Modern Computational Physics I (NEVF160)

Modeling of Many-Particle Systems

Štěpán Roučka

MFF UK

ZS 2022

Modeling of Many-Particle Systems

Course Outline

- Overview
- Monte-Carlo Method – Equilibrium Models, Markov Chains, Kinetic Monte-Carlo
- Molecular Dynamics Method – Integration of Equations of Motion
- Efficient Computation of Force Interaction – Tree Algorithms, Particle-in-Cell, Poisson Equation Solvers
- Continuous Modeling – Brief Overview of Basic Equations, Conditions of Applicability
- Continuous-Particle Hybrid Models – Qualitative Overview
- Program Optimization – Profiling, Vectorization, Parallelization, GPU

[wikimedia: Andromeda galaxy]

Molecular dynamics (MD) vs Monte-Carlo (MC)

Molecular dynamics

- System evolution according to equations of motion
- Thermal interaction with the environment - complicated
- Disparate time scales - complicated - e.g. rare process on the background of fast oscillations of the crystal lattice
- Calculation of forces demanding

Monte-Carlo

- Generates random system configurations and calculates their probability of realization
- Ideal for calculating mean thermal values
- Allows the study of "slow" processes
- Can also study dynamics

Model systems

Classical continuous system

- N particles described by continuous coordinates (e.g. $\boldsymbol{r}^N$, $\boldsymbol{p}^N$) in phase space
- Potential energy typically given by pair interactions

$$U = \sum_{i<j} u^{(2)}(\boldsymbol{r}_i, \boldsymbol{r}_j)$$

- But we also work with multi-particle interactions (e.g. induced dipole moment of molecules) and anisotropic potentials (e.g. molecules or planets...)
- For example gravitational/Coulombic interaction, hard sphere potential, or Lennard-Jones potential

$$U_{\mathrm{LJ}} = 4\epsilon \left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6} \right]$$

Classical continuous system- examples

Star cluster

- Isolated system – microcanonical ensemble, NVE
- Highly non-equilibrium (separated stars due to kinetic energy)
- Suitable description by molecular dynamics

Molecule/cluster in vacuum

- Isolated system – microcanonical ensemble, NVE
- Equilibrium – all configurations with energy E have the same probability

$$\pi(\boldsymbol{r}^N, \boldsymbol{p}^N) = \text{const.} \times \delta(H(\boldsymbol{r}^N, \boldsymbol{p}^N) - E)$$

- Again: MD more suitable - in MC it is a problem to conserve energy

Classical continuous system- examples

Molecule in liquid/gas

- Energy exchange with surroundings – thermalization – canonical ensemble, NVT
- Equilibrium - density π given by Boltzmann distribution

$$\pi(\boldsymbol{r}^N, \boldsymbol{p}^N) = \frac{\exp(-\beta H(\boldsymbol{r}^N, \boldsymbol{p}^N))}{Z(\beta)}$$

  all configurations with energy E have the same probability
- Molecular dynamics and Monte-Carlo can be applied
- Thermostat required in MD

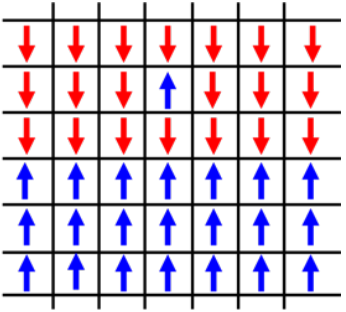
Molecule in liquid, chemical processes

- Exchange of energy and particles with the environment – grandcanonical ensemble, μVT
- Equilibrium - density π again given by Boltzmann distribution
- Molecular dynamics and Monte-Carlo can be applied

Model systems

Classical lattice model

- n particles described by discrete coordinates
- For example, the Ising model of ferromagnetism - see later
- Thin film kinetics - also later
- Lattice models of gases, etc. . .



