Towards an Average Case Runtime Lower Bound of Simulated Annealing on TSP

Bodo Manthey Jesse van Rhijn

University of Twente

October 4, 2022

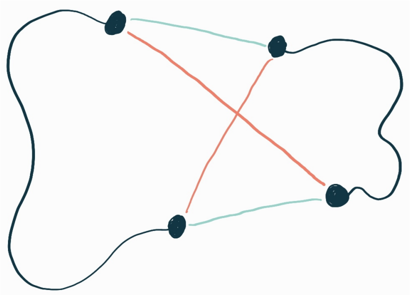
Local Search Heuristics

Many combinatorial problems are hard.

So hard, practitioners usually give up on solving them exactly.

Instead, they often use local search heuristics.

Example: 2-opt for Travelling Salesperson Problem (TSP).



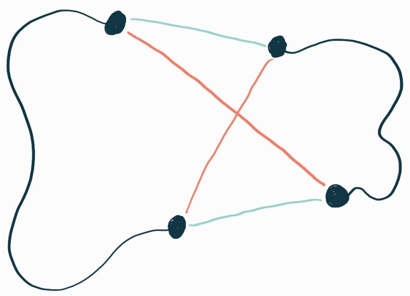
Local Search Heuristics

Many combinatorial problems are hard.

So hard, practitioners usually give up on solving them exactly.

Instead, they often use local search heuristics.

Example: 2-opt for Travelling Salesperson Problem (TSP).



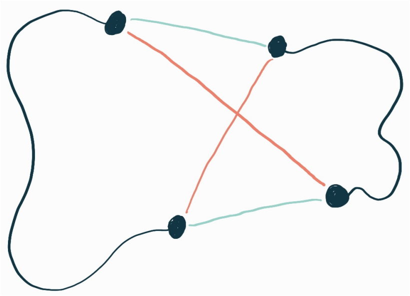
Local Search Heuristics

Many combinatorial problems are hard.

So hard, practitioners usually give up on solving them exactly.

Instead, they often use local search heuristics.

Example: 2-opt for Travelling Salesperson Problem (TSP).



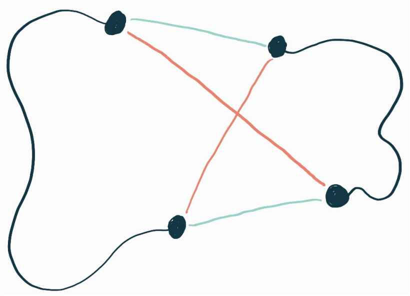
Local Search Heuristics

Many combinatorial problems are hard.

So hard, practitioners usually give up on solving them exactly.

Instead, they often use local search heuristics.

Example: 2-opt for Travelling Salesperson Problem (TSP).

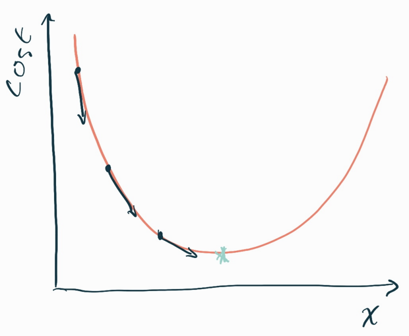


Local Improvement Heuristics

Simplest local search heuristic: local improvement.

Given solution x, select any "neighbor" $y \rightarrow$ if y is better, move to it.

- Simple to implement.
- Can be bad in worst case, but:
- Very effective in practice.

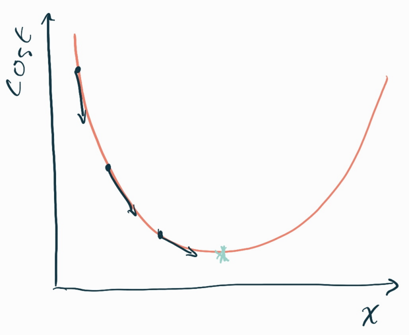


Local Improvement Heuristics

Simplest local search heuristic: local improvement.

Given solution x, select any "neighbor" $y \rightarrow$ if y is better, move to it.

- Simple to implement.
- Can be bad in worst case, but:
- Very effective in practice.

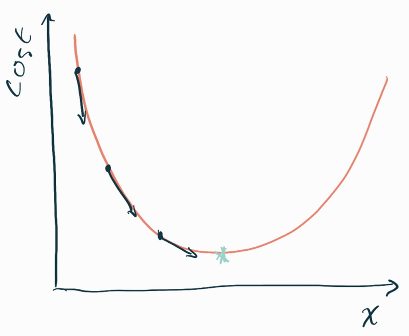


Local Improvement Heuristics

Simplest local search heuristic: local improvement.

Given solution x, select any "neighbor" $y \rightarrow$ if y is better, move to it.

- Simple to implement.
- Can be bad in worst case, but:
- Very effective in practice.

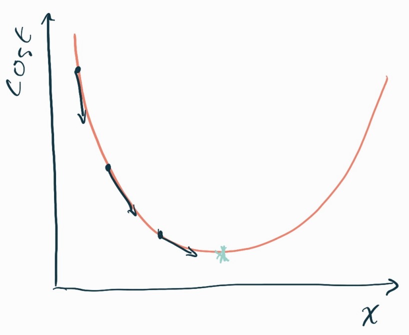


Local Improvement Heuristics

Simplest local search heuristic: local improvement.

