

## B4 ゼミ #9

大上由人

2025 年 7 月 6 日

### 5.3 位相体積の変化としてのエントロピー生成と KL ダイバージェンスによる表現

#### 5.3.1 KL ダイバージェンス

二つの確率分布の違いをはかる指標として KL ダイバージェンスを導入する。

**Def.KL ダイバージェンス**

確率分布  $p$  の  $q$  に対する KL ダイバージェンスは次のように定義される

$$D(p\|q) := \sum_j p_j \ln \frac{p_j}{q_j}. \quad (5.1)$$

たとえば、コイン A（表が出る確率が  $1/2$ ）と、コイン B（表が出る確率が  $2/3$ ）との間の KL ダイバージェンスは以下のように計算される。

$$D(\text{coin A} \|\text{coin B}) = \frac{1}{2} \ln \frac{1/2}{2/3} + \frac{1}{2} \ln \frac{1/2}{1/3} = 0.058 \dots \quad (5.2)$$

KL ダイバージェンスは、距離の公理<sup>\*1</sup>は満たさないが、ある程度距離っぽい性質も持っている。

**Thm.KL ダイバージェンスの非負性**

任意の確率分布  $p, q$  に対して、

$$D(p\|q) \geq 0 \quad (5.3)$$

が成り立つ。

<sup>\*1</sup>

- 非退化性:  $\forall x, y, d(x, y) = 0 \Leftrightarrow x = y$
- 非負性:  $\forall x, y, d(x, y) \geq 0$
- 対称性:  $\forall x, y, d(x, y) = d(y, x)$
- 三角不等式:  $\forall x, y, z, d(x, y) + d(y, z) \geq d(x, z)$

**Thm.KL ダイバージェンスの非退化性**

KL ダイバージェンスは次の性質を満たす。

$$D(p\|q) = 0 \Leftrightarrow p = q \quad (5.4)$$

これらの証明のためには Jensen の不等式を用いる。

**Thm.Jensen の不等式**

$f(x)$  が狭義に下に凸な関数であるとき、次の不等式が成り立つ。

$$\sum_{i=1}^n a_i f(x_i) \geq f\left(\sum_{i=1}^n a_i x_i\right) \quad (5.5)$$

ただし、 $a_1, \dots, a_n \geq 0$ ,  $x_1, \dots, x_n$  は任意で、 $\sum_i a_i = 1$  とする。等号が成り立つのは、 $x_1 = x_2 = \dots = x_n$  のとき、かつそのときに限る。 $f(x)$  が上に凸である場合、不等号の向きは逆になる。

**Prf.(Jensen の不等式)**

$n$  に関する数学的帰納法で示す。 $n = 2$  の場合は、これは狭義に凸な関数の定義そのものである。

次に、 $n = k$  のとき不等式が成り立つと仮定し、 $n = k + 1$  のときも成り立つことを示す。 $n \leq k$  のときの仮定を 2 回用いて、次のように書ける。

$$\sum_{i=1}^{k+1} a_i f(x_i) = \left(\sum_{i=1}^k a_i\right) \cdot \left(\sum_{i=1}^k \frac{a_i}{\sum_{j=1}^k a_j} f(x_i)\right) + a_{k+1} f(x_{k+1}) \quad (5.6)$$

$$\geq \left(\sum_{i=1}^k a_i\right) \cdot f\left(\sum_{i=1}^k \frac{a_i}{\sum_{j=1}^k a_j} x_i\right) + a_{k+1} f(x_{k+1}) \quad (5.7)$$

$$\geq f\left(\left(\sum_{i=1}^k a_i\right) \cdot \frac{\sum_{i=1}^k a_i x_i}{\sum_{j=1}^k a_j} + a_{k+1} x_{k+1}\right) \quad (5.8)$$

$$= f\left(\sum_{i=1}^{k+1} a_i x_i\right) \quad (5.9)$$

よって、 $n = k + 1$  のときも不等式は成り立つ。

さらに、等号が成り立つのは以下の 2 条件が同時に満たされるときに限る。

$$x_1 = \dots = x_k \quad (5.10)$$

$$\frac{\sum_{i=1}^k a_i x_i}{\sum_{i=1}^k a_i} = x_{k+1} \quad (5.11)$$

以上より、 $n = k + 1$  の場合も不等式が成り立つことが示された。□

$\ln(x)$  は  $x$  に関する狭義な凹関数であるので、Jensen の不等式を  $\ln(x)$  に適用することで、以下の不等式が得られる。

$$-D(p\|q) = \sum_j p_j \left( \ln \frac{q_j}{p_j} \right) \leq \ln \left( \sum_j p_j \frac{q_j}{p_j} \right) = \ln \left( \sum_j q_j \right) = \ln 1 = 0 \quad (5.12)$$