Model systems

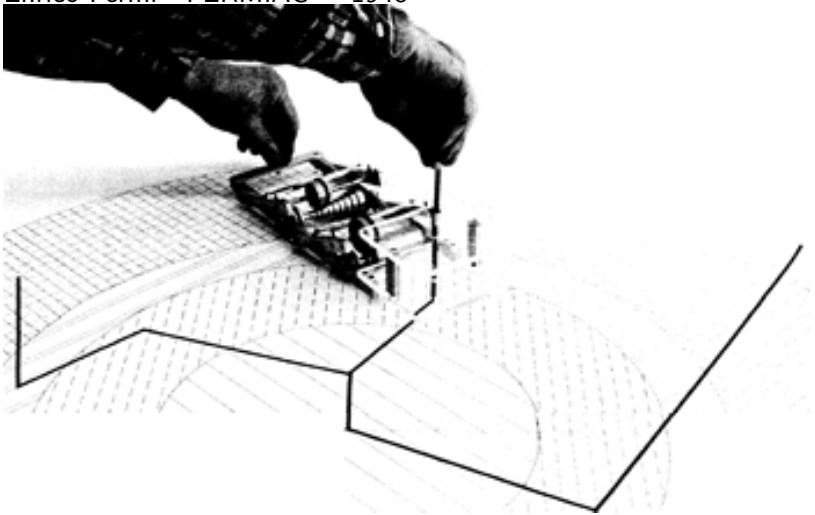
Optimization tasks

- non-physical applications
- Traveling salesman, processor design, internet searches, etc. . .

Basics of Monte Carlo Method

The origins of the Monte Carlo method

Enrico Fermi - FERMIAC $\sim$ 1948



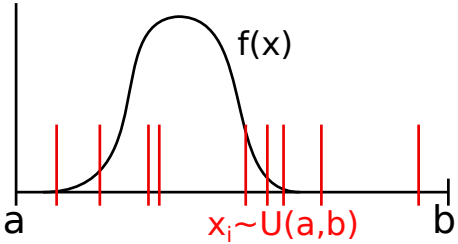
Monte Carlo (MC) - Examples

- MC Principle – parameter estimation using random sampling.

Notation

- $x_i \sim X$ random variable x_i from a distribution X
- $U(a, b)$ uniform distribution on an interval $[a, b]$
- $\mathcal{N}(\mu, \sigma^2)$ Normal distribution
- $E_{x\sim X}[f(x)] = E_X[f(x)]$ Mean $f(x)$ under distribution X

Example: mean of a function



Actual value

$$E_{U(a,b)}[f(x)] = \frac{1}{b-a}\int_a^b f(x)\mathrm{d}x$$

Estimate from n samples

$$\bar{f}_n = \frac{1}{n}\sum_{i=1}^{n} f(x_i)$$

MC models - sampling problem

- Usually we are interested in thermal mean values - we use MC integration.
- The vast majority of randomly generated configurations are extremely improbable
- It is necessary to preferentially sample "interesting regions" - importance sampling

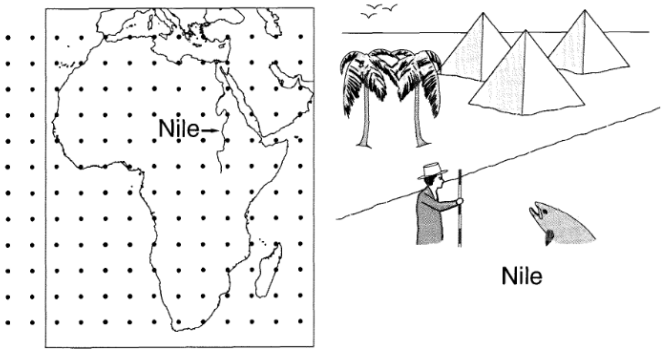


Figure 3.1: Measuring the depth of the Nile: a comparison of conventional quadrature (left), with the Metropolis scheme (right).

[Frenkel & Smit: Understanding Molecular Simulation]

MC error estimation

The investigated quantity

$$\mathrm{E}_p[f(x)] = \int f(x)p(x)\,\mathrm{d}x$$

MC approximation (consistent, unbiased...)

$$\bar{f}_n = \frac{1}{n}\sum_{i=1}^{n} f(x_i) \qquad x_i \sim p$$

MC approximation variance

$$\mathrm{Var}[\bar{f}_n] = \mathrm{Var}\left[\frac{1}{n}\sum_{i=1}^{n} f(x_i)\right] = \frac{1}{n^2}\sum_{i=1}^{n}\mathrm{Var}[f(x_i)] = \frac{1}{n}\mathrm{Var}[f(x)]$$

Importance Sampling - Reduging the MC error

It is advisable to reduce the variance of the sampled quantity $f(x)$, because

$$\mathrm{Var}[\bar{f}_n] = \frac{1}{n}\mathrm{Var}[f(x)]$$

This can be achieved by choosing a different initial distribution than p:

$$\mathrm{E}_p[f(x)] = \int f(x)p(x)\frac{h(x)}{h(x)}\mathrm{d}x = \mathrm{E}_h\left[f(x)\frac{p(x)}{h(x)}\right]$$

If $p = U$ (uniform distribution), then h should satisfy

- $h \propto |f(x)|$
- $f(x) \neq 0 \Rightarrow h > 0$
- Asymptotically, $h \ll |f(x)|$

How to generate $x \sim f$?

