Minimization

PHYS 250 (Autumn 2025) – Lecture 8

David Miller

Department of Physics and the Enrico Fermi Institute
University of Chicago

October 23, 2025

Outline

1. *Introduction to minimization*
   - Minimization is everywhere
   - Statement of the problem

2. *Least squares minimization*
   - Linear regression
   - Curve fitting

Minimization is everywhere

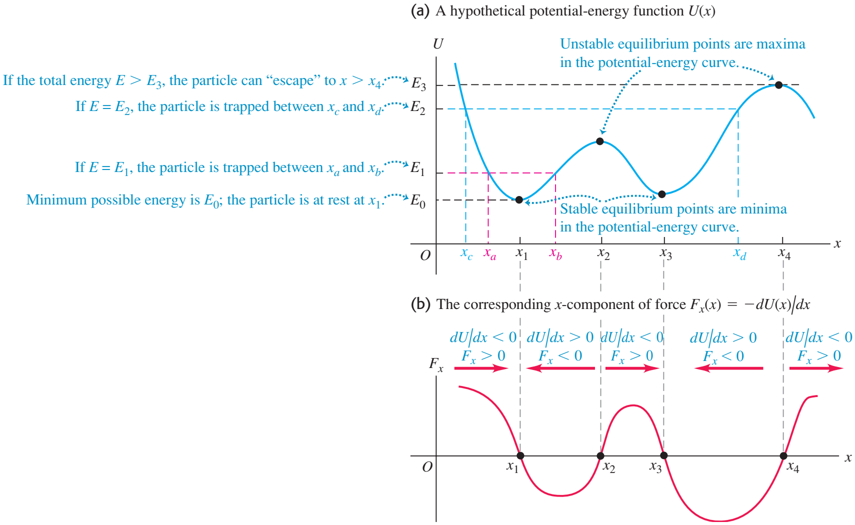
As physicists, we are **constantly attempting to minimize or maximize functions that describe the world around us**.

Examples of minimization

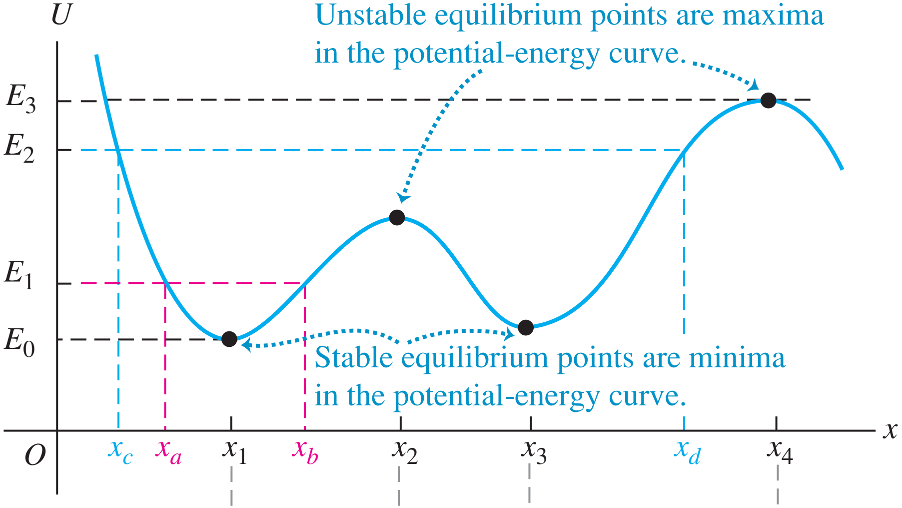
- **Fitting a model to data:** minimize differences between a model and data
- **Second law of thermodynamics:** minimize changes in entropy for a system in thermodynamic equilibrium
- **Conservation of momentum:** establish mechanical equilibrium by minimizing changes in momentum, $\frac{d\vec{p}}{dt} = 0$
- **Principle of least action:** obtain the equations of motion of a system by minimizing (or maximizing!) the variations of the action, S
- **Path integral formulation of quantum mechanics:** sort quantum mechanically possible trajectories by minimizing quantum action
- **Ising model:** minimization the energy of the spin configurations

Energy Minimization from your first year text books

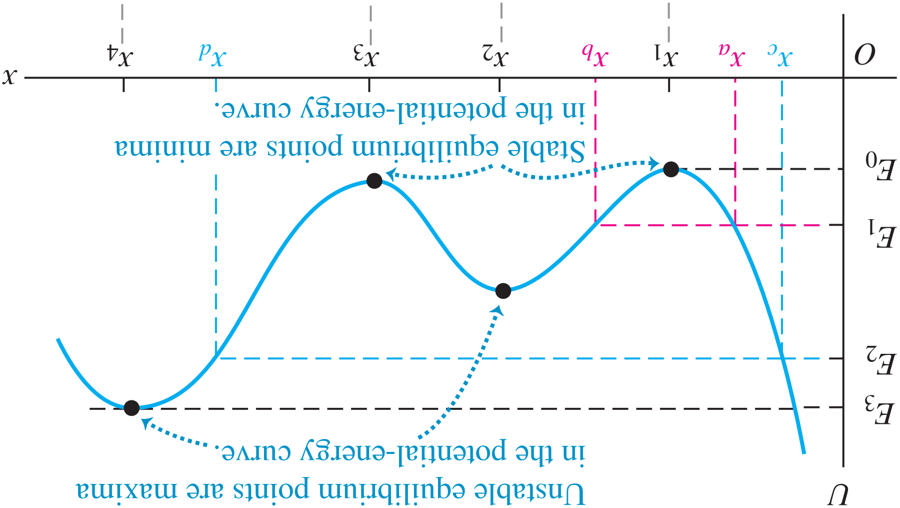
7.24 The maxima and minima of a potential-energy function $U(x)$ correspond to points where $F_x = 0$.



