# 6 Randomized optimization algorithms

## 6.1 Low temperature bounds

Regarding the Metropolis algorithm with large $\theta$ as a randomized optimization algorithm ("simulated annealing at fixed temperature"), one natural finite-time question to ask is "how long does it take to hit the minimum?", and a more technical question is "how large is the relaxation time". Theorem 16 gives crude bounds on these quantities. In the simplest "graph-based" Metropolis construction (Chapter 11 yyy) on a graph of maximal degree $r$, the parameter $\varepsilon$ in Theorem 16 becomes $r^{-1}$, but the theorem doesn't exploit any specific Metropolis construction.

Consider a function $H$ on a $n$-element set $\mathcal{V}$ and the corresponding distribution

$$\pi(v) = z_\theta \exp(-\theta H(v))$$

where $\theta > 0$. Consider an irreducible reversible Markov chain on $\mathcal{V}$, whose stationary distribution is $\pi$, and introduce the graph of possible transitions. Suppose for simplicity that $\min_v H(v)$ is attained at a unique vertex $v_*$. Define the "critical depth" $H\updownarrow$ as follows. Let

$$\hat{H}(v) = \min_{\text{paths } v_*=u_0,u_1,\ldots,v} \ \max_{u\in\text{path}} H(u)$$

$$H\updownarrow = \max_v \hat{H}(v) - H(v).$$

Clearly $H\updownarrow \geq 0$. When the values of $H$ are all distinct, we have $H\updownarrow = 0$ iff $H$ has no local minimum except the global minimum $v_*$.

**Theorem 16** *In the setting above,*

$$\frac{1}{2n}\ \exp(\theta H\updownarrow) \leq \max_v E_v T_{v_*} \leq \varepsilon^{-1} n^2 \exp(\theta H\updownarrow)$$

$$\frac{1}{2n}\ \exp(\theta H\updownarrow) \leq \tau_2 \leq 4\varepsilon^{-1} n^2 \exp(\theta H\updownarrow)$$

*where*

$$\varepsilon = \min\{P(v,x) : P(v,x) > 0, H(x) \leq H(v)\}.$$

The mathematical point of the theorem is that, when $H\updownarrow > 0$, the $\theta \uparrow \infty$ asymptotics are dominated by the $\exp(\theta H updownarrow)$ term. Bounds like this are meaningless algorithmically – the upper bound is larger than the number of states! On the other hand, in the $\theta = 0$ (i.e. random walk on a graph) case we know the $O(rn^2)$ bound is roughly optimal, in the absence of further assumptions.

The proof below uses different arguments for the hitting time and the relaxation time cases. We could alternatively invoke some general inequalities from Chapter 4, as follows. From the definitions we have $\tau_1^{(2.5)} \leq \max_v E_v T_{v_*}$, and from Chapter 4 yyy and yyy we have $\tau_2 \leq \tau_1 \leq K\tau_1^{(2.5)}$, which combine to show that, for reversible chains,

$$\tau_2 \leq K \max_v E_v T_{v_*}, \text{ for some numerical constant } K. \tag{29}$$

Thus the upper bound on $\tau_2$ in Theorem 16, with the $K$ from (29) in place of 4, follows from the upper bound on the mean hitting time.

*Proof of Theorem 16.* We may suppose the values of $H$ are all distinct (by making arbitrarily small changes). The lower bound is easy. Choose $z$ such that $\hat{H}(z) - H(z) = H\updownarrow$. Then for any $v$ with $H(v) \geq \hat{H}(z)$,

$$P_z(X_t = v) = \frac{\pi_v}{\pi_z} P_v(X_t = z) \leq \frac{\pi_v}{\pi_z} \leq \exp(-\theta H\updownarrow)$$

and so $P_z(H(X_t) \geq \hat{H}(z)) \leq n \exp(-\theta H\updownarrow)$. This implies

$$E_z T_{\{v: H(v) \geq \hat{H}(z)\}} \geq \frac{1}{2n} \exp(\theta H\updownarrow).$$

But the left side is a lower bound for $E_z T_{v_*}$.

To argue the upper bound on the mean hitting time, let $\mathcal{S}$ be the set of vertices $v$ such that $\hat{H}(v) = H(v)$. For $v \in \mathcal{S}$ define

$$A_v^< = \{x : \hat{H}(x) < H(v)\},\ A_v^= = \{x : \hat{H}(x) = H(v)\},\ A_v^\geq = \{x : \hat{H}(x) \geq H(v)\}.$$

We shall argue that for $v \in \mathcal{S}$

$$E_v T_{A_v^<} \leq \varepsilon^{-1} n \exp(\theta H\updownarrow) \tag{30}$$

$$\max_{x \in A_v^=} E_x T_{A_v^< \cup \{v\}} \leq \varepsilon^{-1} n (|A_v^=| - 1) \exp(\theta H\updownarrow) \tag{31}$$

These inequalities combine to show

$$\max_{x \in A_v^=} E_x T_{A_v^<} \leq \varepsilon^{-1} n |A_v^=| \exp(\theta H\updownarrow)$$

and the upper bound follows by the obvious iterative argument.