Given solution x, select any "neighbor" $y \rightarrow$ if y is better, move to it.

- Simple to implement.
- Can be bad in worst case, but:
- Very effective in practice.

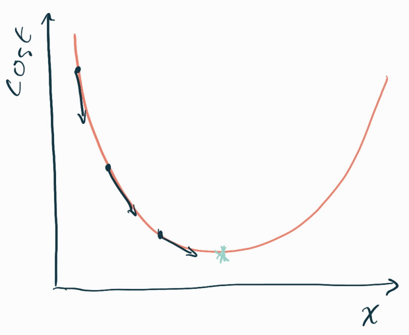


Local Improvement Heuristics

Simplest local search heuristic: local improvement.

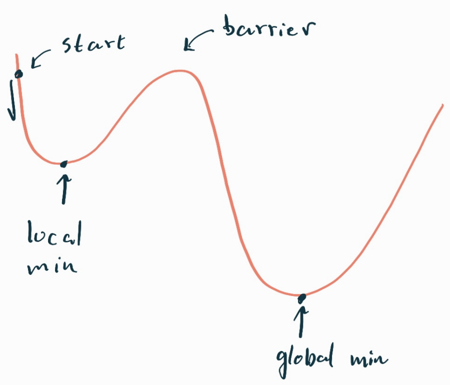
Given solution x, select any "neighbor" y $\rightarrow$ if y is better, move to it.

- Simple to implement.
- Can be bad in worst case, but:
- Very effective in practice.



Local Minima

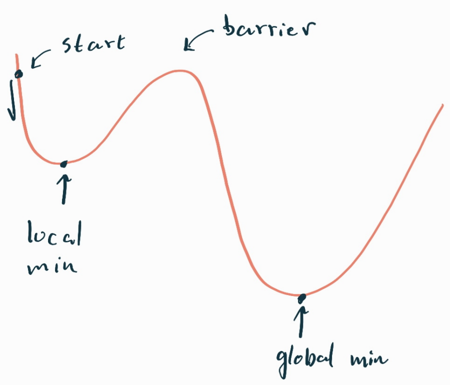
Weakness of local improvement: local minima.



Escape using metaheuristics!

Local Minima

Weakness of local improvement: local minima.



Escape using metaheuristics!

Travelling Salesperson Problem

Definition

Given a graph $G = (V, E)$ with edge weights $w : E \to [0, 1]$, the Travelling Salesperson Problem (TSP) asks for a minimum-weight Hamiltonian cycle on G.

Definition

The solution set S is the set of all Hamiltonian cycles on G.

Definition

The length $L(x \mid w)$ of tour $x \in S$ with respect to the weights $w \in [0, 1]^E$ is the sum of its edge weights, i.e. $L(x \mid w) = \sum_{e \in x} w(e)$.

Travelling Salesperson Problem

Definition

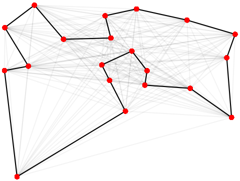
Given a graph $G = (V, E)$ with edge weights $w : E \to [0, 1]$, the Travelling Salesperson Problem (TSP) asks for a minimum-weight Hamiltonian cycle on G.

Definition

The solution set S is the set of all Hamiltonian cycles on G.

Definition

The length $L(x \mid w)$ of tour $x \in S$ with respect to the weights $w \in [0, 1]^E$ is the sum of its edge weights, i.e. $L(x \mid w) = \sum_{e \in x} w(e)$.



Travelling Salesperson Problem

Definition

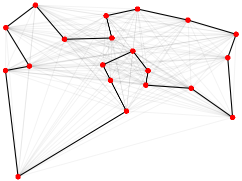
Given a graph $G = (V, E)$ with edge weights $w : E \to [0, 1]$, the Travelling Salesperson Problem (TSP) asks for a minimum-weight Hamiltonian cycle on G.

Definition

The solution set S is the set of all Hamiltonian cycles on G.

Definition

The length $L(x \mid w)$ of tour $x \in S$ with respect to the weights $w \in [0, 1]^E$ is the sum of its edge weights, i.e. $L(x \mid w) = \sum_{e \in x} w(e)$.

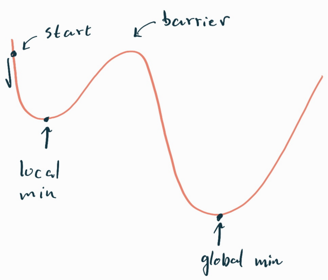


Simulated annealing: outline

Simulated annealing:
metaheuristic.

Generalizes local improvement:
allows "bad" steps.

Defines a Markov chain on S.



Parametrized by "temperature":

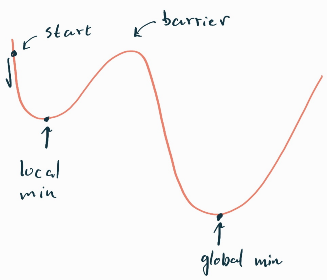
$$\mathbb{P}(\text{bad step}) = e^{-\Delta L/T} = e^{-\beta \Delta L}.$$

Simulated annealing: outline

Simulated annealing:
metaheuristic.

Generalizes local improvement:
allows "bad" steps.

