ΕΛΛΗΝΙΚΗ ΔΗΜΟΚΡΑΤΙΑ
ΠΑΝΕΠΙΣΤΗΜΙΟ ΚΡΗΤΗΣ

Εισαγωγή σε μεθόδους Monte Carlo

Ενότητα 5: Η βελτιστοποίηση Monte Carlo

Βαγγέλης Χαρμανδάρης

Τμήμα Μαθηματικών και Εφαρμοσμένων Μαθηματικών

Optimization

❑ Basic Definitions:

- Differences between the numerical approach and the simulation approach to the problem
$$\max_{\theta \in \Theta} h(\theta)$$
lie in the treatment of the function h.

- In an optimization problem using deterministic numerical methods
  - The analytical properties of the target function (convexity, boundedness, smoothness) are often paramount.

- For the simulation approach
  - We are concerned with h from a probabilistic (rather than analytical) point of view.

Optimization: The Problem

- The problem

$$\max_{\theta \in \Theta} h(\theta)$$

- Deterministic numerical methods $\longrightarrow$ analytical properties

- Simulation approach $\longrightarrow$ probabilistic view.

    - This dichotomy is somewhat artificial
    - Some simulation approaches have no probabilistic interpretation

- Nonetheless, the use of the analytical properties of h plays a lesser role in the simulation approach.

Two Simulation Approaches

- Exploratory Approach
    - Goal: To optimize h by describing its entire range
    - Actual properties of h play a lesser role
- Probabilistic Approach
    - Monte Carlo exploits probabilistic properties of h
    - This approach tied to missing data methods

Stochastic Exploration

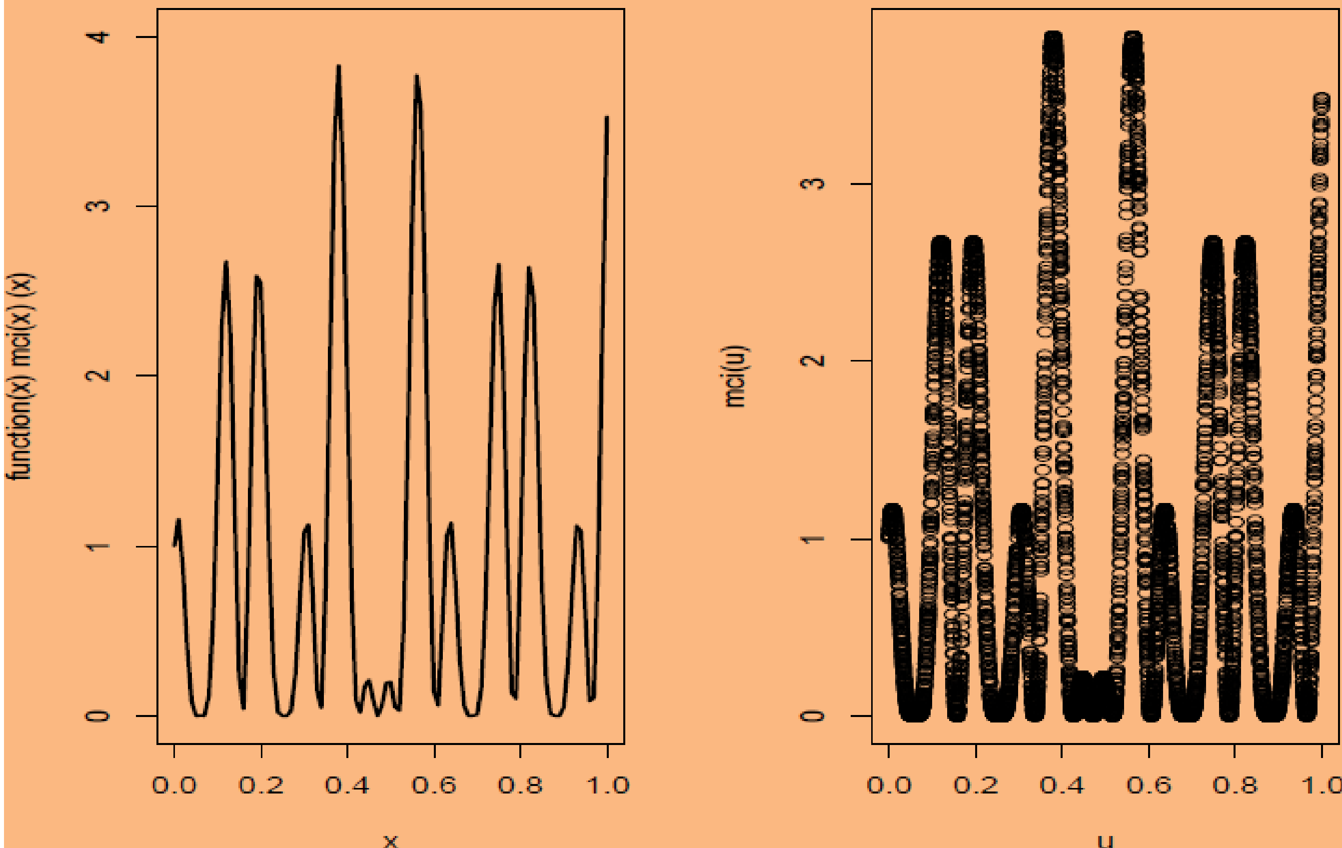
- A first approach is to simulate from a uniform distribution on Θ, $u_1, \ldots, u_m \sim \mathcal{U}_\Theta$,
- Use the approximation