したがって、

$$D(p\|q) \geq 0$$

が示された。

また、Jensen の不等式における等号成立条件により、この不等式が等号となるのは、すべての  $j$  に対して  $p_j = q_j$  が成り立つとき、かつそのときに限る。<sup>\*2</sup> □

#### Thm.KL ダイバージェンスの単調性

2つの確率分布  $p, q$  が、同じマルコフ時間発展方程式に従ってそれぞれ  $p', q'$  に時間発展するならば、次が成り立つ。

$$D(p\|q) \geq D(p'\|q') \quad (5.13)$$

すなわち、マルコフ過程において二つの分布は KL ダイバージェンスの意味で近づく。

**Prf.**

まず、条件付き KL ダイバージェンスを定義する。確率分布  $P$  から  $Q$  に対する条件付き KL ダイバージェンスは次で定義される。

$$D(P(x|y)\|Q(x|y)) := \sum_y P(y) \sum_x P(x|y) \ln \frac{P(x|y)}{Q(x|y)} \quad (5.14)$$

これは非負である（各  $y$  に対して、Jensen の不等式より、 $\sum_x P(x|y) \ln(P(x|y)/Q(x|y)) \geq 0$  だから）。また、以下の連鎖律が成り立つ。

$$D(P(x, y)\|Q(x, y)) = D(P(y)\|Q(y)) + D(P(x|y)\|Q(x|y)) \quad (5.15)$$

( $\because$ ) 簡単な計算により示せる。

---

<sup>\*2</sup> ここで確率の規格化を考えていることに注意。

$$D(P(x, y) \| Q(x, y)) = \sum_{x, y} P(x, y) \ln \frac{P(x, y)}{Q(x, y)} \quad (5.16)$$

$$= \sum_{x, y} P(x|y)P(y) \ln \frac{P(x|y)P(y)}{Q(x|y)Q(y)} \quad (5.17)$$

$$= \sum_{x, y} P(x|y)P(y) \left( \ln \frac{P(x|y)}{Q(x|y)} + \ln \frac{P(y)}{Q(y)} \right) \quad (5.18)$$

$$= \sum_y P(y) \sum_x P(x|y) \ln \frac{P(x|y)}{Q(x|y)} + \sum_y P(y) \ln \frac{P(y)}{Q(y)} \quad (5.19)$$

$$= D(P(x|y) \| Q(x|y)) + D(P(y) \| Q(y)) \quad (5.20)$$

次に、 $P(x^i, x^f)$  を初期状態  $x^i$ 、終状態  $x^f$  にある確率とし、 $Q(x^i, x^f)$  を同様に定義する。

このマルコフ過程の遷移行列を  $K$  とすると、

$$P(x^f) = \sum_{x^i} K(x^f | x^i) P(x^i), \quad Q(x^f) = \sum_{x^i} K(x^f | x^i) Q(x^i) \quad (5.21)$$

すなわち、遷移確率に関して次が成り立つ。

$$P(x^f | x^i) = Q(x^f | x^i) = K(x^f | x^i) \quad (5.22)$$

したがって、

$$D(P(x^f | x^i) \| Q(x^f | x^i)) = 0 \quad (5.23)$$

連鎖律を用いれば、

$$D(P(x^i, x^f) \| Q(x^i, x^f)) = D(P(x^i) \| Q(x^i)) + D(P(x^f | x^i) \| Q(x^f | x^i)) \quad (5.24)$$

$$= D(P(x^f) \| Q(x^f)) + D(P(x^i | x^f) \| Q(x^i | x^f)) \quad (5.25)$$

KL ダイバージェンスの非負性と上の等式を用いれば、

$$D(P(x^i) \| Q(x^i)) = D(P(x^f) \| Q(x^f)) + D(P(x^i | x^f) \| Q(x^i | x^f)) \quad (5.26)$$

