Convergence Proof — Simulated Annealing

Week 7, Lecture 1   03/02/2026

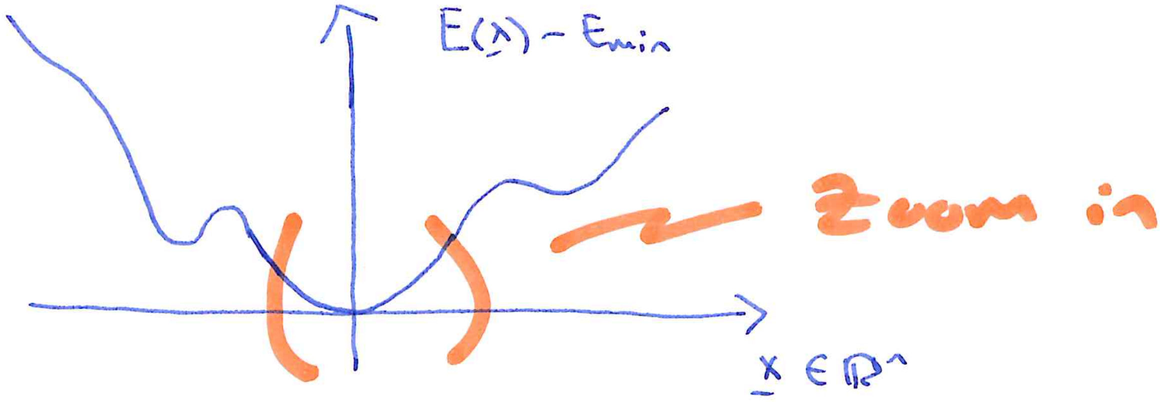
- System, n degrees of freedom
- Characterized by a vector $\underline{x} \in \mathbb{R}^n$, $\mathbb{R}^n$ is the phase space.
- Energy is a function $E(\underline{x})$, with a global minimum.
- Probability distribution function that maximizes entropy for a fixed mean energy $\bar{E}$ is the Boltzman distribution:

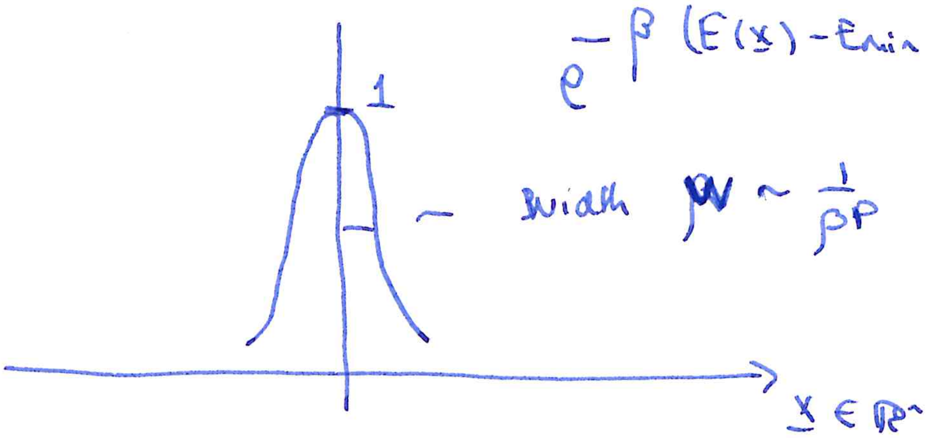
$$P(\underline{x}) = \frac{e^{-\beta E(\underline{x})}}{Z} \quad ; \quad \beta = \frac{1}{T}.$$

Write:

$$P_T(\underline{x}) = \frac{e^{-\beta E(\underline{x})}}{Z},$$

$$Z = \int_{\mathbb{R}^n} e^{-\beta E(\underline{x})} \, d^n x .$$





$e^{-\beta(E(\underline{x})-E_{min})}$

Take $\beta = \frac{1}{T}$, $T \to 0$, $\beta \to \infty$, $w \to 0$

Hence, in the limit as $T \to 0$,

$$\frac{e^{-\beta(E(\underline{x})-E_{min})}}{Z} \longrightarrow \delta(\underline{x} - \underline{x}_*)$$

Where $\underline{x}_*$ is the global min.

Today: turn this into a numerical algorithm, called <u>Simulated Annealing</u> (SA), § 18.2.

- Start with an initial temperature T_0.
- Start with an initial state of the system, $\underline{x}^{(0)} \in \mathbb{R}^n$.
- Generate a <u>proposal</u> $\underline{x}^{(1)}$ to move the system into a new state $\underline{x}^{(1)}$ $\longrightarrow$

- Acceptance of proposal:
  - If the new state reduces $E(x)$, we accept the proposal with probability 1.

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

    $\Delta E = E^{(1)} - E^{(0)}$
  - If the new state fails to reduce the energy ($\Delta E > 0$), we still accept the proposal but with a probability less than 1
- Probability:

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

- After a certain number of steps like this, the temperature is reduced to T_1.
- Continue thus ...
- Temperature is reduced systematically according to the <u>annealing schedule</u>.

<u>Note</u>: This algorithm satisfies detailed balance:

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

This guarantees that all points in the phase space are sampled.

We look at a particular annealing schedule:

$$T_k \leq \frac{T_0}{\ln(k)}, \quad k >> 1.$$

Aim: To show that this is a good annealing schedule (§18.4) We show that with this annealing schedule, we converge on the global min, with probability 1.