$$h_m^* = \max(h(u_1), \ldots, h(u_m)).$$

- This method converges (as m goes to ∞), but it may be very slow since it does not take into account any specific feature of h.
- Distributions other than the uniform, which can possibly be related to h, may then do better.
- In particular, in setups where the likelihood function is extremely costly to compute the number of evaluations of the function h is best kept to a minimum.

Monte Carlo Maximization: A First Example

- Recall the function $h(x) = [\cos(50x) + \sin(20x)]^2$.
- we try our naïve strategy and simulate $u_1, \ldots, u_m \sim \mathcal{U}(0,1)$, and use the approximation $h^*_m = \max(h(u_1), \ldots, h(u_m))$

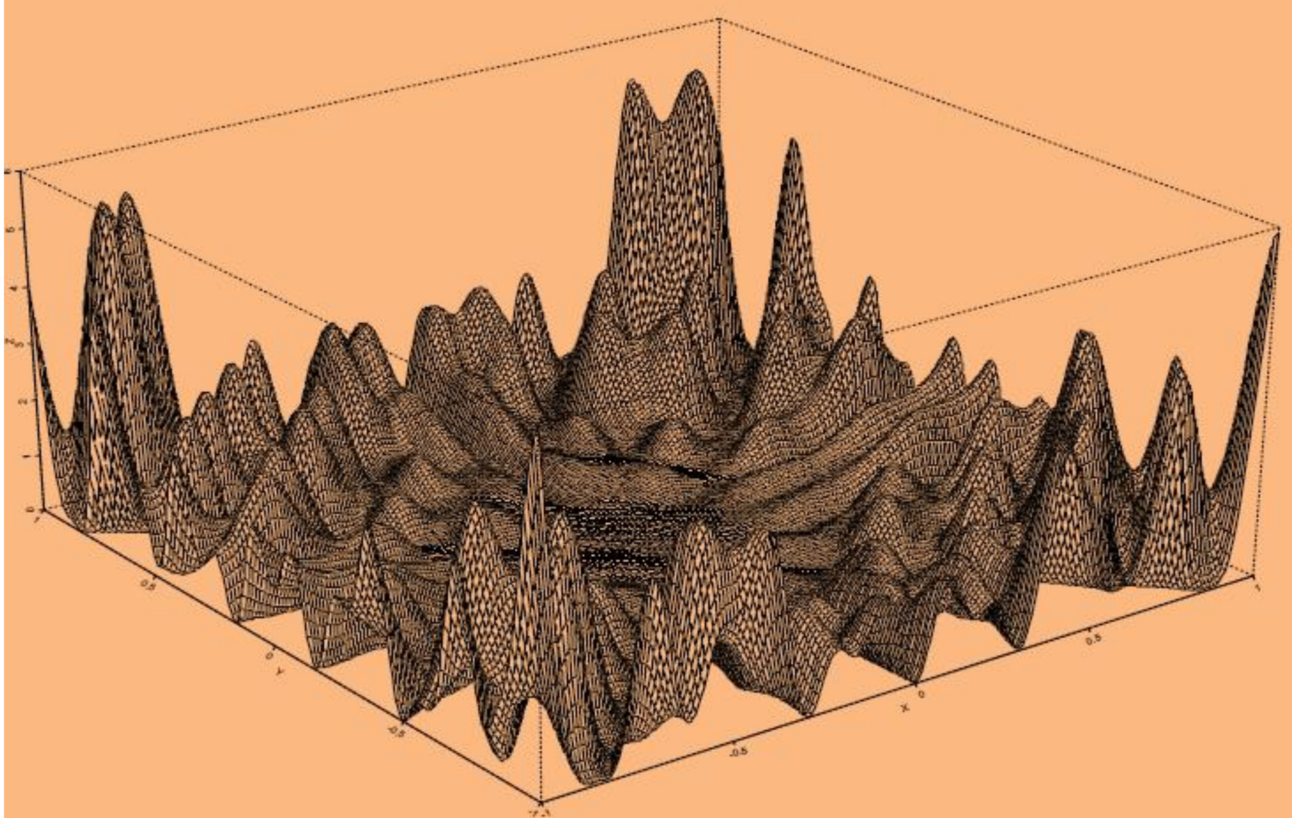


Monte Carlo: A Tough Minimization

- Consider minimizing

$$
\begin{aligned}
h(x,y) &= (x\sin(20y)+y\sin(20x))^2\cosh(\sin(10x)x) \\
&\quad + (x\cos(10y)-y\sin(10x))^2\cosh(\cos(20y)y)\ ,
\end{aligned}
$$

whose global minimum is 0, attained at $(x,y)=(0,0)$



Monte Carlo: A Tough Minimization

Properties

- Many local minima
- Standard methods may not find global minimum
- We can simulate from $\exp(-h(x,y))$
- Get minimum from $\min_i h(x_i, y_i)$
- Can use other methods...

Monte Carlo: A Tough Minimization

Deterministic Gradient Methods

- The *gradient method* is a deterministic numerical approach to the problem
$$\max_{\theta \in \Theta} h(\theta).$$
- It produces a sequence (θ_j) that converges to the maximum when
- ○ the domain $\Theta \subset \Re^d$
  ○ the function $(-h)$

  are both convex.
- The sequence (θ_j) is constructed in a recursive manner through
$$\theta_{j+1} = \theta_j + \alpha_j \nabla h(\theta_j) \ , \qquad \alpha_j > 0 \ ,$$

Here

○ ∇h is the gradient of h

○ α_j is chosen to aid convergence

Monte Carlo: A Tough Minimization

Stochastic Variant

- There are stochastic variants of the gradient method
- They do not always go along the steepest slope
- This is an advantage, as it can avoid local maxima and saddlepoints
- The best, and simple version is Simulated Annealing/Metropolis Algorithm

