Questions: Given any state $S$, how much work can we extract from it? How much work do we need to create a state $S$?

At this stage, "work" is only implicit (via Landauer erasure/Stilard engine). Later: model Hamiltonians explicitly.

Questions for now:

(i) How many pure bits can we extract from a given distribution $p$?

(ii) How many pure bits do we need to create $p$?

(i): We want $p \xrightarrow{\text{noisy}} S_n$, where $S_n = \left( \frac{1}{0} \right)^n \text{ on } n$ pure bits

with $n$ as large as possible.

Lorentz curve of $S_n$: \[1 \quad \text{larger } n \Rightarrow \text{more "valuable" as a resource}\]
Generalize to $\forall k \in \mathbb{N}$:
\[ S_k := \left( \frac{1}{k}, \ldots, \frac{1}{k}, 0, \ldots, 0 \right) \in \mathbb{R}^k \]

where $I = \log \frac{e}{k}$.

e.g. $S_{\log 3} = \left( \frac{1}{3}, \frac{1}{3}, \frac{1}{3}, 0, 0, 0, 0 \right)$. $S_k$ is called a "sharp state".

**Distillable nontriviality**

\[ \Rightarrow \text{We want } p \xrightarrow{\text{approx}} S_k \text{ with } I \text{ as large as possible.} \]

(intuition: then we can extract exactly $I$ kB of work in a Stirling engine)

\[ \Rightarrow \text{Least curve of } S_k \text{ must fit below least curve of } p:\]

\[ \Rightarrow x \geq 1 - 2^{-I} \]

\[ x = 1 - \frac{1}{m} \cdot \# \text{ of nonzero entries of } p = 1 - \frac{1}{m} \text{ rank}(p) \]

\[ \Rightarrow \frac{1}{m} \text{ rank}(p) \leq 2^{-I} \text{. } I \leq \log m - \log \text{ rank}(p) \]

with $H_0(p) = \log \text{ rank}(p)$, writing $\alpha_p = m$, we get
Lemma: $p \xrightarrow{\text{univ.}} S_{I}$ if and only if

$$I \leq \log d_{p} - H_{0}(p).$$

That is, the "distillable non-uniformity" is

$$I_{0}(p) := \log d_{p} - H_{0}(p).$$

**Interpretation:** This is the amount of work that we can extract from $p$ in a Stirling engine with absolute certainty (not just "on average").

**Nonuniformity of formation:**

We want $S_{I} \xrightarrow{\text{univ.}} p$, with $I$ as small as possible.

$\Rightarrow$ Lorentz curve of $p$ must fit below Lorentz curve of $S_{I}$:

$$\Rightarrow \text{initial slope of } L_{S_{I}} \geq \text{initial slope of } L_{p}.$$  

$$\Rightarrow \frac{1}{2^{2^{I}}} \geq \frac{p}{N}$$

$$2^{I} \geq N \max_{i} p_{i}.$$

**Lemma:** $S_{I} \xrightarrow{\text{univ.}} p$ if and only if

$$I \geq \log d_{p} - H_{0}(p)$$
That is, the "nonuniformity of formation" is

\[ I_\infty(p) = \log dp - H_\infty(p). \]

Interpretation: This is the amount of work that we have to invest to create \( p \) perfectly (exactly with zero probability of failure) from "nothing" (i.e., from mutually mixed states), by
- using Landauer erasure to prepare pure bits;
- using noisy operations to prepare \( p \) from these.

It is easy to show that

\[ 0 \leq H_\infty(p) \leq I_\infty(p) \leq \log dp \]

\[ \Rightarrow 0 \leq I_0(p) \leq I_\infty(p) \leq \log dp \]

\[ \Rightarrow \text{In general, it costs more work to create a state than what we can gain/extract from that state.} \]

What about "failure probability" etc.? \( \Rightarrow \) next step.