MCMC - Monte Carlo Markov Chains

So far: Ordinary Monte Carlo (OMC) – directly generated *independent* x_i
Now - Markov chains - sequences of random values
We denote $X_1, X_2, X_3, \ldots$ random variables; $x_1, x_2, x_3, \ldots$ specific values

- They have no memory:
  $\Pr(X_{n+1} = x_{n+1} | X_n = x_n, \ldots, X_1 = x_1) = \Pr(X_{n+1} = x_{n+1} | X_n = x_n)$
- Determined by the initial distribution $\Pr(X_1 = x_1)$ and the transition matrix
  $\Pr(X_{n+1} = x_{n+1} | X_n = x_n)$
- We denote $\Pr(X_{n+1} = x' | X_n = x) = P_n(x \to x')$
- Especially for discrete random variables we can denote
  $P_n(x_i \to x_j) = P_n(i \to j) = P^{(n)}_{i \to j} = P^{(n)}_{ij}$

static

- $P_n(x \to x')$ does not depend on n
- X_n samples the equilibrium distribution

dynamic

- $P_n(x \to x')$ may depend on n
- X_n depends on n - study of dynamics

Static MCMC

Static MCMC

It is often difficult to create a generator e.g. from the distribution under investigation Π
It is easier to create a transition matrix that preserves Π
The distribution X_n then converges to Π
Conditions:

- Stationarity
  $\Pi(x)P(x \to x') = \Pi(x')P(x' \to x)$
  (detailed equilibrium, microreversibility...)
- Ergodicity - aperiodic, recurrent, irreducible (connected)

Discrete case

- $\Pi_i P_{ij} = \Pi_j P_{ji}$
  $\to$We calculate the equations via i:
- $\sum_i \Pi_i P_{ij} = \Pi_j \sum_j P_{ji}$. LHS: matrix product, RHS: stochastic matrix
- $\boldsymbol{\Pi P} = \boldsymbol{\Pi}$. The stationary state is eigenvector of transition matrix

Metropolis-Hastings algorithm I

How to create a transition matrix that preserves Π?
The stationarity condition can be written as

$$\frac{P(x \to x')}{P(x' \to x)} = \frac{\Pi(x')}{\Pi(x)}$$

We define g and A as

$$P(x \to x') = \underbrace{g(x \to x')}_{\text{proposal}} \underbrace{A(x \to x')}_{\text{acceptance}}$$

The stationarity condition is then

$$\frac{A(x \to x')}{A(x' \to x)} = \underbrace{\frac{\Pi(x')}{\Pi(x)} \frac{g(x \to x')}{g(x' \to x)}}_{r(x,x')\text{–Hastingsův poměr}}$$

And it can be satisfied independently of g if we choose

$$A(x \to x') = \min\{1; r(x, x')\}$$

Metropolis-Hastings algorithm II

One step of the MH algorithm is then

- We generate x' from $g(x \to x')$
- Transition to state x' with probability $A(x \to x')$

Advantage of MH algorithm: we do not need to know the normalized $\Pi(x)$, just the ratio $\Pi(x')/\Pi(x)$

Example: Boltzmann distribution $\Pi(x) = \frac{1}{Z}\exp(-\frac{E(x)}{kT})$
$Z = \sum_x \exp(-\frac{E(x)}{kT})$ is difficult to determine for a large system, but

$$\frac{\Pi(x')}{\Pi(x)} = \exp\left(-\frac{E(x') - E(x)}{kT}\right)$$

is easy

Metropolis-Hastings algorithm III

A special choice g: $g(x \to x') = g(x' \to x)$ is called the Metropolis algorithm.