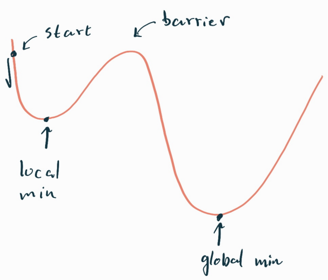
Defines a Markov chain on S.



Parametrized by "temperature":

$$\mathbb{P}(\text{bad step}) = e^{-\Delta L/T} = e^{-\beta \Delta L}.$$

Simulated annealing: outline



Simulated annealing:
metaheuristic.

Generalizes local improvement:
allows "bad" steps.

Defines a Markov chain on S.

Parametrized by "temperature":

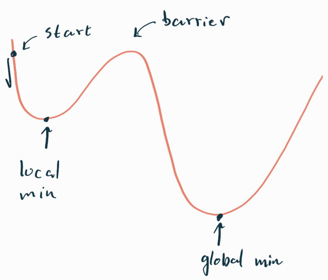
$$\mathbb{P}(\text{bad step}) = e^{-\Delta L/T} = e^{-\beta \Delta L}.$$

Simulated annealing: outline

Simulated annealing:
metaheuristic.

Generalizes local improvement:
allows "bad" steps.

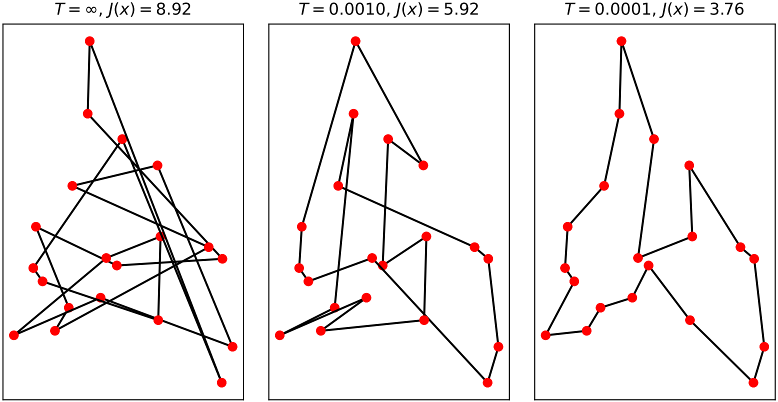
Defines a Markov chain on S.



Parametrized by "temperature":

$$\mathbb{P}(\text{bad step}) = e^{-\Delta L/T} = e^{-\beta \Delta L}.$$

Cooling a Salesperson



Simulated annealing: theoretical guarantee

Theorem (Informal (Hajek, 1989))

Simulated annealing converges to the uniform distribution on the global minima as $t \to \infty$, provided the temperature satisfies

$$T_t = \frac{a}{\log(t)},$$

where a is a problem-dependent constant.

In practice, this *cooling schedule* is too slow.

For general TSP: convergence (w.h.p.) in about $O\left(n^{n^2}\right)$.

Simulated annealing: theoretical guarantee

Theorem (Informal (Hajek, 1989))

Simulated annealing converges to the uniform distribution on the global minima as $t \to \infty$, provided the temperature satisfies

$$T_t = \frac{a}{\log(t)},$$

where a is a problem-dependent constant.

In practice, this *cooling schedule* is too slow.

For general TSP: convergence (w.h.p.) in about $O\left(n^{n^2}\right)$.

Simulated annealing: theoretical guarantee

Theorem (Informal (Hajek, 1989))

Simulated annealing converges to the uniform distribution on the global minima as $t \to \infty$, provided the temperature satisfies

$$T_t = \frac{a}{\log(t)},$$

where a is a problem-dependent constant.

In practice, this *cooling schedule* is too slow.

For general TSP: convergence (w.h.p.) in about $O\left(n^{n^2}\right)$.

Lower bound

Just how bad is this cooling schedule?

- Size of solution set in TSP $< n! \ll n^{n^2}$.
- Held-Karp solves TSP in $O^*(2^n)$.

Q: can SA with log-cooling do better than Held-Karp?

Alternatively: can we find the optimal tour with constant probability?

Spoiler: probably not.

Lower bound

Just how bad is this cooling schedule?

- Size of solution set in TSP $< n! \ll n^{n^2}$.
- Held-Karp solves TSP in $O^*(2^n)$.

Q: can SA with log-cooling do better than Held-Karp?

Alternatively: can we find the optimal tour with constant probability?

Spoiler: probably not.

Lower bound

Just how bad is this cooling schedule?

- Size of solution set in TSP $< n! \ll n^{n^2}$.
- Held-Karp solves TSP in $O^*(2^n)$.

Q: can SA with log-cooling do better than Held-Karp?

Alternatively: can we find the optimal tour with constant probability?

Spoiler: probably not.

Lower bound

Just how bad is this cooling schedule?

- Size of solution set in TSP $< n! \ll n^{n^2}$.
- Held-Karp solves TSP in $O^*(2^n)$.

Q: can SA with log-cooling do better than Held-Karp?

Alternatively: can we find the optimal tour with constant probability?

Spoiler: probably not.

Lower bound

Just how bad is this cooling schedule?

- Size of solution set in TSP $< n! \ll n^{n^2}$.
- Held-Karp solves TSP in $O^*(2^n)$.

Q: can SA with log-cooling do better than Held-Karp?

Alternatively: can we find the optimal tour with constant probability?

Spoiler: probably not.

Lower bound

Just how bad is this cooling schedule?