2.7. Closeness of (quantum) states: the trace distance

Definition: For two density matrices \( S, T \in B(C^n) \), we define their trace distance as

\[ D(S, T) = \frac{1}{2} \| S - T \|_1 \]

This is a special case of the "Schrödinger P-Norm"
\[ \| M \|_p := [\text{tr}(|M|^p)]^{1/p} \quad (p \geq 1); \]
\[ D(S, G) = \frac{1}{2} \| S - G \|_1. \]

This is a distance measure:
\[ 0 \leq D(S, G) \leq 1, \]
\[ D(S, G) = 0 \iff S = G, \]
\[ D(S, G) = 1 \iff S \text{ and } G \text{ are orthogonal, i.e. } S^T G = 0 \]
\[ D(S, T) \leq D(S, G) + D(G, T). \]

Operational meaning as distinguishability of \( S \) and \( G \):

Two-outcome projective measurement:

State \( S \)

\[ \text{Prob(1st outcome)} = \text{tr}(SP), \]
\[ \text{Prob(2nd outcome)} = \text{tr}(SP^2), \]
\[ P^\perp = I - P. \]

Lemma:
\[ D(S, G) = \max \left\{ \text{max}_p \text{max}_{\text{orth proj.}} \left| \text{tr}(SP) - \text{tr}(GP) \right| \right\} \]
\[ = \max \left\{ \text{max}_p \text{max}_{\text{orth proj.}} \left[ P(S - G) \right] \right\} \]
\[ = \max \left\{ \text{max}_p \left[ E(S - G) \right] \right\}. \]

Discuss meaning:
Quantum operations \( \mathcal{N} \) (e.g., open-system evolutions) only ever decrease the distinguishability:
\[ D(\mathcal{N}(S), \mathcal{N}(G)) \leq D(S, G). \]
\[ D(U S U^+, U S U^+) = D(S, G). \]

Unitary evolution preserves trace distance.
What about "success probability"?

- state $5$ has a complicated protocol/measure
  - yes!
  - no!

Classical observation

There is some operation element $0 \leq E \leq 1$ such that $\text{Prob}(\text{success}) = \text{Tr}(E_5)$.

If protocol always succeeds on state $8$, i.e., $\text{Prob}(\text{success} | 8) = 1$, but we have instead $8' \neq 8$ then

$$\text{Prob}(\text{success} | 8') \geq 1 - D(8, 8').$$

Classical analog: the variation distance

$[E_8, E_5] = 0 \Rightarrow 8 = (p_1, \ldots, p_n), \quad 8' = (q_1, \ldots, q_n)$ in some basis.

$$D(8, 8') = \frac{1}{2} \text{Tr} |8 - 8'| = \frac{1}{2} \text{Tr} \left| \left( p_1 - q_1, \ldots, p_n - q_n \right) \right| = \frac{1}{2} \left( \sum_{i=1}^{n} |p_i - q_i| \right).$$

$$= \frac{1}{2} \sum_{i=1}^{n} |p_i - q_i|.$$

For prob. vectors $p, q \in \mathbb{R}^n$,

$$D(p, q) := \frac{1}{2} \sum_{i=1}^{n} |p_i - q_i| = \frac{1}{2} \|p - q\|_1; \quad \text{special case of } l_p\text{-norm } \|x\|_p := \left( \sum_{i=1}^{n} |x_i|^p \right)^{1/p}.$$

**Lemma:**

$$D(p, q) = \max_{S \subseteq \{1, \ldots, n\}} \|p(S) - q(S)\|_1 = \max_{S \subseteq \{1, \ldots, n\}} \left( \sum_{i \in S} |p_i - q_i| \right)^{1/p}.$$

where $p(S) = \sum_{i \in S} p_i$. 
2.8. Approximate state formation and distillation of pure bits

Write $P \xrightarrow{\epsilon\text{-noisy}} \tilde{Q}$ if there exists a prob. vector $\tilde{Q}$ such that $P \xrightarrow{\text{noisy}} \tilde{Q}$ with $D(Q, \tilde{Q}) \leq \epsilon$.

Consistency of quantum and classical case:

Lemma: For quantum states $\tilde{S}, \tilde{T}$, there exists a noisy (quantum) operation $W$ and another state $\tilde{S'}$ with $D(\tilde{S}, \tilde{S'}) \leq \epsilon$ and $W(\tilde{S}) = \tilde{S'}$ if and only if $\lambda(\tilde{S}) \xrightarrow{\epsilon\text{-noisy}} \lambda(\tilde{S'})$.

Proof not here.

Definition ("smooth entropies"): For a prob. vector $p$ and $\epsilon \geq 0$, define

\[ H^E_\infty(p) := \max_{p': D(p, p') \leq \epsilon} H_\infty(p'), \quad \text{and} \]

\[ H^E_0(p) := \min_{p': D(p, p') \leq \epsilon} H_0(p'). \]

Smooth entropies have many applications in classical and quantum information theory.