Minimization can imply maximization → optimization



Minimization can imply maximization → optimization



Optimization, or finding the extrema of a system

Since we are most often interested in maxima or minima of the evolution or behavior of a system as a function of some external parameter, the problem often boils down to the determination of **first and second derivatives** as a function of that parameter.

$$f(x) \approx f(a) + \frac{f'(a)}{1!}(x-a) + \frac{f''(a)}{2!}(x-a)^2 + \frac{f'''(a)}{3!}(x-a)^3 + \cdots, \quad (1)$$

If we focus only on the first three terms, we can write this for a function of n variables $\vec{x} = \sum_{i=0}^{n} x_i$ as:

$$f(\vec{x}) \approx f(\vec{a}) + (\vec{x}-\vec{a})^{\mathrm{T}} \nabla f(\vec{a}) + \frac{1}{2!}(\vec{x}-\vec{a})^{\mathrm{T}} \mathbf{H}(\vec{a})(\vec{x}-\vec{a}) \quad (2)$$

where $\mathbf{H}$ is the **Hessian matrix**, describing the **curvature** of $f(\vec{x})$ by

$$\mathbf{H}_{i,j} = \frac{\partial^2 f(\vec{a})}{\partial x_i \partial x_j} \quad (3)$$

(Note: The determinant of $\mathbf{H}$ is also sometimes referred to as **the Hessian**.)

Optimization methods and approaches

There are many details associated with the **existence, feasibility, and constraints** on the optimization problem for finding and describing extrema.

Assuming that these are generally satisfied, we can categorize the approaches into two primary groups and specific implementations of each:

- **Evaluate second derivatives (Hessians):** Newton's method is the most famous and widely used
- **Evaluate first derivatives (gradients):** Gradient descent is perhaps the most widely used

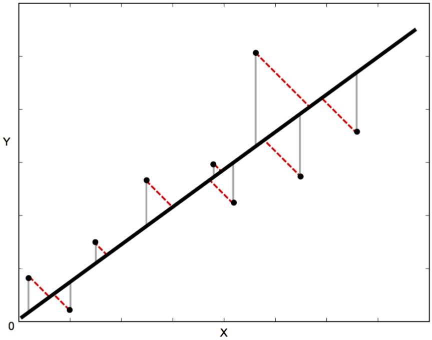
Then, there is a kind of "hybrid" approach which is referred to as **quasi-Newton** wherein the Hessian matrix is approximated using updates specified by gradient evaluations.

We will come back to these methods on Thursday, but for now, let's discuss a simple minimization: **Least squares**

Outline

Linear regression (II)

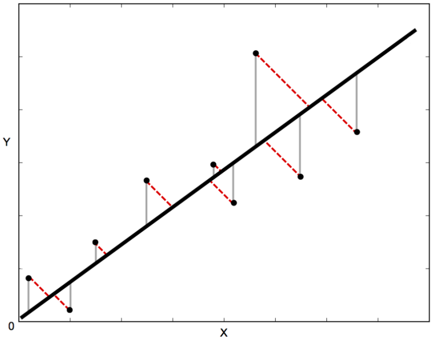
Suppose that you want to fit a set data points (x_i, y_i), where $i = 1, 2, \ldots, N$, to a straight line, $y = ax + b$.



The process of determining the best-fit line is called **linear regression**. This involves choosing the parameters a and b to minimize the sum of the squares of the differences between the data points and the linear function.

Linear regression (II)

Suppose that you want to fit a set data points (x_i, y_i), where $i = 1, 2, \ldots, N$, to a straight line, $y = ax + b$.



If there are only uncertainties in the y direction, then the differences in the vertical direction (the gray lines in the figure) are used. If there are uncertainties in both the x and y directions, the orthogonal (perpendicular) distances from the line (the dotted red lines in the figure) are used.

Using the χ^2 (again!)

For the case where there are only uncertainties in the y direction, there is an analytical solution to the problem.

If the uncertainty in y_i is σ_i, then the difference squared for each point is weighted by $w_i = 1/\sigma_i^2$. The function to be minimized with respect to variations in the parameters, a and b, is

$$\chi^2 = \sum_{i=1}^{N} w_i \left[y_i - (ax_i + b)\right]^2 . \tag{4}$$

The analytical solutions for the best-fit parameters that minimize χ^2 are those that satisfy $\dfrac{\partial(\chi^2)}{\partial a} = 0$ (and similarly for b).

Uncertainties

From the above equation for the χ^2, we can obtain a and b from:

$$a = \frac{\sum w_i \sum w_i x_i y_i - \sum w_i x_i \sum w_i y_i}{\sum w_i \sum w_i x_i^2 - \left(\sum w_i \sum w_i x_i\right)^2} \tag{5}$$

and

$$b = \frac{\sum w_i y_i \sum w_i x_i^2 - \sum w_i x_i \sum w_i x_i y_i}{\sum w_i \sum w_i x_i^2 - \left(\sum w_i \sum w_i x_i\right)^2}. \tag{6}$$

The uncertainties in the parameters are

$$\sigma_a = \sqrt{\frac{\sum w_i}{\sum w_i \sum w_i x_i^2 - \left(\sum w_i \sum w_i x_i\right)^2}} \tag{7}$$

$$\sigma_b = \sqrt{\frac{\sum w_i x_i^2}{\sum w_i \sum w_i x_i^2 - \left(\sum w_i \sum w_i x_i\right)^2}}. \tag{8}$$

All of the sums in the four previous equations are over i from 1 to N.