- Size of solution set in TSP $< n! \ll n^{n^2}$.
- Held-Karp solves TSP in $O^*(2^n)$.

Q: can SA with log-cooling do better than Held-Karp?

Alternatively: can we find the optimal tour with constant probability?

Spoiler: probably not.

Outline of Analysis

Random TSP model

Let $G = (V, E)$ be a complete graph on n vertices. Assign weights to the edges by drawing them from $\mu = U[0,1]^E$.

Theorem

For the random TSP model, the optimal tour length is $\Theta(1)$ w.h.p.

Use this model to obtain average case predictions.

Average case $\leq$ worst case, so lower bounds transfer.

Outline of Analysis

Random TSP model

Let $G = (V, E)$ be a complete graph on n vertices. Assign weights to the edges by drawing them from $\mu = U[0, 1]^E$.

Theorem

For the random TSP model, the optimal tour length is $\Theta(1)$ *w.h.p.*

Use this model to obtain average case predictions.

Average case $\leq$ worst case, so lower bounds transfer.

Outline of Analysis

Random TSP model

Let $G = (V, E)$ be a complete graph on n vertices. Assign weights to the edges by drawing them from $\mu = U[0, 1]^E$.

Theorem

For the random TSP model, the optimal tour length is $\Theta(1)$ *w.h.p.*

Use this model to obtain average case predictions.

Average case $\leq$ worst case, so lower bounds transfer.

Outline of Analysis

Random TSP model

Let $G = (V, E)$ be a complete graph on n vertices. Assign weights to the edges by drawing them from $\mu = U[0, 1]^E$.

Theorem

For the random TSP model, the optimal tour length is $\Theta(1)$ *w.h.p.*

Use this model to obtain average case predictions.

Average case $\leq$ worst case, so lower bounds transfer.

Outline of Analysis

Fix a temperature $T =: \beta^{-1}$, and draw edge weights $W \sim \mu = U[0,1]^E$.

Start from arbitrary tour $x \in S$, and run until distribution over S converges.

Stationary distribution for this random instance:

$$\pi_\beta(x \mid W) = \frac{e^{-\beta L(x \mid W)}}{\sum_{y \in S} e^{-\beta L(y \mid W)}} = \frac{e^{-\beta L(x \mid W)}}{Z(\beta \mid W)}$$

where $L(x \mid W) = \sum_{e \in x} W(e)$, the length of tour x.

Note: $\pi_\beta(x \mid W)$ is a random measure!

Outline of Analysis

Fix a temperature $T =: \beta^{-1}$, and draw edge weights $W \sim \mu = U[0,1]^E$.

Start from arbitrary tour $x \in S$, and run until distribution over S converges.

Stationary distribution for this random instance:

$$\pi_\beta(x \mid W) = \frac{e^{-\beta L(x \mid W)}}{\sum_{y \in S} e^{-\beta L(y \mid W)}} = \frac{e^{-\beta L(x \mid W)}}{Z(\beta \mid W)}$$

where $L(x \mid W) = \sum_{e \in x} W(e)$, the length of tour x.

Note: $\pi_\beta(x \mid W)$ is a random measure!

Outline of Analysis

Fix a temperature $T =: \beta^{-1}$, and draw edge weights $W \sim \mu = U[0,1]^E$.

Start from arbitrary tour $x \in S$, and run until distribution over S converges.

Stationary distribution for this random instance:

$$\pi_\beta(x \mid W) = \frac{e^{-\beta L(x \mid W)}}{\sum_{y \in S} e^{-\beta L(y \mid W)}} = \frac{e^{-\beta L(x \mid W)}}{Z(\beta \mid W)}$$

where $L(x \mid W) = \sum_{e \in x} W(e)$, the length of tour x.

Note: $\pi_\beta(x \mid W)$ is a random measure!

Outline of Analysis

Fix a temperature $T =: \beta^{-1}$, and draw edge weights $W \sim \mu = U[0,1]^E$.

Start from arbitrary tour $x \in S$, and run until distribution over S converges.

Stationary distribution for this random instance:

$$\pi_\beta(x \mid W) = \frac{e^{-\beta L(x \mid W)}}{\sum_{y \in S} e^{-\beta L(y \mid W)}} = \frac{e^{-\beta L(x \mid W)}}{Z(\beta \mid W)}$$

where $L(x \mid W) = \sum_{e \in x} W(e)$, the length of tour x.

Note: $\pi_\beta(x \mid W)$ is a random measure!

Outline of Analysis

We study the statistics of $L(x \mid W)$. A nice fact:

$$\mathbb{E}_{\pi_\beta}(L \mid W) = -\frac{\mathrm{d}}{\mathrm{d}\beta} \ln Z(\beta \mid W).$$

Proof.

$$\mathbb{E}_{\pi_\beta}(L \mid W) = \sum_{x \in S} \pi_\beta(x \mid W) L(x \mid W) = \sum_{x \in S} \frac{e^{-\beta L(x \mid W)}}{Z(\beta \mid W)} L(x \mid w)$$

$$= -\frac{1}{Z(\beta \mid W)} \frac{\mathrm{d}}{\mathrm{d}\beta} \sum_{x \in S} e^{-\beta L(x \mid W)} = -\frac{\mathrm{d}}{\mathrm{d}\beta} \ln Z(\beta \mid W).$$

