Optimizing Multiple Constraints

Mar. 9th 2012

- Recall, in the method of Lagrange multipliers we studied, we were extremizing a function $f(\underline{x})$ subject to a constraint $g(\underline{x}) = 0$
- What if we have more than one constraint?

Problem Let $f : \mathbb{R}^n \to \mathbb{R}$ be a C^1 function and $\phi_j : \mathbb{R}^n \to \mathbb{R}$; $j = 1, \dots, m$ be m C^1 functions. Let $S = \{ \underline{x} \in \mathbb{R}^n : \phi_j(\underline{x}) = 0 \text{ for } j = 1, \dots, m \}$. Which points in S extremize f over S?

i.e. Extremize $f(x_1, \dots, x_n)$ subject to constraints
$$\begin{cases} \phi_1(x_1, \dots, x_n) = 0 \\ \vdots \\ \phi_m(x_1, \dots, x_n) = 0 \end{cases}$$

Key Idea: If we only consider points $\underline{x}$ that satisfy the constraints, then $(\underline{x}, f(\underline{x}))$ is an extremum iff the constraints at that point do not allow movement in a direction that changes f

→ otherwise, we could move in that direction to make f bigger or smaller

→ in other words, $df = 0$ for all $d\underline{x} = (dx_1, dx_2, \dots, dx_n)$

where $df \approx f(\underline{x} + d\underline{x}) - f(\underline{x})$
$$\approx \nabla f(\underline{x}) \cdot d\underline{x}$$

- Consider the level sets of f at $\underline{x}$
  - let $\{\hat{v}_L\} \subset \mathbb{R}^n :=$ the set of directions $(\|\hat{v}_L\| = 1)$ in which we can move and still remain in the level set
  - i.e. these are the unit vectors for which $df = 0$

$$\therefore \ \forall \ \hat{v}_L \text{ in this set } \nabla f \cdot \hat{v}_L = 0 \qquad (\mathrm{I})$$

- Consider the constraints $\phi_j(\underline{x}) = 0$, $j = 1, \dots, m$ at $\underline{x}$
  - let $\{\hat{v}_c\} \subset \mathbb{R}^n :=$ the set of directions $(\|\hat{v}_c\| = 1)$ in which we can move and still satisfy the constraint
  - since the constraints are $\phi_j = 0$ (a constant), we have just like for level sets $f = \text{constant}$ that the $\{\hat{v}_c\}$ are those directions for which $d\phi_j = 0 \quad \forall j$

or, $\nabla \phi_j \cdot \hat{v}_c = 0 \quad \forall j \qquad (\mathrm{II})$

"perpendicular to"

(I) ⇒ Directions that don't change the value of f are $\perp \nabla f$

⇒ ∇f changes the value of f (provided $\nabla f \neq 0$)

(II) ⇒ All directions that still satisfy the constraints are $\perp \nabla \phi_j$, $j = 1, \ldots, m$

(I) + (II)

> A point $\underset{\sim}{x}$ is a constrained extremum of f iff the direction that changes f violates at least one of the constraints.

⇔ The direction that changes f the most — namely, ∇f — violates at least one of the constraints

- a direction $\hat{v}$ violates a constraint ϕ_i ⇔ $\nabla \phi_i \cdot \hat{v} \neq 0$
- i.e. each constraint yields a "forbidden direction" given by $\nabla \phi_i$
- set of forbidden directions $\{\nabla \phi_j\}_{j=1}^{m}$

Then the direction that changes f violates at least one constraint

$$\Leftrightarrow \quad \nabla f \in \operatorname{span}\{\nabla \phi_j\}_{j=1}^{m}$$

Lagrange multipliers

$$\Leftrightarrow \quad \nabla f = \lambda_1 \nabla \phi_1 + \cdots + \lambda_m \nabla \phi_m \quad \text{for } \lambda_1, \lambda_2, \ldots, \lambda_m \in \mathbb{R}$$

or

$$\nabla f - \sum_{j=1}^{m} \lambda_j \nabla \phi_j = 0$$

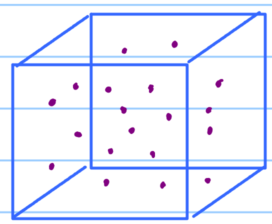
Thm (Method of Lagrange Multipliers) Extremal points $\underset{\sim}{x}$ of f on the constraints $\phi_1, \phi_2, \ldots, \phi_m$ satisfy

$$\nabla f(\underset{\sim}{x}) - \sum_{j=1}^{m} \lambda_j \phi_j(\underset{\sim}{x}) = 0 \quad \text{and} \quad \phi_j(\underset{\sim}{x}) = 0 \quad \text{for } j = 1, \ldots, m$$

for some $\lambda_1, \lambda_2, \ldots, \lambda_m \in \mathbb{R}$ called Lagrange Multipliers.

Notes:
- we still also need to check points where $\nabla \phi_j(\underset{\sim}{x}) = \underset{\sim}{0}$
- we still also need to check endpoints of the constraints.