より

$$D(P(x^i) \| Q(x^i)) \geq D(P(x^f) \| Q(x^f)) \quad (5.27)$$

が示され、定理の主張が得られる。  $\square$

KL ダイバージェンスの単調性の重要な応用の一つは「縮約性」である。周辺状態への縮約  $(x_i, y_j) \rightarrow x_i$  がマルコフ過程であるから TODO このこと確認 KL ダイバージェンスは常に減少することが以下のように表される。

$$D(P(x_i, y_j) \| Q(x_i, y_j)) \geq D(P(x_i) \| Q(x_i)) \quad (5.28)$$

ここで、 $P(x_i) := \sum_j P(x_i, y_j)$ ,  $Q(x_i) := \sum_j Q(x_i, y_j)$  と定義する。

なお、KL ダイバージェンスはすべての距離の公理を満たすわけではない。距離関数  $d(x, y)$  に期待される公理は以下の 4 つである。

- 非負性： $d(x, y) \geq 0$
- 同一性： $d(x, y) = 0 \iff x = y$
- 対称性： $d(x, y) = d(y, x)$
- 三角不等式： $d(x, y) + d(y, z) \geq d(x, z)$

KL ダイバージェンスは上のうち非負性と同一性を満たすが、後の 2 つ（対称性と三角不等式）は満たさない。<sup>\*3</sup>

**Prf.** ( $H(x) \geq H(x|y)$ )

エントロピーの不等式  $H(x) \geq H(x|y)$  を KL ダイバージェンスを用いて証明する。

$$H(x) - H(x|y) = - \sum_i P(x_i) \ln P(x_i) + \sum_{i,j} P(x_i, y_j) \ln P(x_i|y_j) \quad (5.32)$$

$$= - \sum_{i,j} P(x_i, y_j) \ln P(x_i) + \sum_{i,j} P(x_i, y_j) \ln P(x_i|y_j) \quad (5.33)$$

$$= \sum_{i,j} P(x_i, y_j) \ln \frac{P(x_i|y_j)}{P(x_i)} \quad (5.34)$$

$$= \sum_{i,j} P(x_i, y_j) \ln \frac{P(x_i, y_j)}{P(x_i)P(y_j)} \quad (5.35)$$

$$= D(P(x, y) \| P(x)P(y)) \quad (5.36)$$

$$\geq 0 \quad (5.37)$$

よって、主張される不等式  $H(x) \geq H(x|y)$  が得られる。  $\square$

---

<sup>\*3</sup> 対称性を満たさないことは自明なので、後者の反例を示しておく、 $p = (0.2, 0.8)$ ,  $q = (0.5, 0.5)$ ,  $r = (0.8, 0.2)$  とすると、

$$D(p \| q) \simeq 0.192745 \quad (5.29)$$

$$D(q \| r) \simeq 0.223144 \quad (5.30)$$

$$D(p \| r) \simeq 0.831777 \quad (5.31)$$

したがって、 $D(p \| q) + D(q \| r) \geq D(p \| r)$  は成り立たない。

補足として、KL ダイバージェンスは確率分布に対してだけでなく、 $\sum_i p_i = \sum_i q_i$  を満たす一般の非負ベクトルに対しても定義が拡張され、このときも KL ダイバージェンスの非負性および単調性は成り立つ。(教科書付録参照)

### 5.3.2 Phase volume

ゆらぎの定理は、位相体積の変化に関する関係式としても理解できる。ここでは、その基本的なアイデアを紹介する。

ゆらぎの定理と位相体積の関係を理解するために、 $0 \leq t \leq \tau$  の時間区間における決定論的な力学系を考える。全系の初期状態  $w^{\text{tot}}(0)$  から終状態  $w^{\text{tot}}(\tau)$  への写像を  $M$  とする。時刻  $t$  における状態  $w^{\text{tot}}_s$  の位相密度を  $\rho(w^{\text{tot}}, t)$  とし、参照分布を  $\rho^{\text{ref}}(w^{\text{tot}})$  とする。

DFT の場合、参照分布は以下のように定義される\*4

$$\rho_t^{\text{ref}} \left( w \otimes \bigotimes_{\nu} w^{\text{B}, \nu} \right) := \rho(w, t) \otimes \bigotimes_{\nu} \rho^{\text{can}, \nu}(w^{\text{B}, \nu}) \quad (5.38)$$

ここで  $\rho^{\text{can}, \nu}$  は  $\nu$  番目の熱浴のカノニカル分布である。

力学系が Liouville の定理に従って位相体積を保存するとすれば、エントロピー生成は以下のよう位相空間密度の比の対数で定義される。

$$\hat{\sigma} = \ln \frac{\rho(w^{\text{tot}}, 0)}{\rho_t^{\text{ref}}(M(w^{\text{tot}}))} \quad (5.39)$$

この表現は、エントロピー生成が参照分布と比べて、位相密度がどれだけ縮小または拡大したかを定量化している。

熱浴の参照分布をカノニカル分布とすることで、位相密度の変化は以下のようにエネルギー変化と関係づけられる。

$$\Delta E^{\nu} := E^{\nu}(w^{\text{B}, \nu}(\tau)) - E^{\nu}(w^{\text{B}, \nu}(0)) \quad (5.40)$$