□

So if we want to compute $\mathbb{E}(L)$:

$$\mathbb{E}(L) = -\frac{\mathrm{d}}{\mathrm{d}\beta} \mathbb{E}_\mu (\ln Z(\beta \mid W)).$$

Outline of Analysis

We study the statistics of $L(x \mid W)$. A nice fact:

$$\mathbb{E}_{\pi_\beta}(L \mid W) = -\frac{\mathrm{d}}{\mathrm{d}\beta} \ln Z(\beta \mid W).$$

Proof.

$$\mathbb{E}_{\pi_\beta}(L \mid W) = \sum_{x \in S} \pi_\beta(x \mid W) L(x \mid W) = \sum_{x \in S} \frac{e^{-\beta L(x \mid W)}}{Z(\beta \mid W)} L(x \mid w)$$

$$= -\frac{1}{Z(\beta \mid W)} \frac{\mathrm{d}}{\mathrm{d}\beta} \sum_{x \in S} e^{-\beta L(x \mid W)} = -\frac{\mathrm{d}}{\mathrm{d}\beta} \ln Z(\beta \mid W).$$

□

So if we want to compute $\mathbb{E}(L)$:

$$\mathbb{E}(L) = -\frac{\mathrm{d}}{\mathrm{d}\beta} \mathbb{E}_\mu \left(\ln Z(\beta \mid W) \right).$$

Outline of Analysis

We study the statistics of $L(x \mid W)$. A nice fact:

$$\mathbb{E}_{\pi_\beta}(L \mid W) = -\frac{\mathrm{d}}{\mathrm{d}\beta} \ln Z(\beta \mid W).$$

Proof.

$$\mathbb{E}_{\pi_\beta}(L \mid W) = \sum_{x \in S} \pi_\beta(x \mid W) L(x \mid W) = \sum_{x \in S} \frac{e^{-\beta L(x \mid W)}}{Z(\beta \mid W)} L(x \mid w)$$

$$= -\frac{1}{Z(\beta \mid W)} \frac{\mathrm{d}}{\mathrm{d}\beta} \sum_{x \in S} e^{-\beta L(x \mid W)} = -\frac{\mathrm{d}}{\mathrm{d}\beta} \ln Z(\beta \mid W).$$

□

So if we want to compute $\mathbb{E}(L)$:

$$\mathbb{E}(L) = -\frac{\mathrm{d}}{\mathrm{d}\beta} \mathbb{E}_\mu (\ln Z(\beta \mid W)).$$

Main Obstacle

Problem: cannot compute $\mathbb{E}_\mu(\ln Z(\beta \mid W))$.

We *can* compute $\ln \mathbb{E}_\mu(Z(\beta \mid W))$, but that is not what we have. . .

Luckily, $\mathbb{E}_\mu(Z(\beta \mid W))$ contains some useful information still.

Vague outline: define another, easier-to-analyze Markov chain related to SA.

Analyze this simpler chain instead and compare expected tour lengths.

Main Obstacle

Problem: cannot compute $\mathbb{E}_\mu(\ln Z(\beta \mid W))$.

We *can* compute $\ln \mathbb{E}_\mu(Z(\beta \mid W))$, but that is not what we have...

Luckily, $\mathbb{E}_\mu(Z(\beta \mid W))$ contains some useful information still.

Vague outline: define another, easier-to-analyze Markov chain related to SA.

Analyze this simpler chain instead and compare expected tour lengths.

Main Obstacle

Problem: cannot compute $\mathbb{E}_\mu(\ln Z(\beta \mid W))$.

We *can* compute $\ln \mathbb{E}_\mu(Z(\beta \mid W))$, but that is not what we have...

Luckily, $\mathbb{E}_\mu(Z(\beta \mid W))$ contains some useful information still.

Vague outline: define another, easier-to-analyze Markov chain related to SA.

Analyze this simpler chain instead and compare expected tour lengths.

Main Obstacle

Problem: cannot compute $\mathbb{E}_\mu(\ln Z(\beta \mid W))$.

We *can* compute $\ln \mathbb{E}_\mu(Z(\beta \mid W))$, but that is not what we have...

Luckily, $\mathbb{E}_\mu(Z(\beta \mid W))$ contains some useful information still.

Vague outline: define another, easier-to-analyze Markov chain related to SA.

Analyze this simpler chain instead and compare expected tour lengths.

Main Obstacle

Problem: cannot compute $\mathbb{E}_\mu(\ln Z(\beta \mid W))$.

We *can* compute $\ln \mathbb{E}_\mu(Z(\beta \mid W))$, but that is not what we have. . .

Luckily, $\mathbb{E}_\mu(Z(\beta \mid W))$ contains some useful information still.

Vague outline: define another, easier-to-analyze Markov chain related to SA.

Analyze this simpler chain instead and compare expected tour lengths.

Auxiliary Distribution

We define the auxiliary distribution as

$$\pi_\beta^A(x, w) = \frac{e^{-\beta L(x \mid w)} \mu(w)}{\mathbb{E}_\mu(Z(\beta \mid W))}.$$

$\rightarrow$ stationary distribution of auxiliary chain.

Let $X, W \sim \pi_\beta^A$ and $L_A = \sum_{e \in X} W(e)$. Then

$$\mathbb{E}(L_A) = -\frac{\mathrm{d}}{\mathrm{d}\beta} \ln \mathbb{E}_\mu(Z(\beta \mid W)).$$

This lower bounds the real tour length!