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

We take

$$\sigma_k = \sqrt{T_k}, \quad \sigma_k^2 = T_k.$$

This is called Boltzmann annealing.
We look at the probability that the new proposal is in a region R of phase space:

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

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

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

We have: $T_k \leq T_0 / \ln(k)$, $k >> 1$.

$$g_k \simeq \left[\frac{1}{(2\pi\sigma_k^2)^{1/2}} e^{-(x_{*1} - x_1)^2/2\sigma_k^2} \Delta x \right]$$

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

$$= \frac{1}{(2\pi\sigma_k^2)^{n/2}} e^{-\|\underline{x}_* - \underline{x}_0\|_2^2/2\sigma_k^2} \Delta V .$$

We compute the probability that the system will NOT be in the region $R(\underline{x}_*, \Delta V)$ after k cooling steps, and we will show that this probability goes to zero.

<u>To show</u>:

$$\prod_{i=1}^{k} (1 - g_i) \longrightarrow 0 ,$$

as $k \to \infty$.

If and only if

$$\log \prod_{i=1}^{k} (1 - g_i) \longrightarrow -\infty$$

If and only if

$$\cancel{\log} \sum_{i=1}^{k} \log(1 - g_k) \longrightarrow -\infty .$$

Since $g_k \propto \Delta V$ (small), we can Taylor expand. So the previous statements will be true if and only if:

$$\boxed{\sum_{i=1}^{k} g_k \longrightarrow \infty}, \quad \text{as } k \to \infty .$$

We look at the probability that the new proposal is in a region of size ΔV centred at the global min $\underline{x}_*$: We call this g_k.

$$g_k = \mathbb{P}(\text{New proposal is in } R(\underline{x}_*, \Delta V))$$
$$= \int_{R(\underline{x}_*, \Delta V)} p(\underline{x})\, d^n x$$

We have:

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

We may take $R(\underline{x}_*, \Delta V)$ to be a hypercube:

$$g_k = \left[\int_{x_{*1} - \Delta x/2}^{x_{*1} + \Delta x/2} \frac{1}{(2\pi\sigma_k^2)^{1/2}} e^{-(x_1 - x_{01})^2 / 2\sigma_k^2} dx_1 \right]$$
$$\cdots \left[\int_{x_{*n} - \Delta x/2}^{x_{*n} + \Delta x/2} \frac{1}{(2\pi\sigma_k^2)^{1/2}} e^{-(x_n - x_{0n})^2 / 2\sigma_k^2} dx_n \right]$$

We may take ΔV to be an arbitrarily small volume. We can therefore apply the trapezoidal rule to these integrals (one function evaluation at the midpoint) $\longrightarrow$

i.e. we need to show:

$$\sum_{i=1}^{k} \frac{1}{(2\pi T_i)^{n/2}} e^{-\|\underline{x}_* - \underline{x}_0\|_2^2 / 2T_i}$$

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

But $T_i \leq T_0 / \ln(i)$.

So it suffices to show:

$$\cancel{\sum_{i=1}^{k} \frac{1}{(2\pi T_i)^{n/2}}}$$

$$\cancel{\sum_{i=1}^{k} \left(\frac{1}{2\pi T_0}\right)^{n/2} \left[\ln(i)\right]^{n/2} e^{-\ln \|\underline{x}_* - \underline{x}_0\|_2^2 / 2}}$$

$$\sum_{i=1}^{k} \left(\frac{1}{2\pi T_0}\right)^{n/2} \left[\ln(i)\right]^{n/2} e^{-\ln(i) \, \|\underline{x}_* - \underline{x}_0\|_2^2 / 2T_0}$$

$$\longrightarrow \infty , \quad \text{as} \quad k \to \infty .$$

We may choose T_0 sufficiently large, such that:

$$\frac{\|\underline{x}_* - \underline{x}_0\|_2^2}{2T_0} \leq 1 .$$

Thus, it suffices to show:

$$\sum_{i=1}^{k} \ln(i)^{n/2} e^{-\ln(i) \underbrace{\|x_\infty - x_0\|_2^2 / 2T_0}_{\leq 1}}$$

$$\geq \sum_{i=1}^{k} [\ln(i)]^{n/2} e^{-\ln(i)}$$

$$= \sum_{i=1}^{k} [\ln(i)]^{n/2} \cdot \frac{1}{i}$$

It suffices to look at k large, so that:

$$\sum_{i=1}^{k} \ln(i)^{n/2} e^{-\ln(i) \|x_\infty - x_0\|_2^2 / 2T_0}$$

$$\geq \sum_{i=1}^{k} [\ln(i)]^{n/2} \cdot \frac{1}{i}$$

$$\geq \sum_{i=1}^{k} \frac{1}{i} \quad , \text{ for } k \text{ large.}$$

This is the harmonic series, which is divergent.

Hence:

$$\sum_{i=1}^{k} \ln(i)^{n/2} e^{-\ln(i) \|x_\infty - x_0\|_2^2 / 2T_0}$$

$$\geq \dots \sum_{i=1}^{k} \frac{1}{i} \longrightarrow \infty, \text{ as } k \to \infty.$$

Reasoning back up, we see that:

$$\prod_{i=1}^{k} (1 - g_n) \longrightarrow 0 \quad \text{as } k \to \infty.$$

Hence, probability of being within ΔV (small) of the global min tends to 1, as $k \to \infty$.

S.A. converges to global min with probability 1. ∎