To argue (30,31) it is convenient to use the weighted graphs - electrical network interpretation. Assign edge $(x, y)$ the weight $w_{x,y} = \exp(-\theta H(x))p_{x,y}$. Fix $v$ and consider this weighted graph truncated to $A_v^{\geq}$. The total weight $W$ satisfies

$$W \leq \sum_{x \in A_v^{\geq}} \exp(-\theta H(x)) \leq n \exp(\theta(H\updownarrow - H(v))). \tag{32}$$

Given $x \in A_v^{=}$, there exists a path $x = u_0, u_1, \ldots, u_m = v$ in $A_v^{=}$ of length $l \leq |A_v^{=}| - 1$. Each edge $(u, u')$ of this path has weight at least $\varepsilon \exp(-\theta \max(H(u), H(u'))) \geq \varepsilon \exp(-\theta H(v))$, and so the effective resistance $r(x, v) \leq l\varepsilon^{-1} \exp(\theta H(v))$. Thus by the commute interpretation of resistance, the mean hitting time from $x$ to $v$ is at most $r(x, v)W \leq n(|A_v^{=}| - 1)\varepsilon^{-1} \exp(\theta H\updownarrow)$. But this mean hitting time to $v$ for the truncated chain is clearly an upper bound for the mean hitting time to $\{v\} \cup A_v^{<}$ in the original chain, establishing (31). For the remaining inequality (30), given $v \neq v_* \in \mathcal{S}$ there exists $z \in A_v^{<}$ with $p_{v,z} \geq \varepsilon$. So consider the graph truncated to $A_v^{\geq}$ together with the edge $(v, z)$ with weight $w_0 = \exp(-\theta H(v))p_{v,z}$. The mean hitting time $E_v T_z$ on the truncated graph is an upper bound for the desired mean hitting time on the original graph. But xxx gives the first identity in

$$E_v T_z = \frac{W}{w_0} \leq W \exp(\theta H(v))\varepsilon^{-1}$$

and then (30) follows from (32).

We now turn to the relaxation time. For the upper bound, write $w_{x,y} = \pi(x)P(x, y)$. The bound from Chapter 4 yyy, crudely replacing the term $1_{(e \in \gamma_{xy})}$ by 1, becomes

$$\tau_2 \leq 2 \sum_x \sum_y \pi(x)\pi(y)r(\gamma_{x,y}) \tag{33}$$

where $\gamma_{x,y}$ is a path from $y$ that we are free to choose, and $r(\gamma)$ is the resistance of the path $\gamma$. Now $w_{x,y} \geq \varepsilon \min(\pi(x), \pi(y))$ and so

$$r(\gamma) \leq \frac{n}{\varepsilon \min_{z \in \gamma} \pi(z)}. \tag{34}$$

Now by definition of $H\updownarrow$ we may choose a path $\gamma = \gamma_{x,y}$ such that $\max_{z \in \gamma} H(z) \leq H\updownarrow + \max(H(x), H(y))$ and so

$$\min_{z \in \gamma} \pi(z) \geq \exp(-\theta H\updownarrow)\ \min(\pi(x), \pi(y)).$$

Combining with (34),

$$r(\gamma_{x,y}) \min(\pi(x), \pi(y)) \leq \varepsilon^{-1} n \exp(\theta H\updownarrow).$$

Writing $\pi(x)\pi(y) = \max(\pi(x), \pi(y)) \min(\pi(x), \pi(y))$, we deduce from (33) that

$$\tau_2 \leq 2\varepsilon^{-1} n \exp(\theta H\updownarrow) \quad \left( \sum_x \sum_y \max(\pi(x), \pi(y)) \right)$$

and the double sum is at most $2n$.

To get the lower bound, we apply the general inequality (Chapter 4 yyy)

$$\tau_2 \geq \frac{\pi(A)\pi(A^c)}{Q(A, A^c)}$$

to the subset $A \equiv \{v : \hat{H}(v) < \hat{H}(z)\}$, where as before we choose $z$ such that $\hat{H}(z) - H(z) = H\updownarrow$. Note that if $x \in A, y \in A^c$ and $P(x,y) > 0$ then $y$ must be in $B \equiv \{v : H(v) \geq \hat{H}(z)\}$. Thus $Q(A, A^c) = Q(A, B) = Q(B, A) \leq \pi(B)$, and then

$$\begin{aligned} \tau_2 &\geq \frac{\pi(A)\pi(A^c)}{\pi(B)} \\ &\geq \frac{1}{2} \min\left( \frac{\pi(A)}{\pi(B)}, \frac{\pi(A^c)}{\pi(B)} \right) \\ &\geq \frac{1}{2} \min\left( \frac{\pi(v_*)}{\pi(B)}, \frac{\pi(z)}{\pi(B)} \right) \\ &= \frac{\pi(z)}{2\pi(B)} \\ &\geq \frac{\exp(\theta H\updownarrow)}{2|B|}. \end{aligned}$$

xxx interesting example not available

xxx snakes; bound in terms of rel time of contour subgraphs.

xxx intersting math questions about H with no local minima, but not applied