Auxiliary Distribution

We define the auxiliary distribution as

$$\pi_\beta^A(x, w) = \frac{e^{-\beta L(x \mid w)} \mu(w)}{\mathbb{E}_\mu(Z(\beta \mid W))}.$$

$\rightarrow$ stationary distribution of auxiliary chain.

Let $X, W \sim \pi_\beta^A$ and $L_A = \sum_{e \in X} W(e)$. Then

$$\mathbb{E}(L_A) = -\frac{\mathrm{d}}{\mathrm{d}\beta} \ln \mathbb{E}_\mu(Z(\beta \mid W)).$$

This lower bounds the real tour length!

Auxiliary Distribution

We define the auxiliary distribution as

$$\pi_\beta^A(x, w) = \frac{e^{-\beta L(x \mid w)}\mu(w)}{\mathbb{E}_\mu(Z(\beta \mid W))}.$$

$\rightarrow$ stationary distribution of auxiliary chain.

Let $X, W \sim \pi_\beta^A$ and $L_A = \sum_{e \in X} W(e)$. Then

$$\mathbb{E}(L_A) = -\frac{\mathrm{d}}{\mathrm{d}\beta} \ln \mathbb{E}_\mu(Z(\beta \mid W)).$$

This lower bounds the real tour length!

Auxiliary Distribution

We define the auxiliary distribution as

$$\pi_\beta^A(x, w) = \frac{e^{-\beta L(x \mid w)} \mu(w)}{\mathbb{E}_\mu(Z(\beta \mid W))}.$$

$\rightarrow$ stationary distribution of auxiliary chain.

Let $X, W \sim \pi_\beta^A$ and $L_A = \sum_{e \in X} W(e)$. Then

$$\mathbb{E}(L_A) = -\frac{\mathrm{d}}{\mathrm{d}\beta} \ln \mathbb{E}_\mu(Z(\beta \mid W)).$$

This lower bounds the real tour length!

Summarizing the Strategy

To summarize: we define

$$\pi_\beta(x \mid W) = \frac{e^{-\beta L(x \mid W)}}{\sum_{y \in S} e^{-\beta L(y \mid W)}} = \frac{e^{-\beta L(x \mid W)}}{Z(\beta \mid W)},$$

and also

$$\underbrace{\pi_\beta^P(x, w) = \pi_\beta(x \mid w)\mu(w)}_{\text{primary distribution}} \quad \text{and} \quad \underbrace{\pi_\beta^A(x, w) = \pi_\beta(x \mid w)\nu_\beta(w)}_{\text{auxiliary distribution}},$$

where $\nu_\beta(w) = \frac{\mu(w)Z(\beta \mid w)}{\mathbb{E}_\mu(Z(\beta \mid W))}$. Then let $X_{A/P}, W_{A/P} \sim \pi_\beta^{A/P}$.

We compare the variables $L_{A/P} = \sum_{e \in X_{A/P}} W_{A/P}(e)$.

Summarizing the Strategy

To summarize: we define

$$\pi_\beta(x \mid W) = \frac{e^{-\beta L(x \mid W)}}{\sum_{y \in S} e^{-\beta L(y \mid W)}} = \frac{e^{-\beta L(x \mid W)}}{Z(\beta \mid W)},$$

and also

$$\underbrace{\pi_\beta^P(x, w) = \pi_\beta(x \mid w)\mu(w)}_{\text{primary distribution}} \quad \text{and} \quad \underbrace{\pi_\beta^A(x, w) = \pi_\beta(x \mid w)\nu_\beta(w)}_{\text{auxiliary distribution}},$$

where $\nu_\beta(w) = \frac{\mu(w)Z(\beta \mid w)}{\mathbb{E}_\mu(Z(\beta \mid W))}$. Then let $X_{A/P}, W_{A/P} \sim \pi_\beta^{A/P}$.

We compare the variables $L_{A/P} = \sum_{e \in X_{A/P}} W_{A/P}(e)$.

Summarizing the Strategy

To summarize: we define

$$\pi_\beta(x \mid W) = \frac{e^{-\beta L(x \mid W)}}{\sum_{y \in S} e^{-\beta L(y \mid W)}} = \frac{e^{-\beta L(x \mid W)}}{Z(\beta \mid W)},$$

and also

$$\underbrace{\pi_\beta^P(x, w) = \pi_\beta(x \mid w)\mu(w)}_{\text{primary distribution}} \quad \text{and} \quad \underbrace{\pi_\beta^A(x, w) = \pi_\beta(x \mid w)\nu_\beta(w)}_{\text{auxiliary distribution}},$$

where $\nu_\beta(w) = \frac{\mu(w)Z(\beta \mid w)}{\mathbb{E}_\mu(Z(\beta \mid W))}$. Then let $X_{A/P}, W_{A/P} \sim \pi_\beta^{A/P}$.

We compare the variables $L_{A/P} = \sum_{e \in X_{A/P}} W_{A/P}(e)$.

