10 \]

\[A = 100\]

Solve a closely related problem:

Given a rectangular area = 100, enclosing which rectangular shape requires the shortest rope?
Generalizing Lagrange multipliers

If you have more than 2 variables:

\[
\frac{\partial f}{\partial x_1} - \lambda \frac{\partial g}{\partial x_1} = 0
\]

\[
\frac{\partial f}{\partial x_2} - \lambda \frac{\partial g}{\partial x_2} = 0
\]

\[
\vdots
\]

\[
\frac{\partial f}{\partial x_M} - \lambda \frac{\partial g}{\partial x_M} = 0
\]

If you have more than 1 constraint:

\[
\frac{\partial f}{\partial x_1} - \lambda \frac{\partial g}{\partial x_1} - \beta \frac{\partial h}{\partial x_1} = 0
\]

\[
\frac{\partial f}{\partial x_2} - \lambda \frac{\partial g}{\partial x_2} - \beta \frac{\partial h}{\partial x_2} = 0
\]

\[
\frac{\partial f}{\partial x_M} - \lambda \frac{\partial g}{\partial x_M} - \beta \frac{\partial h}{\partial x_M} = 0
\]
The chain rule

If the variable(s) of a function $f$ are themselves function(s) of other variables:

$$df = \frac{\partial f}{\partial x} dx + \frac{\partial f}{\partial y} dy$$

and

$$x = x(u); \ y = y(u)$$

We can use the chain rule:

$$\frac{df}{du} = \frac{\partial f}{\partial x} \frac{\partial x}{\partial u} + \frac{\partial f}{\partial y} \frac{\partial y}{\partial u}$$

\[\downarrow\]

$$df = \frac{\partial f}{\partial x} \frac{\partial x}{\partial u} du + \frac{\partial f}{\partial y} \frac{\partial y}{\partial u} du$$

Example: Assume that we are changing slowly $T$ and $V$ for an ideal gas: $pV = nRT$

Knowing $\frac{dT}{dt} = \alpha$ and $\frac{dV}{dt} = \beta$, what is $\frac{dp}{dt}$?

Answer:

$$\frac{dp}{dt} = \frac{\partial p}{\partial V} \frac{dV}{dt} + \frac{\partial p}{\partial T} \frac{dT}{dt} = \alpha \frac{nR}{V} - \beta \frac{nRT}{V^2}$$
The Boltzmann Law

Boltzmann’s grave. Zentralfriedhof, Vienna
A bit of history: Ludwig Boltzmann

Europe during Boltzmann’s life

Ludwig Boltzmann
1844 – 1906

Founder of statistical physics
Definition of thermodynamic entropy, $S$

$$S = k_B \ln(W)$$

$W$ = “Wahrscheinlichkeit”
Old German meaning: multiplicity
New German meaning: probability
We mean multiplicity by $W$.

$k_B \approx 1.38 \times 10^{-23} \text{JK}^{-1} =$ Boltzmann’s constant
Another bit of history: Claude Shannon

Imagine sending a binary sequence of messages. Which message sequence is more informative?

Message #1: 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
Message #2: 0 1 1 0 0 0 1 0 1 0 0 1 1 1 0 1 1 0 0

Shannon Entropy $H$ (in bits):

$$H = - \sum_{i} p_i \log_2(p_i)$$

$p_i$ = probability of message $i$ taken from all possible messages

The higher $H$, the more informative each message. Many symbols, uniformly distributed are great!
What relates $S$ to $H$? Both refer to distributions.

Imagine casting an $M$-faced die $N$ times. If outcome $i$ occurs $n_i$ times, the multiplicity $W$ is:

$$W = \frac{N!}{n_1! n_2! n_3! \ldots n_M!}$$

Use Stirling’s approximation: $n! \approx \left(\frac{n}{e}\right)^n$

$$W = \frac{N^N}{n_1^{n_1} n_2^{n_2} n_3^{n_3} \ldots n_M^{n_M}}$$

$$W = \frac{1}{p_1^{n_1} p_2^{n_2} p_3^{n_3} \ldots p_M^{n_M}}$$

$$\frac{S}{k} = \frac{\ln(W)}{N} = -\sum_i p_i \ln(p_i)$$
Why the logarithm in $k_B \ln(W)$?

Take two subsystems, A and B.

The entropy of the joint system should be the sum of individual entropies.

$$W_{A+B} = W_A W_B$$

$$S(W_{A+B}) = S(W_A W_B) = k \ln(W_A W_B) = S(W_A) + S(W_B)$$