Simulated Annealing

- This name is borrowed from Metallurgy:
- A metal manufactured by a slow decrease of temperature (*annealing*) is stronger than a metal manufactured by a fast decrease of temperature.
- The fundamental idea of simulated annealing methods is that a change of scale, called *temperature*, allows for faster moves on the surface of the function h to maximize.
    - Rescaling partially avoids the trapping attraction of local maxima.
    - As T decreases toward 0, the values simulated from this distribution become concentrated in a narrower and narrower neighborhood of the local maxima of h

Simulated Annealing/Metropolis Algorithm

- **Basic Idea:**

- Simulation method proposed by Metropolis *et al.* (1953)
- Starting from θ_0, ζ is generated from

$$\zeta \sim \text{Uniform in a neighborhood of } \theta_0.$$

- The new value of θ is generated as

$$\theta_1 = \begin{cases} \zeta & \text{with probability } \rho = \exp(\Delta h/T) \wedge 1 \\ \theta_0 & \text{with probability } 1 - \rho, \end{cases}$$

  - $\Delta h = h(\zeta) - h(\theta_0)$
  - If $h(\zeta) \geq h(\theta_0)$, ζ is accepted
  - If $h(\zeta) < h(\theta_0)$, ζ may still be accepted
  - This allows escape from local maxima

Simulated Annealing/Metropolis Algorithm

Metropolis/Simulated Annealing Algorithm

- In its most usual implementation, the simulated annealing algorithm modifies the temperature T at each iteration
- It has the form

1. Simulate ζ from an instrumental distribution with density $g(|\zeta - \theta_i|)$;
2. Accept $\theta_{i+1} = \zeta$ with probability

$$\rho_i = \exp\{\Delta h_i / T_i\} \wedge 1;$$

   take $\theta_{i+1} = \theta_i$ otherwise.

3. Update T_i to T_{i+1}.

Simulated Annealing/Metropolis Algorithm

- **Important Comments:**

1. Simulate ζ from an instrumental distribution with density $g(|\zeta - \theta_i|)$;
2. Accept $\theta_{i+1} = \zeta$ with probability
$$\rho_i = \exp\{\Delta h_i / T_i\} \wedge 1;$$
take $\theta_{i+1} = \theta_i$ otherwise.
3. Update T_i to T_{i+1}.

