Simulated Annealing

Global Optimization

Sequoia Andrade

Computational Statistics, December 12th, 2023

Table of Contents

Motivation

Goal: Find the global maxima

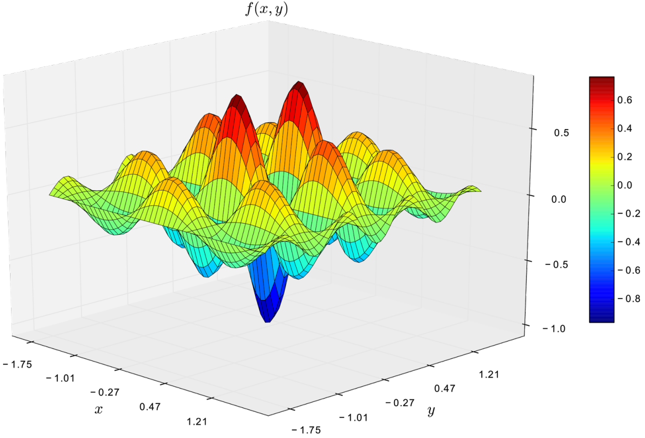


Figure: Example function

Motivation

The primary motivation of simulated annealing is to identify global maxima

- What are some methods used to identify global maxima?

Motivation

The primary motivation of simulated annealing is to identify global maxima

- What are some methods used to identify global maxima?
    - differentiation
    - gradient decent
    - brute force
    - hill climbing

Motivation

The primary motivation of simulated annealing is to identify global maxima

- What are some methods used to identify global maxima?
  - differentiation
  - gradient decent
  - brute force
  - hill climbing
- What are some limitations of these methods?

Motivation

The primary motivation of simulated annealing is to identify global maxima

- What are some methods used to identify global maxima?
  - differentiation
  - gradient decent
  - brute force
  - hill climbing
- What are some limitations of these methods?
  - need explicit function
  - stuck on local maxima
  - computational efficiency

Motivation

The primary motivation of simulated annealing is to identify global maxima

- What are some methods used to identify global maxima?
  - differentiation
  - gradient decent
  - brute force
  - hill climbing
- What are some limitations of these methods?
  - need explicit function
  - stuck on local maxima
  - computational efficiency

Simulated Annealing approximates the global maxima in cases where exact methods fail

Background

- Published my Metropolis, Rosenbluth, Rosenbluth, Teller, and Teller alongside the Metropolis-Hastings MCMC method [Metropolis et al., 1953]
- The term "annealing" comes from metallurgy and is a method to make materials more pliable by altering the **temperature**
- In the family of Monte Carlo Optimization algorithms [Robert et al., 1999]
- Can be used to solve the traveling salesman problem [Wikipedia contributors, 2023]

Simulated Annealing: Idea

Recall: What are the key ideas of the MH algorithm?

Simulated Annealing: Idea

Recall: What are the key ideas of the MH algorithm?

- Monte Carlo method - involves random sampling
- Uses adaptive acceptance/rejection method to move towards samples of higher probability
- Lower probability states are still achievable

Simulated Annealing: Idea

Recall: What are the key ideas of the MH algorithm?

- Monte Carlo method - involves random sampling
- Uses adaptive acceptance/rejection method to move towards samples of higher probability
- Lower probability states are still achievable

Simulated Annealing follows the same general idea:

- Randomly sample proposal locations
- Accept or reject proposal based on acceptance probability, which is a function of the current conditions
- Move towards areas of higher probability, leads to convergence

Simulated Annealing: Idea

The variable that controls the acceptance probability is called **temperature**

Simulated Annealing: Idea

The variable that controls the acceptance probability is called **temperature**

- Inspired by molecule movement

Simulated Annealing: Idea

The variable that controls the acceptance probability is called **temperature**

- Inspired by molecule movement
- High temperature allows the simulation to jump around to different states more freely

Simulated Annealing: Idea

The variable that controls the acceptance probability is called **temperature**

- Inspired by molecule movement
- High temperature allows the simulation to jump around to different states more freely
- Lower temperature results in smaller jumps around the state space

