Week 7, Lecture 1

29 February 2024 17:05

In this lecture we look at the theory behind Simulated Annealing, and we will show that the method is guaranteed to converge - with probability 1 - for any starting value.

Simulated Annealing (§ 17.2).

- Start with an initial temperature T_0.
- Start with an initial state of the system, $\underline{x}^{(0)}$.
- Generate a proposal to move the system into a new state $\underline{x}^{(1)}$.
- If the new state reduces the cost function, the proposal is accepted with probability 1.

$$E^{(1)} = E(\underline{x}^{(1)}), \quad E^{(0)} = E(\underline{x}^{(0)})$$

$$\Delta E = E^{(1)} - E^{(0)}.$$

$$\mathbb{P}(\text{Accept } \underline{x}^{(0)} \to \underline{x}^{(1)}) = \begin{cases} 1 & \text{if } \Delta E < 0. \\ e^{-\Delta E / T_0} & \text{if } \Delta E > 0. \end{cases}$$

- After a certain number of steps like this, the temperature is reduced to T_1.
- Continue thus .....
- Temperature is reduced in a systematic way according to an annealing schedule.

The algorithm satisfies <u>detailed balance</u>:

$$p_T(E^{(k)})\,\mathbb{P}(\underline{x}^{(k)} \to \underline{x}^{(k+1)})$$

$$= p_T(E^{(k+1)})\,\mathbb{P}(\underline{x}^{(k+1)} \to \underline{x}^{(k)}).$$

where $p_T(E) = \dfrac{e^{-\beta E}}{Z}$.

Detailed balance guarantees that all parts of the phase space are sampled.

Algorithm — § 17.3

Annealing Schedule — § 17.4

<u>Aim</u>: Justify the fact that

$$T_k \leq \frac{T_0}{\ln(k)}$$

is a good annealing schedule.

We start with an initial guess $\underline{x}_0$. We generate a new guess ($\underline{x}_1$) — called a proposal — by drawing a vector from a normal distribution:

$$\underline{x}_1 \sim N(\underline{x}_0, \sigma_k)$$

We take $\sigma_k = \sqrt{T_k}$ $(\sigma_k^2 = T_k)$. Other references use $\sigma_k^2 = 2T_k$. Both choices references use $\sigma_k^2 = 2T_k$. Both choices work. The choice $\sigma_k^2 = T_k$ is called <u>Boltzmann annealing</u>.

The probability that the new proposal is in a particular region R of phase space is:

$$\mathbb{P}(\text{New proposal in } R) = \int_R p(\underline{x})\, d^n \underline{x}$$

where $p(\underline{x})$ is the normal distribution

$$p(\underline{x}) = \frac{1}{(2\pi \sigma_k^2)^{n/2}} e^{-\|\underline{x} - \underline{x}_0\|_2^2 / 2\sigma_k^2}$$

As indicated previously, we propose the annealing schedule

$$T_k \leq \frac{T_0}{\ln k},$$

where T_0 is the initial temperature which is supposed to be "sufficiently large" (TBC).

We now look at the probability g_k that the new proposal is to be found in a small volume ΔV around the global min, and we denote this region as $R(\underline{x}_*, \Delta V)$. Hence:

$g_k = \mathbb{P}(\text{New proposal in } R(x_*, \Delta V))$

$$g_k = \mathbb{P}(\text{New proposal in } \mathcal{R}(\underline{x}_*, \Delta V))$$

$$= \int_{\mathcal{R}(\underline{x}_*, \Delta V)} p(\underline{x})\, d^n \underline{x} .$$

We have:

$$g_k = \int_{\mathcal{R}(\underline{x}_*, \Delta V)} \frac{1}{(2\pi\sigma_k^2)^{n/2}} e^{-\|\underline{x}-\underline{x}_0\|_2^2/2\sigma_k} d^n x$$

$$= \left[\int_{x_{*1}-\Delta x/2}^{x_{*1}+\Delta x/2} \frac{1}{(2\pi\sigma_k^2)^{1/2}} e^{-(x_1-x_{01})^2/2\sigma_k^2} dx_1\right] \times \cdots$$