Summarizing the Strategy

To summarize: we define

$$\pi_\beta(x \mid W) = \frac{e^{-\beta L(x \mid W)}}{\sum_{y \in S} e^{-\beta L(y \mid W)}} = \frac{e^{-\beta L(x \mid W)}}{Z(\beta \mid W)},$$

and also

$$\underbrace{\pi_\beta^P(x,w) = \pi_\beta(x \mid w)\mu(w)}_{\text{primary distribution}} \quad \text{and} \quad \underbrace{\pi_\beta^A(x,w) = \pi_\beta(x \mid w)\nu_\beta(w)}_{\text{auxiliary distribution}},$$

where $\nu_\beta(w) = \frac{\mu(w)Z(\beta \mid w)}{\mathbb{E}_\mu(Z(\beta \mid W))}$. Then let $X_{A/P}, W_{A/P} \sim \pi_\beta^{A/P}$.

We compare the variables $L_{A/P} = \sum_{e \in X_{A/P}} W_{A/P}(e)$.

Summarizing the Strategy

To summarize: we define

$$\pi_\beta(x \mid W) = \frac{e^{-\beta L(x \mid W)}}{\sum_{y \in S} e^{-\beta L(y \mid W)}} = \frac{e^{-\beta L(x \mid W)}}{Z(\beta \mid W)},$$

and also

$$\underbrace{\pi_\beta^P(x, w) = \pi_\beta(x \mid w)\mu(w)}_{\text{primary distribution}} \quad \text{and} \quad \underbrace{\pi_\beta^A(x, w) = \pi_\beta(x \mid w)\nu_\beta(w)}_{\text{auxiliary distribution}},$$

where $\nu_\beta(w) = \frac{\mu(w)Z(\beta \mid w)}{\mathbb{E}_\mu(Z(\beta \mid W))}$. Then let $X_{A/P}, W_{A/P} \sim \pi_\beta^{A/P}$.

We compare the variables $L_{A/P} = \sum_{e \in X_{A/P}} W_{A/P}(e)$.

Main Result

Theorem

For $\beta > 0$,

$$\mathbb{E}(L_P) \geq \mathbb{E}(L_A) = n\left(\frac{1}{\beta} - \frac{1}{e^\beta - 1}\right).$$

Corollary

Assuming SA is in equilibrium at iteration t, the logarithmic cooling schedule with parameter $a > 0$ yields

$$\mathbb{E}(L_P) = \Omega\left(\frac{an}{\log t}\right) \quad \text{as} \quad t \to \infty.$$

$\rightarrow$ Getting $\mathbb{E}(L) = O(1)$ requires $t = 2^{\Omega(n)}$.

Main Result

Theorem

For $\beta > 0$,

$$\mathbb{E}(L_P) \geq \mathbb{E}(L_A) = n\left(\frac{1}{\beta} - \frac{1}{e^{\beta} - 1}\right).$$

Corollary

Assuming SA is in equilibrium at iteration t, the logarithmic cooling schedule with parameter $a > 0$ yields

$$\mathbb{E}(L_P) = \Omega\left(\frac{an}{\log t}\right) \quad \textit{as} \quad t \to \infty.$$

$\to$ Getting $\mathbb{E}(L) = O(1)$ requires $t = 2^{\Omega(n)}$.

Main Result

Theorem

For $\beta > 0$,

$$\mathbb{E}(L_P) \geq \mathbb{E}(L_A) = n\left(\frac{1}{\beta} - \frac{1}{e^{\beta} - 1}\right).$$

Corollary

Assuming SA is in equilibrium at iteration t, the logarithmic cooling schedule with parameter $a > 0$ yields

$$\mathbb{E}(L_P) = \Omega\left(\frac{an}{\log t}\right) \quad \textit{as} \quad t \to \infty.$$

$\rightarrow$ Getting $\mathbb{E}(L) = O(1)$ requires $t = 2^{\Omega(n)}$.

Towards a tail bound

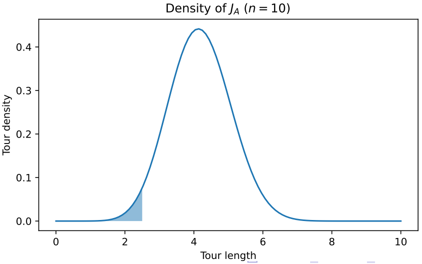
Only have a lower bound on the expected tour length.

But over many iterations, could we sometimes sample better solutions?

Want to prove a tail bound for L_P, i.e.

$$\mathbb{P}(L_P \leq j) \leq \text{(something small)}.$$

Can prove this for L_A:



Towards a tail bound

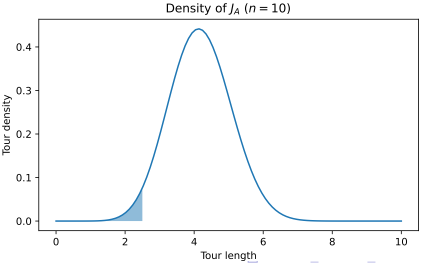
Only have a lower bound on the expected tour length.

But over many iterations, could we sometimes sample better solutions?

Want to prove a tail bound for L_P, i.e.

$$\mathbb{P}(L_P \leq j) \leq (\text{something small}).$$

Can prove this for L_A:



Towards a tail bound

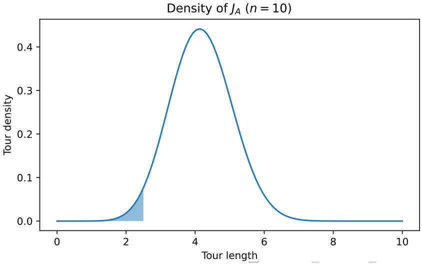
Only have a lower bound on the expected tour length.