- Typically $x' = x + e; \quad e \sim \mathcal{N}(0, \sigma)$
- It is appropriate to choose σ so that $\mathcal{N}(x, \sigma) \approx \Pi$
- It can be diagnosed by the ratio of rejected attempts:
  - $\bar{A} \to 1$: σ too small - slow movement in the state space
  - $\bar{A} \to 0$: σ too large - in one step we skip the whole Π
- Ideally $A \approx 0.23$ for a multidimensional Gaussian distribution Π

Boltzmann distribution:

$$A(x \to x') = \min\left\{1; \exp\left(-\frac{\Delta E}{kT}\right)\right\}$$

Example: Ising model I

"Simple,, model of ferromagnetism

- Spins $\sigma \in \{-1; 1\}$ defined on a lattice with N elements
- A parallel configuration of neighboring spins and at the same time parallel to the external field H is advantageous:

$$\mathcal{H} = -J \sum_{\langle i,j \rangle} \sigma_i \sigma_j - H \sum_{i}^{N} \sigma_i$$

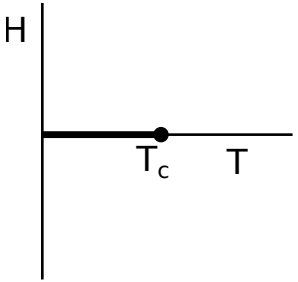
$\langle i, j \rangle$ denotes nearest neighbor pairs, $J > 0$ interacting constant

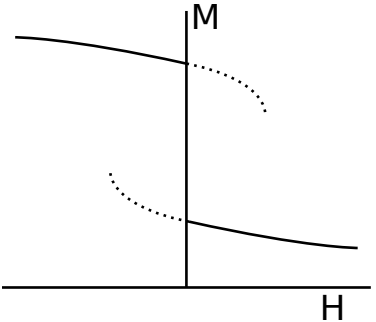
Example: Ising model II

"Simple„ model of ferromagnetism

$$\mathcal{H} = -J \sum_{\langle i,j \rangle} \sigma_i \sigma_j - H \sum_{i}^{N} \sigma_i$$

- Contains phase transition
- Metastable states
- Magnetization $M = \sum \sigma_i$





Example: Ising model III

MH simulation with Gibbs sampling algorithm - in each step we manipulate only one element of the N-dimensional state vector - spin flipping.
Test model for

- Phase transitions
- Metastable states
- Models of condensation and evaporation of liquids

MCMC error and convergence

Variance of correlated data:

$$\mathrm{Var}_\Pi[\bar{f}_n(x)] = \sigma^2/n$$

$$\sigma^2 = \mathrm{Var}_\Pi[f(X_1)] + 2\sum_{i=1}^{\infty} \mathrm{cov}_\Pi[f(X_1), f(X_i)]$$

Batch means of length $b\ldots$
Convergence issue - hard

- Possibility of pseudoconvergence
- Batch means must be uncorrelated
- Different starting positions can be tried x_1
- Other sampling
- It is reliable to wait...

Example - Google PageRank

A fictitious Google bot randomly crawls the entire website. At each step

- with 85% probability it will go to a random link
- with 15% probability. goes to a completely random unrelated page.

Markov chain - equilibrium distribution is PageRank (probability of visiting a page)
Can be solved matrix-wise - Searched PageRank x is an eigenvector of the transition matrix A,

$$x = Ax$$

http://www.mathworks.com/company/newsletters/news_notes/clevescorner/oct02_cleve.html

Simulated annealing

Simulated annealing

The idea of using MCMC to minimize complex functions

- Kirkpatrick et al. 1983 and Vladimír Černý 1985 (Kom. Uni. Blava)
- Inspiration by annealing - heating and subsequent cooling of the material leads to relaxation - minimization of free energy.
- Classical "Boltzmann" algorithm for finding the ground state of a physical system - we use the MH algorithm with a high temperature and gradually "cool"
- Transition to dynamic MCMC
- The system should end up in the ground state at $T = 0$.