- All positive moves accepted
- As $T \downarrow 0$
  - Harder to accept downward moves
  - No big downward moves
- Not a Markov Chain - difficult to analyze

Simulated Annealing/Metropolis Algorithm

❑ **Simple Example (as before):**

- Recall the function $h(x) = [\cos(50x) + \sin(20x)]^2$
- The specific algorithm we use is

  Starting at iteration t, the iteration is at $(x^{(t)}, h(x^{(t)}))$:

  1. Simulate $u \sim \mathcal{U}(a_t, b_t)$ where $a_t = \max(x^{(t)} - r, 0)$ and $b_t = \min(x^{(t)} + r, 1)$
  2. Accept $x^{(t+1)} = u$ with probability

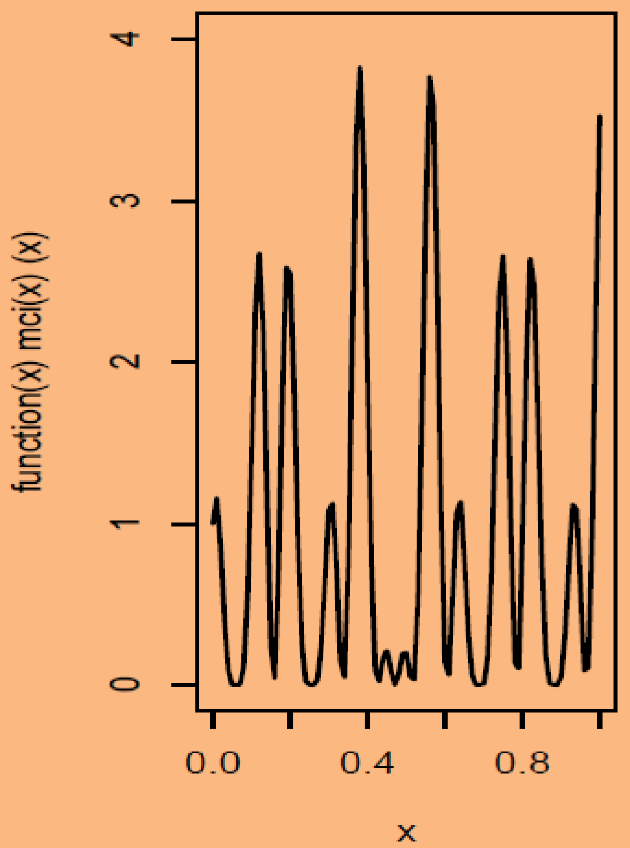
  $$\rho^{(t)} = \min\left\{\exp\left(\frac{h(u) - h(x^{(t)})}{T_t}\right), 1\right\},$$

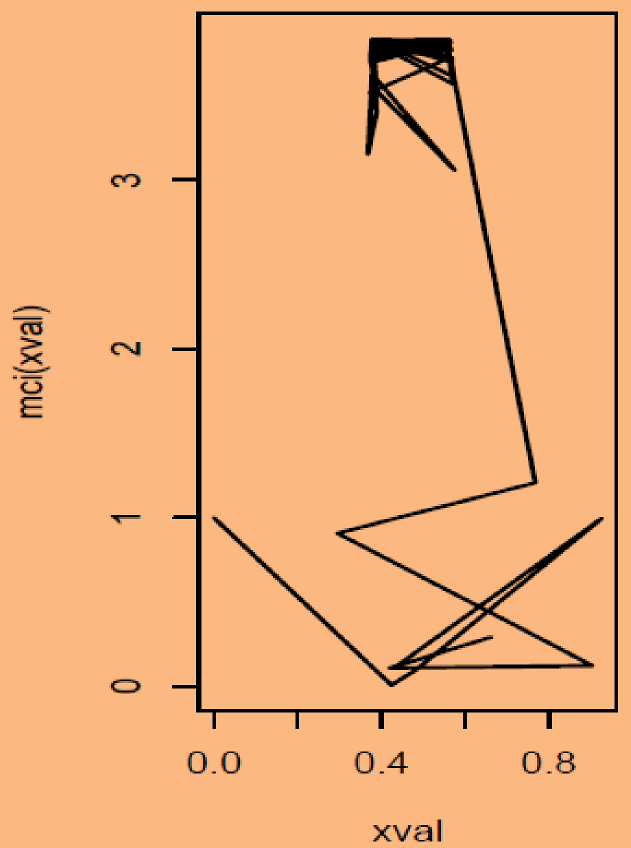
     take $x^{(t+1)} = x^{(t)}$ otherwise.
  3. Update T_t to T_{t+1}.