But over many iterations, could we sometimes sample better solutions?

Want to prove a tail bound for L_P, i.e.

$$\mathbb{P}(L_P \le j) \le \text{(something small)}.$$

Can prove this for L_A:



Towards a tail bound

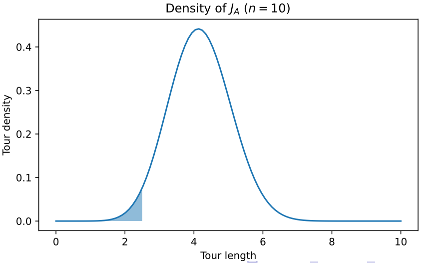
Only have a lower bound on the expected tour length.

But over many iterations, could we sometimes sample better solutions?

Want to prove a tail bound for L_P, i.e.

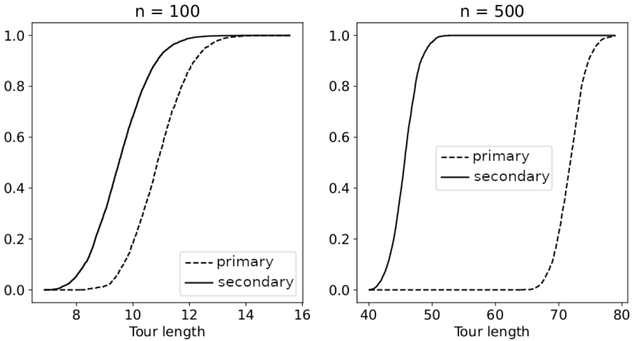
$$\mathbb{P}(L_P \leq j) \leq \text{(something small)}.$$

Can prove this for L_A:



Towards a tail bound

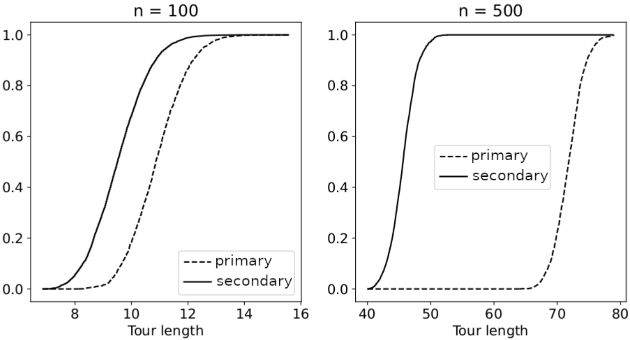
Compare $L_{A/P}$ again:



Hints that $\mathbb{P}(L_P \leq j) \leq \mathbb{P}(L_A \leq j) \rightarrow$ Tail bound for L_P!

Towards a tail bound

Compare $L_{A/P}$ again:



Hints that $\mathbb{P}(L_P \leq j) \leq \mathbb{P}(L_A \leq j) \rightarrow$ Tail bound for L_P!

Consequences

Next step is to prove that $\mathbb{P}(L_P \leq j) \leq \mathbb{P}(L_A \leq j)$.

Then for log-cooling, SA finds global optimum with probability $o_n(1)$ in $2^{o(n)}$ iterations (assuming equilibrium).

Further directions:

- Extend to non-equilibrium situations.
- Consider other problems besides TSP.
- Consider different or more general probabilistic models.

Consequences

Next step is to prove that $\mathbb{P}(L_P \leq j) \leq \mathbb{P}(L_A \leq j)$.

Then for log-cooling, SA finds global optimum with probability $o_n(1)$ in $2^{o(n)}$ iterations (assuming equilibrium).

Further directions:

- Extend to non-equilibrium situations.
- Consider other problems besides TSP.
- Consider different or more general probabilistic models.

Consequences

Next step is to prove that $\mathbb{P}(L_P \leq j) \leq \mathbb{P}(L_A \leq j)$.

Then for log-cooling, SA finds global optimum with probability $o_n(1)$ in $2^{o(n)}$ iterations (assuming equilibrium).

Further directions:

- Extend to non-equilibrium situations.
- Consider other problems besides TSP.
- Consider different or more general probabilistic models.

Consequences

Next step is to prove that $\mathbb{P}(L_P \leq j) \leq \mathbb{P}(L_A \leq j)$.

Then for log-cooling, SA finds global optimum with probability $o_n(1)$ in $2^{o(n)}$ iterations (assuming equilibrium).

Further directions:

- Extend to non-equilibrium situations.
- Consider other problems besides TSP.
- Consider different or more general probabilistic models.

Consequences

Next step is to prove that $\mathbb{P}(L_P \leq j) \leq \mathbb{P}(L_A \leq j)$.

Then for log-cooling, SA finds global optimum with probability $o_n(1)$ in $2^{o(n)}$ iterations (assuming equilibrium).

Further directions:

- Extend to non-equilibrium situations.
- Consider other problems besides TSP.
- Consider different or more general probabilistic models.

Consequences

Next step is to prove that $\mathbb{P}(L_P \leq j) \leq \mathbb{P}(L_A \leq j)$.

Then for log-cooling, SA finds global optimum with probability $o_n(1)$ in $2^{o(n)}$ iterations (assuming equilibrium).

Further directions:

- Extend to non-equilibrium situations.
- Consider other problems besides TSP.
- Consider different or more general probabilistic models.