DOI:10.1007/BF00940812 Černý 1985, Traveling salesman

DOI:10.1126/science.220.4598.671 Kirkpatrick et al. 1983, Chip design

Simulated annealing - algorithm

Optimized quantity E, Transition probability $g_T(s \to s')$, acceptance probability $A_T(\Delta E)$, schedule $T(k)$, system state s

- $s = s_0$
- for $k = 0 : k_{\max}$
  - $T = T(k)$
  - $s_{\text{proposal}} \sim g(s \to s')$
  - if $A_T(E(s_{\text{proposal}}) - E(s)) > u \sim U(0, 1)$
    - $s = s_{\text{proposal}}$

Simulated Annealing - parameters

Classical algorithm - Boltzmann annealing, imitates physics

- $A_T \approx \exp(-\Delta E/T)$
- $g_T = (2\pi T)^{-D/2} \exp(-\Delta x^2/2T)$

Convergence is guaranteed for

$$T(k) = \frac{T_0}{\ln k}$$

Sometimes used

$$T(k) = T_0 \exp(-\epsilon k)$$

simulated quenching (? quenching), result undefined - bad

Fast Simulated Annealing - parameters

Cauchy jump distribution with large tails - faster sampling of the entire parameter space

- $A_T \approx \exp(-\Delta E/T)$
- $g_T = \frac{T}{(\Delta x^2 + T^2)^{(D+1)/2}}$

Convergence is guaranteed for

$$T(k) = \frac{T_0}{k}$$

There is a very fast annealing (Ingber, JMCM, 1989)

$$T(k) = T_0 \exp(-\epsilon k)$$

and many other variants DOI:10.1016/0895-7177(93)90204-C Ingber 1993

Simulated Annealing - Applications

- Physical models, searching for the ground state of the system - statistical physics, materials
- General optimization, traveling salesman problem - processor interconnection design etc...

Kinetic MCMC

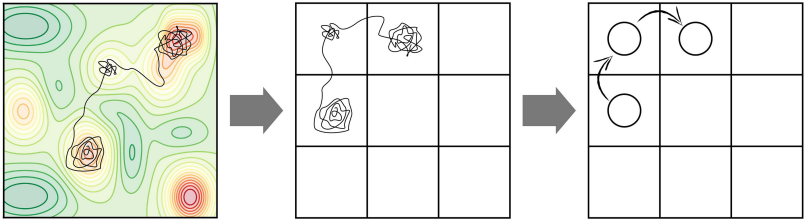
Kinetic/Dynamic MC

- Distribution is not in equilibrium: $\Pi \to \Pi_n$
- Typically continuous time: Markov chain $(n) \to$ Markov process (t)
- Described by a master equation

$$\frac{\mathrm{d}\pi_j(t)}{\mathrm{d}t} = \sum_i \pi_i(t) R_{ij} - \sum_j \pi_j(t) R_{ji}$$

- If the initial system had continuous physical states, we must discretize the state space (for example, continuous motion of an atom $\to$ jumps on a lattice.

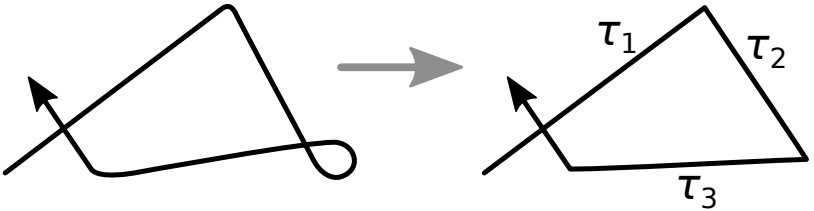
KMC Example 1 - surface motion of atoms



Andersen et al. 2019 DOI: 10.3389/fchem.2019.00202

- Transition from MD calculation to PES $\rightarrow$ identification of stationary states and transition rates (elements R_{ij}) $\rightarrow$ KMC simulation.
- E.g.: lattice $100 \times 100 \rightarrow 10000$ elements of the state vector $\rightarrow 2^{10000} \approx 2 \times 10^{3010}$ possible configurations and $(2^{10000})^2 \approx 4 \times 10^{6020}$ elements of the transition matrix
- Further simplification is necessary - locality - e.g. in each step only one atom moves by at most one position $\rightarrow$ number of non-zero elements $\boldsymbol{R} \sim$ number of atoms.
- Locality + Symmetry $\rightarrow$ The transition probability depends only on the state in the nearest neighborhood - only e.g. 6 different values R_{ij}

KMC Example 2 - MCC - Monte Carlo Collisions in gas simulations



- Weakly bound plasma - continuous motion and interaction with the surroundings can be separated into free motion in the mean potential and occasional collisions.
- Molecular chaos assumption - particle positions are uncorrelated $\rightarrow$ collision probability depends only on velocities and local concentration
- ex: $X + B \xrightarrow{k_{ij}} C + D; \qquad R_{ij} = k_{ij}[B] = 1/\tau$