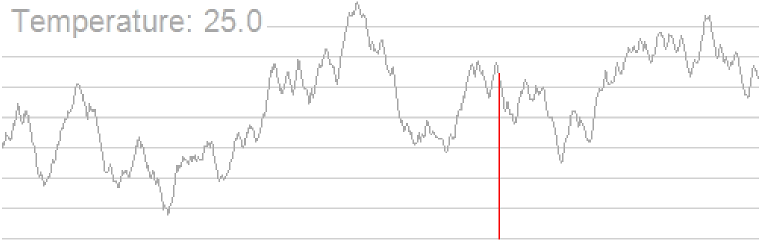
Simulated Annealing: Idea

The variable that controls the acceptance probability is called **temperature**

- Inspired by molecule movement
- High temperature allows the simulation to jump around to different states more freely
- Lower temperature results in smaller jumps around the state space
- Temperature **decreases** over each iteration, resulting in convergence

Simulated Annealing: Idea

Simple example



[Wikipedia contributors, 2023]

Notation

- i - ith iteration
- n - total iterations
- $f(x)$ - function to optimize
- T_i - temperature, a function of i
- X_i - state at step i
- u_i - proposal state at step i
- $A(X_i, X_{i+1}, T_i)$ - acceptance probability to move from state i to state i+1 given temperature

Simulated Annealing: Notation

Notation

- i - ith iteration
- n - total iterations
- $f(x)$ - function to optimize
- T_i - temperature, a function of i
- X_i - state at step i
- u_i - proposal state at step i
- $A(X_i, X_{i+1}, T_i)$ - acceptance probability to move from state i to state i+1 given temperature

Definition

$$A(X_i, X_{i+1}, T_i) = \begin{cases} 1 \text{ if } f(x_{i+1}) \geq f(x_i) \\ e^{\frac{f(x_{i+1}) - f(x_i)}{T_i}} \end{cases}$$

Simulated Annealing: Algorithm

Algorithm

for i in (1:n) **do**
  $u_i \sim U(a, b)$
  $p_i \sim U(0, 1)$
  if $f(u_i) \geq f(x_i)$ **then**
    $x_{i+1} = u_i$
  else
    $A(x_i, u_i) = min(e^{\frac{f(u_i)-f(x_i)}{T_i}}, 1)$
    if $p_i < A(x_i, u_i)$ **then**
      $x_{i+1} = u_i$
    else
      $x_{i+1} = x_i$
    end if
  end if
  $T_{i+1} = T(i + 1)$
end for

Simulated Annealing: Algorithm

Algorithm

for i in (1:n) **do**
  $u_i \sim U(a, b)$
  $p_i \sim U(0, 1)$
  if $f(u_i) \geq f(x_i)$ **then**
    $x_{i+1} = u_i$
  else
    $A(x_i, u_i) = min(e^{\frac{f(u_i)-f(x_i)}{T_i}}, 1)$
    if $p_i < A(x_i, u_i)$ **then**
      $x_{i+1} = u_i$
    else
      $x_{i+1} = x_i$
    end if
  end if
  $T_{i+1} = T(i + 1)$
end for

- Where (a, b) is the support of $f(x)$
- Note that (a, b) can be specified to be a neighborhood, r, around a given x_i, e.g., $a_i = max(a, x_i - r), b_i = min(b, x_i + r)$

Theory: Temperature

Consider the acceptance probability, $A(x_i, x_{i+1}) = e^{\frac{f(x_{i+1}) - f(x_i)}{T_i}}$

- What happens if the temperature is very large?

Theory: Temperature

Consider the acceptance probability, $A(x_i, x_{i+1}) = e^{\frac{f(x_{i+1}) - f(x_i)}{T_i}}$

- What happens if the temperature is very large?
    - The difference between the proposal and current is dominated by the temperature $\rightarrow$ more likely to accept lower values

Theory: Temperature

Consider the acceptance probability, $A(x_i, x_{i+1}) = e^{\frac{f(x_{i+1}) - f(x_i)}{T_i}}$

- What happens if the temperature is very large?
  - The difference between the proposal and current is dominated by the temperature $\rightarrow$ more likely to accept lower values
- What happens if the temperature is small?