Example : Statistical Mechanics



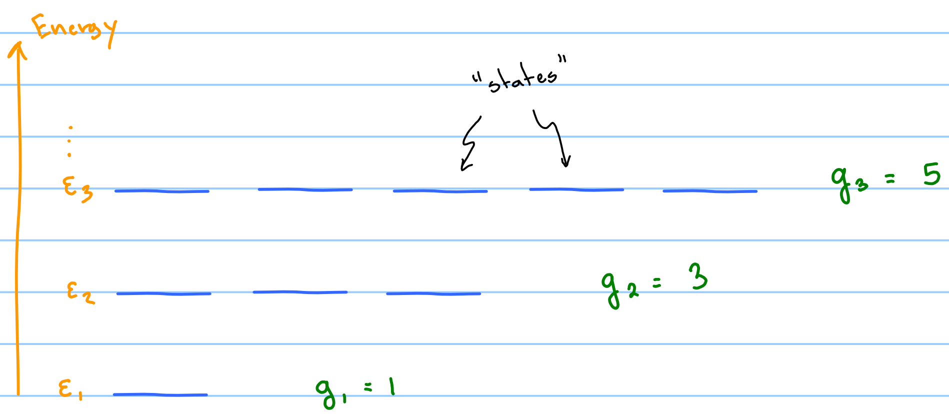
- Suppose we have some collection of N particles (very large, $\sim 10^{26}$) where each particle can have as its energy one of

$$0 < \varepsilon_1 < \varepsilon_2 < \cdots < \varepsilon_n$$

and that there are g_j states available to a particle if it has energy ε_j

(i.e. there are g_j ways for a particle to have energy ε_j)

Pictorially,



- The configuration of the system is specified by the distribution of the particles among the system's energy levels

  i.e. the numbers $N_j :=$ # of particles with energy ε_j

- Notice, we must have $\sum_{j=1}^{n} N_j = N$ — (1)

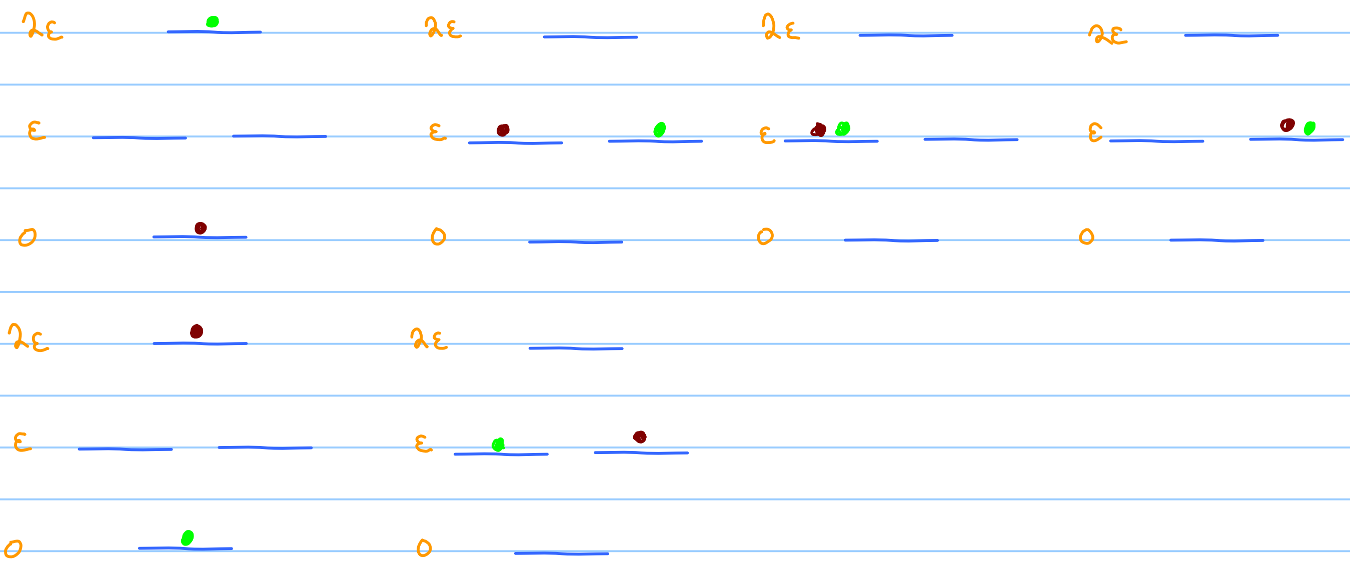
- We pose the following question :

  Suppose our system has a total energy $U = \sum_{j=1}^{n} N_j \varepsilon_j$. — (2)

  What are the values taken by the N_j , $j = 1, \ldots, n$?
  Or, what is the configuration the system assumes ?