KMC - implementation

- Time distribution between events EXP(R_{ij}), $f(\tau_{ij}) = R_{ij}\exp(-R_{ij}\tau)$
- We generate $\tau = -(1/R_{ij})\ln(\xi), \xi \sim U(0,1)$
- Can be simulated as a discrete Markov process with probability of events $1 - \exp(-R_{ij}\Delta t)$
- Different calculation procedures:
  - Variable time step $\tau \sim$ EXP(w)
    - Direct method (τ_{total} and reaction selection 2RN), First reaction method (min$\{\tau_k\}$ n RN), Next reaction method (all τ in memory),
  - Fixed short time step, each reaction occurs with $\Pr \approx 1 - \exp(-w\Delta t)$
  - Fixed long time step τ (τ-leaping), each reaction occurs k-times in time step $\Pr(k) = \frac{(w\tau)^k}{k!}\exp(-w\tau)$; $x' = x + ks$

Cao & Samuels 2009

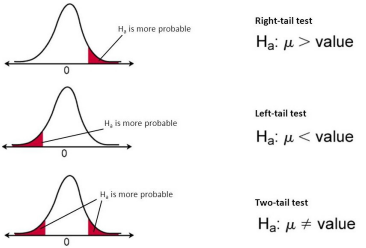
Ilie et al. 2009

MC data processing

Hypothesis Testing

Frequentist, Eg:

- Null hypothesis, H_0: data are from a certain distribution
  e.g. $x_1 \dots x_n \sim \mathcal{N}(\mu, \sigma^2)$
- Errors type I (rejection H_0) and II (acceptance H_0)
- We choose the significance level α
  probability of type I error



Test

- We calculate the statistic $\hat{\theta}$ from our data
- We generate random data according to the hypothesis H_0 and calculate their distribution statistic θ
- For the distribution θ we determine the quantiles (α, $1-\alpha$, or $\alpha/2$ and $1-\alpha/2$ for left-/right-/two-sided test)
- If $\hat{\theta}$ is more extreme than the specified quantiles, we reject H_0 with a significance level of α.

Hypothesis Testing II

- Be careful with sequential testing - we cannot collect data until we get significance
- Multiple comparison problem, alpha spending, Bonferroni correction...

Bootstrap

E.g. from the measured data $x_1 \dots x_n$ we determine $\hat{\theta}$. What is the variance and confidence intervals for θ?

- We generate a bootstrap sample $x_1^* \dots x_n^*$ by random sampling with repetition from $x_1 \dots x_n$.
- From the bootstrap sample we determine $\hat{\theta}^*$.
- By repeating the procedure we determine the distribution $\hat{\theta}^*$.
- From the distribution we determine the parameters of interest.

MCMC in model fitting and Bayesian statistic

- Gaussian uncertainty of data - least squares optimization
- Arbitrary uncertainties - Maximum likelihood (MLE) -
  Likelihood function of θ: probability of data x given parameters θ:
  $\mathcal{L}(\theta|x) = f(x|\theta) = P(X = x|\theta)$
- Including prior knowledge - Posterior probability and maximum a posteriori estimated (MAP)
- Posterior probability: Bayes -

$$f(\theta|x) = \frac{f(x|\theta)f(\theta)}{f(x)} = \frac{f(x|\theta)f(\theta)}{\int f(x|\theta')f(\theta')\mathrm{d}\theta'}$$

- The normalization can be extremely difficult to calculate, but MCMC can still sample the posterior probability

Exercise tasks

1. Generate a random variable $x_i \sim 0.3\mathcal{N}(0.1, 0.1^2) + 0.7\mathcal{N}(0.5, 0.07^2)$using the Metropolis algorithm
   - Compare different step choices
2. Find the global minimum of the function $x^2 - \cos(20(\cos^8(4x) - 1))$
3. Or $\sin(1/\mathrm{abs}(x + 10^{-16}) + 100x) + \sin(10x)^2 + 0.1x^2$
   - Hungry (greedy) algorithm
   - Simulated annealing
4. Apply simulated annealing to a more complex problem
   - Finding the basic configuration of a Van der Waals bound cluster, more complex potentials can be used
   - Solving the traveling salesman problem
   - Other applications of your choice... Ising model of ferromagnetism: M-H curves; Arbitrary (PIC, fluid, hybrid) plasma model; Physical applications of machine learning; Tree algorithm for calculating force...