Nonequilibrium work relations

III. Dissipation and the Arrow of Time

second law of thermodynamics (macroscopic) \( \rightarrow W \geq \Delta F \)
nonequilibrium work relations (microscopic) \( \rightarrow \langle e^{-\beta W} \rangle = e^{-\beta \Delta F} \), etc.

What is the precise relationship?
What does the 2nd law “look like” at the microscale?

**Inequalities:**

Jensen’s inequality \( \rightarrow \quad \langle e^x \rangle \geq e^{\langle x \rangle} \)

holds for any real, convex function of a real variable

Simple derivation of Jensen’s inequality:

convex function \( A(x) \quad , \quad A''(x) \geq 0 \quad \forall x \)

probability distribution

\( p(x) \quad , \quad 1 = \int dx \ p(x) \quad , \quad \bar{x} = \int dx \ x \ p(x) \)

define a new function \( B(x) = A(\bar{x}) + (x - \bar{x}) A'(\bar{x}) \quad ( \text{tangent to} \ A(x) \ ) \)

by construction: \( A(x) \geq B(x) \quad \forall x \)

\( \langle A(x) \rangle \geq \langle B(x) \rangle = \int dx \ p(x) \left[ A(\bar{x}) + (x - \bar{x}) A'(\bar{x}) \right] = A(\bar{x}) \)
\[ e^{-\beta \Delta F} = \langle e^{-\beta W} \rangle \geq e^{-\beta \langle W \rangle} \quad \Rightarrow \quad \langle W \rangle \geq \Delta F \quad \text{(as expected)} \]

now let’s obtain a somewhat stronger result …

\[ P(W \leq \Delta F - n \beta^{-1}) = \text{probability that the 2nd law is “violated” by at least } nk_B T \]

\[ = \int_{-\infty}^{\Delta F-n \beta^{-1}} dW \rho(W) \]

\[ \leq \int_{-\infty}^{\Delta F-n \beta^{-1}} dW \rho(W) e^{\beta(\Delta F-n \beta^{-1}-W)} \]

\[ \leq e^{\beta(\Delta F-n \beta^{-1})} \int_{-\infty}^{+\infty} dW \rho(W) e^{-\beta W} = \exp(-n) \]

area underneath tail decays exponentially (or faster)
Guessing the direction of Time's Arrow:

In forward/reverse processes, trajectories come in conjugate pairs: if \( X \) denotes a possible realization of the forward process, then its conjugate twin \( X^+ \) is a possible realization of the reverse process.

Suppose you are shown a movie depicting the microscopic evolution of the system as \( \lambda : A \to B \) (forward process). How can you tell whether you are viewing (1) the events in the order in which they actually occurred, or (2) a movie of the reverse process, run backward?

exercise in statistical inference:
given the observed data, which hypothesis is more likely?

Bayes' theorem

\[
L(hyp \mid dat) \propto P(dat \mid hyp) \cdot P_0(hyp)
\]

e.g.

\[
L(F \mid X) \propto P^F(X) \cdot P_0(F)
\]

assume equal priors: \( P_0(F) = P_0(R) = 1/2 \)

\[
L(F \mid X) = \frac{P^F(X)}{P^F(X) + P^R(X^+)} = \frac{1}{1 + \frac{P^R(X^+) / P^F(X)}{P^F(X)}}
\]

\[
= \frac{1}{1 + e^{-\beta(W-\Delta F)} }
\]

[Shi03,Mar07]
\[ L(F \mid X) = \frac{1}{1 + e^{-\beta(W - \Delta F)}} \]

- when \( W > \Delta F \), it is more likely that we are seeing the events in the correct order ("forward"), while for \( W < \Delta F \) it is the other way around.

- the transition from “almost certainly reverse” (\( L \approx 0 \)) to “almost certainly forward” (\( L \approx 1 \)) happens over a few \( k_B T \) (consistent w/ earlier results: very low probability to see second law “violated” by more than a few \( k_B T \))
Relative entropy and dissipation: [Mae99, Gas04, Jar06, Kaw07]

Given two normalized probability distributions \( p(x) \) & \( q(x) \), the relative entropy of \( p \) with respect to \( q \) is

\[
D(p \mid q) = \int dx \ p(x) \ln \frac{p(x)}{q(x)} \geq 0
\]

(aka Kullback-Leibler divergence)… provides a measure of the degree to which one distribution is distinguishable from the other.

Let’s use this to quantify thermodynamic irreversibility.

\( P^F(X) = \) distribution of forward trajectories
\( P^R(X^+) = \) distribution of reverse trajectories

The relative entropy between these two distributions measures time-reversal asymmetry. (How differently does the system respond in the two processes?)

\[
D(P^F \mid P^R) = \int dX \ P^F(X) \ln \frac{P^F(X)}{P^R(X^+)}
\]

\[
= \int dX \ P^F(X) \beta \left[ W^F(X) - \Delta F \right]
\]

\[
= \beta \left( \langle W \rangle^F - \Delta F \right) \equiv \beta W_{diss}^F
\]

This result relates a physical measure of irreversibility (dissipated work) to an information-theoretic measure of time-reversal asymmetry (relative entropy).

Consistent w/ macroscopic experience:

\( W_{diss} \gg k_B T \), \( D \gg 1 \)
References

Below is a short list of papers cited in these lecture notes. For a more comprehensive reference list, see the bibliography in my mini-review in *Eur. Phys. J. B* 64, 331-340 (2008)