- Notice, there be many different ways to produce a given configuration, or macrostate $(N_1, \dots, N_n)$

ex: $N = 2$, $U = 2\varepsilon$, possible energies $\varepsilon_0 = 0$, $\varepsilon_1 = \varepsilon$, $\varepsilon_2 = 2\varepsilon$, $g_0 = 1$, $g_1 = 2$, $g_2 = 1$. Then, possible states are



→ 6 microstates
(Note: here we are assuming the particles are distinguishable)

Def$^{\text{n}}$: $\omega(N_1, \dots, N_n) :=$ # of microstates that produce the macrostate $(N_1, N_2, \dots, N_n)$

So, in the above, eg. $\omega(N_0 = 0, N_1 = 2, N_2 = 0) = 4$
$\omega(N_0 = 1, N_1 = 0, N_2 = 1) = 2$
$\omega(\text{anything else}) = 0$

- Statistical mechanics postulates that how the particles in a system arrange themselves is completely random and that each microstate is equally likely

$\Rightarrow$ The most likely, or equilibrium configuration $(N_1, N_2, \ldots, N_n)$ is that for which $\omega(N_1, N_2, \ldots, N_n)$ is maximal.

So, to answer the question,

Maximize $\omega(N_1, \ldots, N_n)$ subject to the constraints

(1) $\sum N_j = N$ [conservation of particle number]

(2) $\sum N_j \varepsilon_j = U$ [conservation of energy]

Solution For N distinguishable particles with energy levels $\varepsilon_1, \ldots, \varepsilon_n$ and degeneracies $g_1, \ldots, g_n$, one can show using combinatorics that

$$\omega(N_1, \ldots, N_n) = N! \prod_{i=1}^{n} \frac{g_i^{N_i}}{N_i!} \qquad (3)$$

- since ln is a monotonically-increasing function, lets maximize $\ln \omega$ instead of ω

$$\ln \omega = \ln N! + \sum_{i=1}^{n} \left[N_i \ln g_i - N_i! \right]$$

$$\approx N \ln N - N + \sum \left[N_i \ln g_i - N_i \ln N_i + N_i \right]$$

$$= N \ln N + \sum \left[N_i \ln g_i - N_i \ln N_i \right]$$

Stirling's approximation
$\ln x! \approx x \ln x - x$
for large x
(recall $N \sim 10^{26}$)

Constraints (1): $\phi_1(N_1, \ldots, N_n) = \sum N_i - N = 0$

(2): $\phi_2(N_1, \ldots, N_n) = \sum N_i \varepsilon_i - U = 0$

Now, we look for $(N_1, \ldots, N_n)$ satisfying

$$\nabla(\ln \omega) + \lambda_1 \nabla \phi_1 + \lambda_2 \nabla \phi_2 = 0$$

$$\Rightarrow \quad \frac{\partial}{\partial N_j}\ln\omega + \lambda_1 \frac{\partial}{\partial N_j}\phi_1 + \lambda_2 \frac{\partial}{\partial N_j}\phi_2 = 0 \qquad \text{for} \quad j = 1, \ldots, n$$

We have $\frac{\partial}{\partial N_j}\ln\omega = \ln g_j - \ln N_j - 1$

$$\approx \ln g_j - \ln N_j \qquad \text{for large } N_j,\ g_j$$

$$\frac{\partial\phi_1}{\partial N_j} = 1$$

$$\frac{\partial\phi_2}{\partial N_j} = \varepsilon_j$$

$$\Rightarrow \quad \ln g_j - \ln N_j + \lambda_1 + \lambda_2\varepsilon_j = 0$$

$$\ln\left(\frac{N_j}{g_j}\right) = \lambda_1 + \lambda_2\varepsilon_j$$

$$\boxed{\frac{N_j}{g_j} = e^{\lambda_1 + \lambda_2\varepsilon_j}} \qquad (*)$$

This is the famous Maxwell - Boltzmann distribution – it tells us the values of N_j, i.e. the system's configuration.

- To eliminate λ_1 and λ_2, we would need to use the constraints.

From $\sum N_j = N$ and $(*)$,

$$\sum N_j = \sum g_j e^{\lambda_1 + \lambda_2\varepsilon_j} = e^{\lambda_1}\sum g_j e^{\lambda_2\varepsilon_j} = N$$

The "partition function"

$$\Rightarrow \quad e^{\lambda_1} = \frac{N}{\sum g_j e^{\lambda_2\varepsilon_j}} \equiv \frac{N}{Z} \qquad \text{where } Z = \sum g_j e^{\lambda_2\varepsilon_j}$$

- Solving for λ_2 is typically done using certain results from thermodynamics – one could use the second constraint, but it's pretty messy. A good exercise, however!

Propⁿ $\lambda_2 = -\frac{1}{kT}$ where T : system temperature
k : Boltzmann's constant

Then, the equilibrium configuration of the system is given by

$$\frac{N_j}{g_j} = \frac{N}{Z} e^{-\varepsilon_j / kT}$$

The Maxwell-Boltzmann Distribution.