- ◦ The value of r controls the size of the interval around the current point (staying in $(0, 1)$)
- ◦ The value of T_t controls the cooling.

Simulated Annealing/Metropolis Algorithm

The Trajectory





- Left Panel is the function
- Right Panel is the Simulated Annealing Trajectory

Simulated Annealing

Simulated Annealing Property

- Theorem 5.7: Under mild assumptions, the Simulated Annealing algorithm is guaranteed to find the global maximum

❑ **Maximization problem (as before):**

Return to the difficult maximization

- Apply simulated Annealing
- Different choices of T_i
  - Results dependent on choice of T_i
  - $T_i \propto 1/\log(i+1)$ preferred

Simulated Annealing: Examples

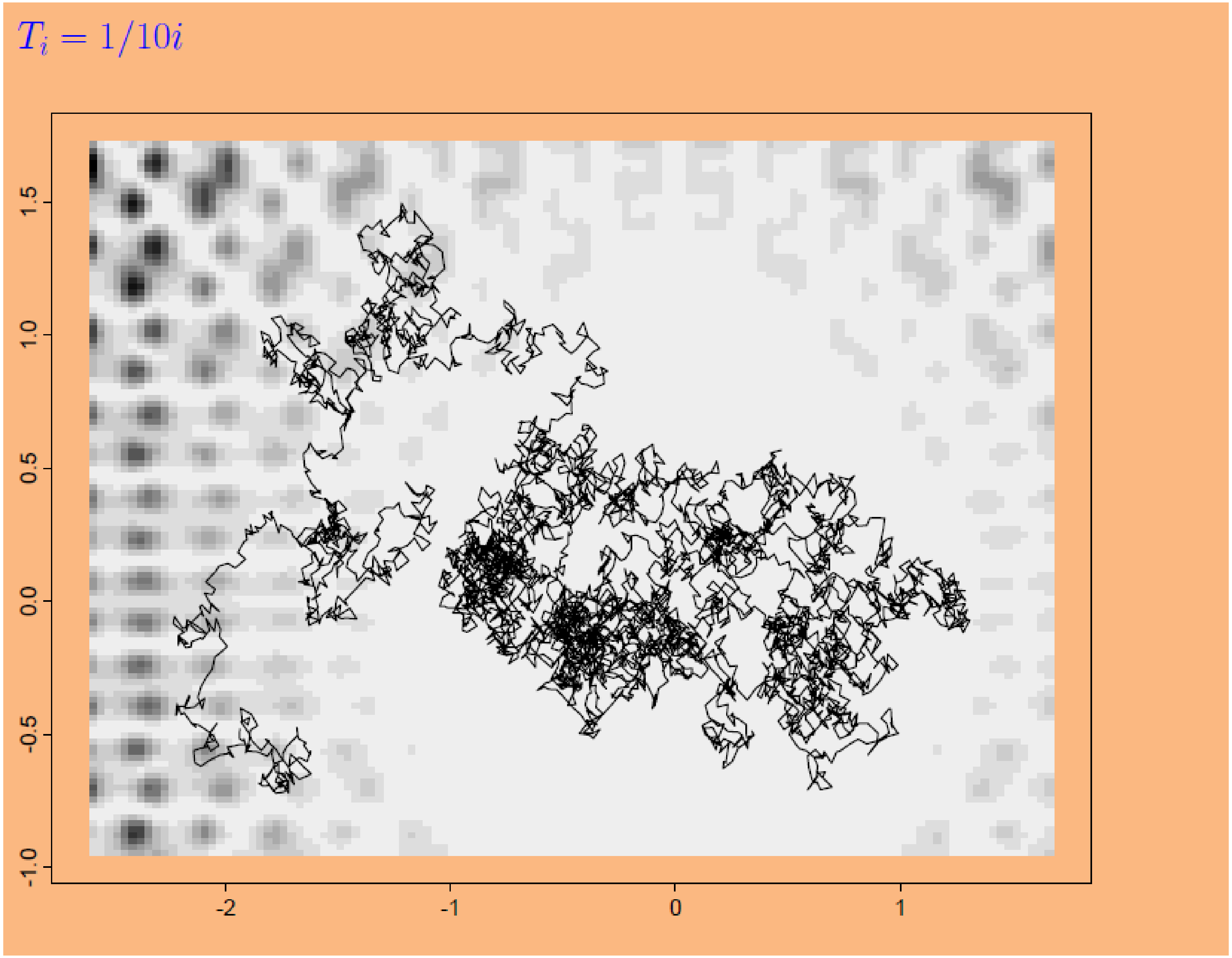
Simulated Annealing Runs

- $g \sim \text{Uniform}(-.1, .1)$
- Starting point $(0.5, 0.4)$

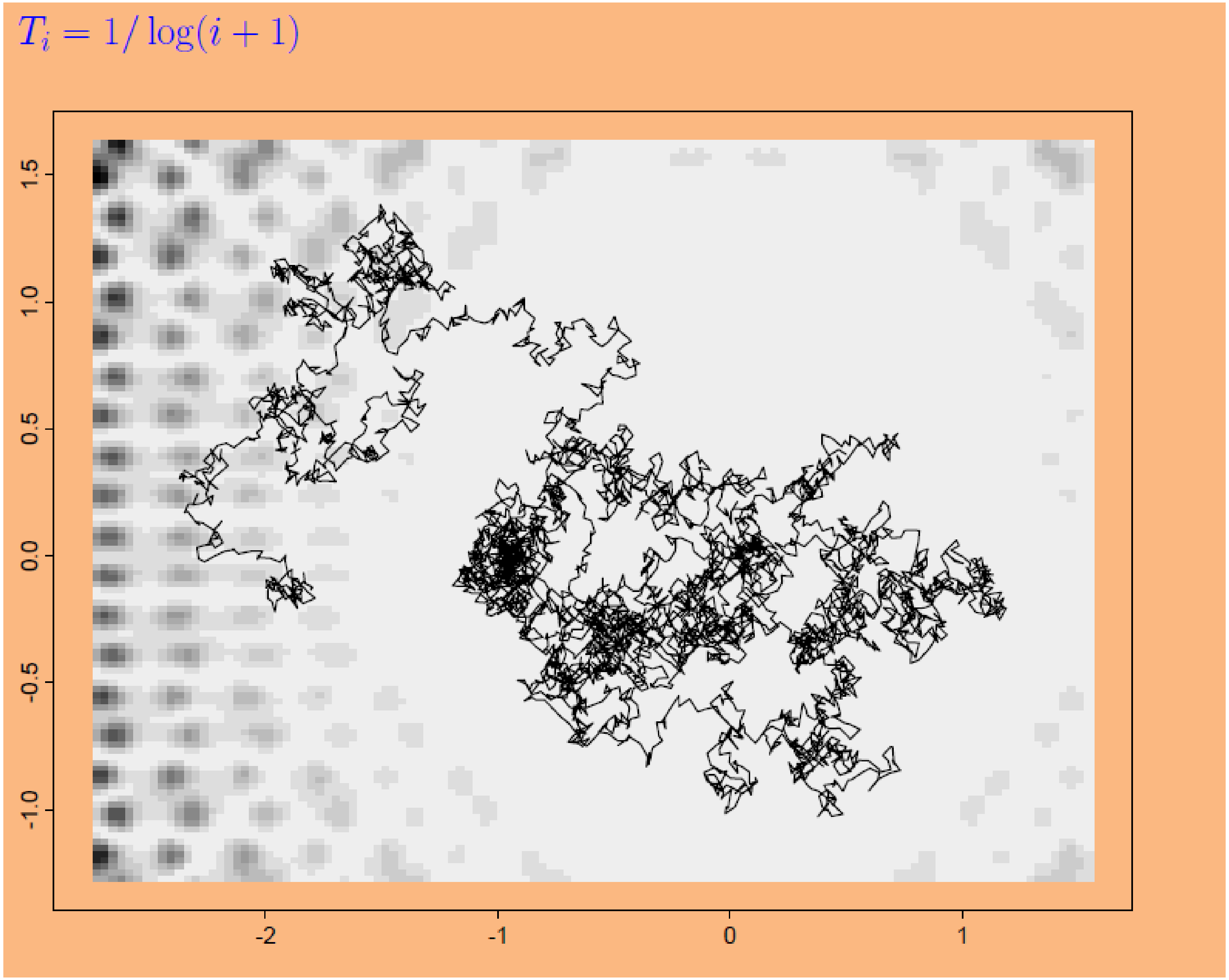
Case	T_i	θ_T	$h(\theta_T)$	$\min_t h(\theta_t)$	Accept. rate
1	$1/10i$	$(-1.94, -0.480)$	0.198	$4.02\,10^{-7}$	0.9998
2	$1/\log(1+i)$	$(-1.99, -0.133)$	3.408	3.823×10^{-7}	0.96
3	$100/\log(1+i)$	$(-0.575, 0.430)$	0.0017	4.708×10^{-9}	0.6888
4	$1/10\log(1+i)$	$(0.121, -0.150)$	0.0359	2.382×10^{-7}	0.71