$$\cdots \times \left[\int_{x_{*n}-\Delta x/2}^{x_{*n}+\Delta x/2} \frac{1}{(2\pi\sigma_k^2)^{1/2}} e^{-(x_n-x_{0n})^2/2\sigma_k^2} dx_n\right]$$

$$\simeq \left[\frac{1}{(2\pi\sigma_k^2)^{1/2}} e^{-(x_{*1}-x_{01})^2/2\sigma_k^2} \Delta x\right] \times \cdots$$

$$\cdots \times \left[\frac{1}{(2\pi\sigma_k^2)^{1/2}} e^{-(x_{*n}-x_{0n})^2/2\sigma_k^2} \Delta x\right]$$

$$= \frac{(\Delta x)^n}{(2\pi\sigma_k^2)^{n/2}} e^{-\|\underline{x}_*-\underline{x}_0\|_2^2/2\sigma_k^2}$$

$$= \Delta V \quad e^{-\|\underline{x}_*-\underline{x}_0\|_2^2/2\sigma_k^2} .$$

$$= \frac{\Delta V}{(2\pi \sigma_k^2)^{n/2}} e^{-\|\underline{x}_* - \underline{x}_0\|_2^2 / 2\sigma_k^2}.$$

Here, we have used the TRAPEZOIDAL RULE:

$$\int_{a-\Delta x/2}^{a+\Delta x/2} f(x)\,dx \simeq f(a)\,\Delta x$$

valid for smooth functions.

Summarizing where we are so far:

$$\boxed{g_k = \frac{\Delta V}{(2\pi T_k)^{n/2}} e^{-\|\underline{x}_* - \underline{x}_0\|_2^2 / 2T_k}}$$

We compute the probability that the system will NOT enter this region, and we will show that this probability goes to zero.

<u>To show</u>: $\prod_k (1 - g_k) \longrightarrow 0$ as $k \to \infty$.

Take logs on both sides:

To show:

$$\log \prod (1-g_k) \longrightarrow -\infty \quad \text{as} \quad k \to \infty .$$

Or:

$$\sum_k \log(1-g_k) \longrightarrow -\infty \quad \text{as} \quad k \to \infty .$$

$\propto \Delta V$

Assuming g_k is small, and Taylor-expand:

$$\log(1-g_k) \simeq -g_k .$$

Hence, it suffices to show:

$$\sum_k g_k \longrightarrow +\infty \quad \text{as} \quad k \to \infty .$$

Or

$$\sum_k \frac{1}{(2\pi T_k)^{n/2}} e^{-\|\underline{x}_* - \underline{x}_0\|_2^2 / 2T_k} \longrightarrow \infty .$$

But:

$$T_k \leq \frac{T_0}{\ln(k)} .$$

Hence, it suffices to show:

$$\sum_k (\log k)^{n/2} e^{-\ln k \, \|\underline{x} - \underline{x}_0\|_2^2 / 2T_0} \longrightarrow \infty .$$

We choose T_0 large, such that:

$$\frac{\|\underline{x}_* - \underline{x}_0\|_2^2}{2T_0} \leq \Delta .$$

Then, it suffices to show:

$$\sum_k (\log k)^{n/2} e^{-\overbrace{(\log k)\|\underline{x}_* - x_0\|_2^2 / 2T_0}} \geq \sum_k (\log k)^{n/2} e^{-\log k}.$$

We look at the tail of the series, $k \geq k_0$:

$$\sum_{k=k_0}^{\infty} [\log(k)]^{n/2} e^{-(\log k)\|\underline{x}_* - x_0\|_2^2 / 2T_0} \geq \sum_{k=k_0}^{\infty} (\log k)^{n/2} e^{-\log k}$$

$$= \sum_{k=k_0}^{\infty} (\log k)^{n/2} \cdot \frac{1}{k}$$

$$\geq \sum_{k=k_0}^{\infty} \frac{1}{k}.$$

This is the divergent <u>harmonic series</u>.

$$\ldots \geq \infty .$$

Hence, $\sum_k (\log k)^{n/2} e^{-\|\underline{x}_* - x_0\|_2^2 / 2T_k} \longrightarrow \infty$

as $k \to \infty$.

Reasoning back up the chain,

$$\prod_k (1 - g_k) \longrightarrow 0 .$$

Probability of reaching the global min (ΔV around the global min) goes to 1.