Theory: Temperature

Consider the acceptance probability, $A(x_i, x_{i+1}) = e^{\frac{f(x_{i+1}) - f(x_i)}{T_i}}$

- What happens if the temperature is very large?
  - The difference between the proposal and current is dominated by the temperature $\rightarrow$ more likely to accept lower values
- What happens if the temperature is small?
  - The difference between the proposal and current must be more significant $\rightarrow$ less likely to accept lower values

Theory: Temperature

Consider the acceptance probability, $A(x_i, x_{i+1}) = e^{\frac{f(x_{i+1})-f(x_i)}{T_i}}$

- What happens if the temperature is very large?
  - The difference between the proposal and current is dominated by the temperature $\rightarrow$ more likely to accept lower values
- What happens if the temperature is small?
  - The difference between the proposal and current must be more significant $\rightarrow$ less likely to accept lower values

Example temperature functions include:

- $T(i) = 1/log(i)$
- $T(i+1) = T_i/\alpha$

Theory: Acceptance Probability

The probability of a specified end state is related to the Boltzman distribution and is given by:

$$P(x) = \frac{e^{f(x)/T}}{NC}$$

Where $NC = \sum_{s \in X} e^{f(s)/T}$
Hence, the state, x^*, which maximizes f is the most probable end state.

Theory: Acceptance Probability

The probability of a specified end state is related to the Boltzman distribution and is given by:

$$P(x) = \frac{e^{f(x)/T}}{NC}$$

Where $NC = \sum_{s \in X} e^{f(s)/T}$
Hence, the state, x^*, which maximizes f is the most probable end state.

When considering the acceptance probability, $A(x_i, x_{i+1}) = e^{\frac{f(x_{i+1}) - f(x_i)}{T_i}}$:

- Moves to new states are directly related to the probability of the end state
- If $f(x_{i+1})$ is much smaller than $f(x_i)$, the acceptance probability is much lower

Theory: Convergence Theorem

It can be proven that Simulated Annealing converges for the discrete state space [Robert et al., 1999]

Definition (Hajec, 1988)

Given a state space S and a function f to be maximized, let O be the set of local maxima on S, O^* be the set of global maxima on S.
Define:

- a state s_j that *can be reached at altitude h* from state s_i if there exists a sequence of states, $s_1, ...s_n$ linking s_i and s_j such that $h(s_k) \geq h$ for $k = 1, ..., n$.
- the *height of maximum* s_i is the largest d_i such that there exists a state s_j where $f(s_j) > f(s_i)$ can be reached at altitude $h(s_i) + d_i$ from s_i.

.

Theory: Convergence Theorem

It can be proven that Simulated Annealing converges for the discrete state space [Robert et al., 1999]

Theorem (Hájec, 1988)

Consider a system in which it is possible to link two arbitrary states by a finite sequence of states. If, for every $h > 0$ and every pair (s_i, s_j), s_i can be reached at altitude h from s_j if and only if s_j can be reached at altitude h from s_i, and if (T_i) decreases toward 0, the sequence X_i defined by simulated annealing satisfies:

$$\lim_{i \to \infty} P(x_i \in O^*) = 1$$

if and only if

$$\sum_{i=1}^{\infty} exp(-D/T_i) = +\infty$$

With $D = mind_i : s_i \in O - O^*$

Theory: Convergence Theorem

It can be proven that Simulated Annealing converges for the discrete state space [Robert et al., 1999]

Theorem (Hájec, 1988)

Consider a system in which it is possible to link two arbitrary states by a finite sequence of states. If, for every $h > 0$ and every pair (s_i, s_j), s_i can be reached at altitude h from s_j if and only if s_j can be reached at altitude h from s_i, and if (T_i) decreases toward 0, the sequence X_i defined by simulated annealing satisfies:

$$\lim_{i \to \infty} P(x_i \in O^*) = 1$$

if and only if

$$\sum_{i=1}^{\infty} exp(-D/T_i) = +\infty$$

With $D = mind_i : s_i \in O - O^*$

Hence, the simulated annealing algorithm converges to the global maxima so long as the temperature decreases according to the constraint on D.

Simulated Annealing: Example

Example 1: Find the global maximum of $h(x) = [cos(50x) + sin(20x)]^2$ using Simulated Annealing.

(example 5.5 in Monte Carlo Statistical Methods textbook)

Example 2: Using Simulated Annealing, find the global minimum of:

$$h(x, y) = (xsin(20y) + ysin(20x))^2 cosh(sin(10x)x) + (xcos(10y) - ysin(10x))^2 cosh(cos(20y)y)$$

(example 5.3 in Monte Carlo Statistical Methods textbook)

References

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., and Teller, E. (1953).
Equation of state calculations by fast computing machines.
The journal of chemical physics, 21(6):1087–1092.

Robert, C. P., Casella, G., and Casella, G. (1999).
Monte Carlo statistical methods, volume 2.
Springer.

Wikipedia contributors (2023).
Simulated annealing — Wikipedia, the free encyclopedia.
[Online; accessed 11-December-2023].