- Case 3 explores "valley" near the minimum
- Recommended $T_i \approx \Gamma/\log(i+1)$ for large Γ

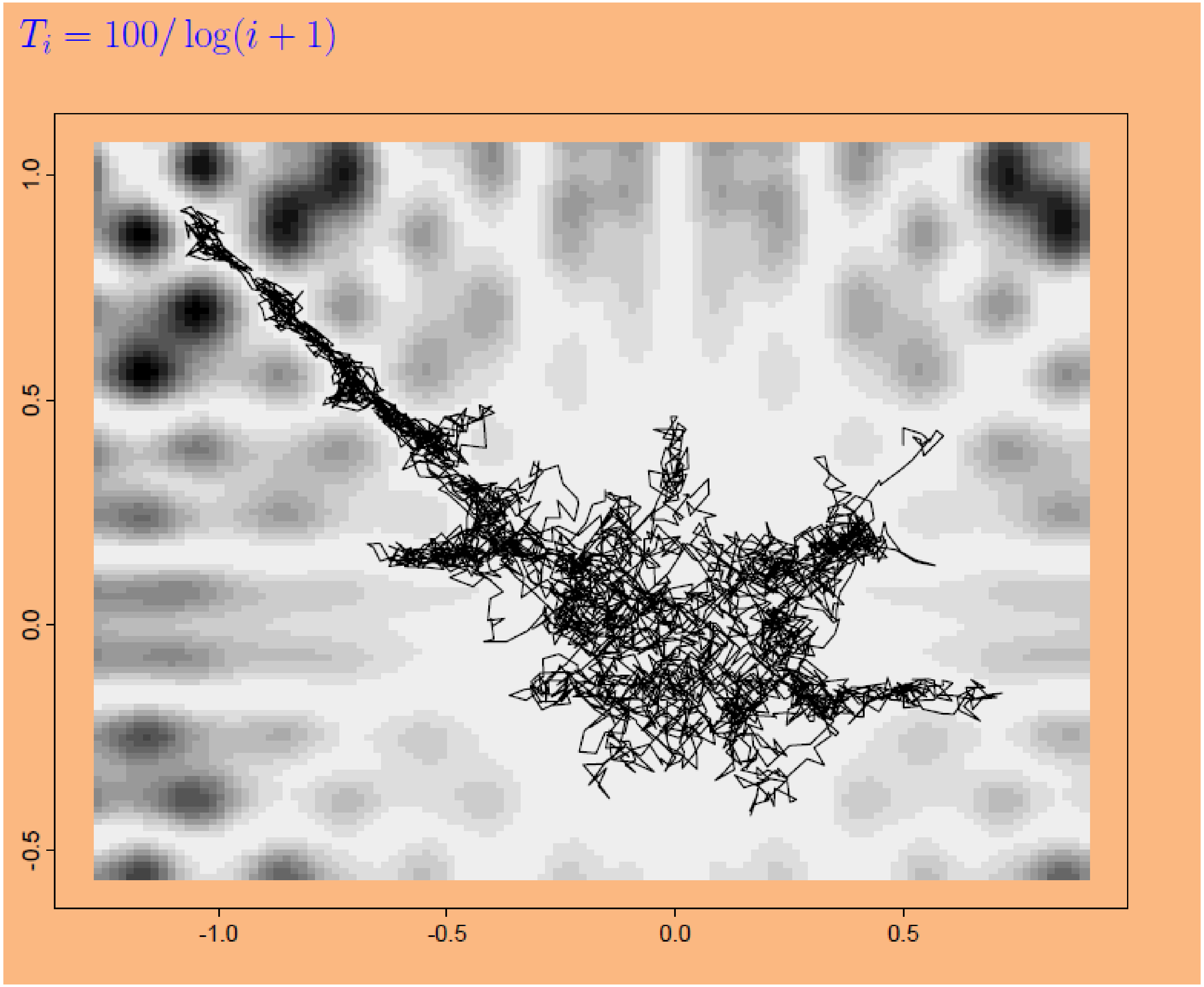
Simulated Annealing: Examples



Simulated Annealing: Examples



Simulated Annealing: Examples



Bibliography

- *Monte Carlo Strategies in Scientific Computing*, J. Liu, Springer, New York, 2001.
- *Monte Carlo Statistical Methods*, C. Robert, G. Casella, Springer, New York, 2004.
- *Stochastic Methods: A Handbook for the Natural and Social Sciences,* C. Gardiner, Springer, New York, 2009.
- *A Guide to Monte Carlo Simulations in Statistical Physics*, D. Landau, K. Binder, Cambridge, New York, 2009.

Τέλος Ενότητας

Ευρωπαϊκή Ένωση
Ευρωπαϊκό Κοινωνικό Ταμείο
ΕΠΙΧΕΙΡΗΣΙΑΚΟ ΠΡΟΓΡΑΜΜΑ
ΕΚΠΑΙΔΕΥΣΗ ΚΑΙ ΔΙΑ ΒΙΟΥ ΜΑΘΗΣΗ
επένδυση στην κοινωνία της γνώσης
ΥΠΟΥΡΓΕΙΟ ΠΑΙΔΕΙΑΣ & ΘΡΗΣΚΕΥΜΑΤΩΝ, ΠΟΛΙΤΙΣΜΟΥ & ΑΘΛΗΤΙΣΜΟΥ
ΕΙΔΙΚΗ ΥΠΗΡΕΣΙΑ ΔΙΑΧΕΙΡΙΣΗΣ
ΕΣΠΑ 2007-2013
πρόγραμμα για την ανάπτυξη
ΕΥΡΩΠΑΪΚΟ ΚΟΙΝΩΝΙΚΟ ΤΑΜΕΙΟ
Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης

Χρηματοδότηση

- Το παρόν εκπαιδευτικό υλικό έχει αναπτυχθεί στα πλαίσια του εκπαιδευτικού έργου του διδάσκοντα.
- Το έργο «**Ανοικτά Ακαδημαϊκά Μαθήματα στο Πανεπιστήμιο Κρήτης**» έχει χρηματοδοτήσει μόνο τη αναδιαμόρφωση του εκπαιδευτικού υλικού.
- Το έργο υλοποιείται στο πλαίσιο του Επιχειρησιακού Προγράμματος «Εκπαίδευση και Δια Βίου Μάθηση» και συγχρηματοδοτείται από την Ευρωπαϊκή Ένωση (Ευρωπαϊκό Κοινωνικό Ταμείο) και από εθνικούς πόρους.

Ευρωπαϊκή Ένωση
Ευρωπαϊκό Κοινωνικό Ταμείο
ΕΠΙΧΕΙΡΗΣΙΑΚΟ ΠΡΟΓΡΑΜΜΑ
ΕΚΠΑΙΔΕΥΣΗ ΚΑΙ ΔΙΑ ΒΙΟΥ ΜΑΘΗΣΗ
επένδυση στην κοινωνία της γνώσης
ΥΠΟΥΡΓΕΙΟ ΠΑΙΔΕΙΑΣ ΚΑΙ ΘΡΗΣΚΕΥΜΑΤΩΝ
ΕΙΔΙΚΗ ΥΠΗΡΕΣΙΑ ΔΙΑΧΕΙΡΙΣΗΣ
ΕΣΠΑ 2007-2013
πρόγραμμα για την ανάπτυξη
ΕΥΡΩΠΑΪΚΟ ΚΟΙΝΩΝΙΚΟ ΤΑΜΕΙΟ
Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης

Σημειώματα

Σημείωμα αδειοδότησης

.

Σημείωμα Αναφοράς

Copyright Πανεπιστήμιο Κρήτης, Βαγγέλης Χαρμανδάρης 2015. «Εισαγωγή σε μεθόδους Monte Carlo. Ενότητα 5: Η βελτιστοποίηση Monte Carlo». Έκδοση: 1.0. Ηράκλειο 2015. Διαθέσιμο από τη δικτυακή διεύθυνση: https://opencourses.uoc.gr/courses/course/view.php?id=228.

Διατήρηση Σημειωμάτων

Οποιαδήποτε αναπαραγωγή ή διασκευή του υλικού θα πρέπει να συμπεριλαμβάνει:

- το Σημείωμα Αναφοράς
- το Σημείωμα Αδειοδότησης
- τη δήλωση Διατήρησης Σημειωμάτων
- το Σημείωμα Χρήσης Έργων Τρίτων (εφόσον υπάρχει)

μαζί με τους συνοδευόμενους υπερσυνδέσμους.