このとき、以下が成り立つ。

$$\frac{\rho^{\text{ref}, \text{B}}(w^{\text{B}}(\tau))}{\rho^{\text{ref}, \text{B}}(w^{\text{B}}(0))} = e^{-\sum_{\nu} \beta^{\nu} \Delta E^{\nu}} \quad (5.41)$$

ここで  $\rho^{\text{ref}, \text{B}}$  は熱浴の参照分布であり、 $w^{\text{B}}(t)$  は時刻  $t$  における浴の状態である。

この観点は、IFT と Jarzynski 等式の違いも明らかにする。IFT では、順過程の終状態の分布を参照分布とし、Jarzynski 等式ではカノニカル分布を参照分布とする。すなわち、両者の違いは参照分布の選び方にある。次節では IFT の場合を詳しく見ていく。

\*4 記号がわかりにくいのが、要するに系と  $\nu$  個の熱浴の直積をとっているだけである。

### 5.3.3 決定論的な場合

$0 \leq t \leq \tau$  において、決定論的な力学系を考える。 $w^{\text{tot}}$  を系と熱浴から成る複合系の状態、 $M$  を初期状態  $w^{\text{tot}}(0)$  から終状態  $w^{\text{tot}}(\tau)$  への写像とする。

順過程の初期分布を、

$$P(w^{\text{tot}}, 0) := P(w, 0) \prod_{\nu} P^{\text{can}, \nu}(w^{\text{B}, \nu}, 0) \quad (5.42)$$

とし、逆過程の初期分布を

$$P^{\dagger}(w^{\text{tot}}, 0) := P^{\dagger}(\bar{w}, 0) \prod_{\nu} P^{\text{can}, \nu \dagger}(\bar{w}^{\text{B}, \nu}, 0) \quad (5.43)$$

とする。このとき、決定論的な平均エントロピー生成は次のように KL ダイバージェンスの形式で表される。

**Thm. 平均エントロピー生成（決定論の場合）**

$$\langle \hat{\sigma} \rangle = \int dw^{\text{tot}} P(w^{\text{tot}}, 0) \ln \frac{P(M(w^{\text{tot}}), \tau)}{P^{\dagger}(\tilde{M}(w^{\text{tot}}), 0)} = D(P(w^{\text{tot}}, \tau) \| P^{\dagger}(\tilde{w}^{\text{tot}}, \tau - t)) \quad (5.44)$$

ここで、 $\tilde{M}(w^{\text{tot}})$  は  $M(w^{\text{tot}})$  の時間反転、であり、

$$D(P(w^{\text{tot}}, \tau) \| P^{\dagger}(\tilde{w}^{\text{tot}}, \tau - t)) := \int dw^{\text{tot}} P(w^{\text{tot}}, \tau) \ln \frac{P(w^{\text{tot}}, \tau)}{P^{\dagger}(\tilde{w}^{\text{tot}}, \tau - t)} \quad (5.45)$$

である。

一本目の等号はゆらぎの定理とまったく同じである。<sup>\*5</sup> また、二本目の等号も、Liouville の定理により積分測度を保存することを利用すれば示せる。

この式は、サイクル過程のエネルギー変換効率についても興味深い示唆を与える。2つの異なる温度  $\beta_H^{-1}$ ,  $\beta_L^{-1}$  の熱浴と系からなる複合系を考える ( $\beta_H < \beta_L$ )。

初期分布を任意な系の分布と 2 つの熱浴のカノニカル分布の積とすると、時間反転性より  $P^{\dagger}(\tilde{w}, 0) = P(w, \tau)$  が成り立つ。

このとき、平均エントロピー生成は

$$\langle \hat{\sigma} \rangle = -\beta_H Q_H + \beta_L Q_L$$

で与えられる。ただし、 $Q_H$  は高温熱浴からの吸熱量、 $Q_L$  は低温熱浴への放熱量。

---

<sup>\*5</sup> 決定論的な力学系においては、

$$P(w^{\text{tot}}, 0) = P(M(w^{\text{tot}}), \tau) \quad (5.46)$$

となることを用いている。

したがって、熱機関の効率  $\eta$  は次のように書ける。

$$\begin{aligned}\eta &:= \frac{W}{Q_H} = 1 - \frac{Q_L}{Q_H} = 1 - \frac{1}{Q_H} \frac{\langle \hat{\sigma} \rangle + \beta_H Q_H}{\beta_L} \\ &= 1 - \frac{\beta_H}{\beta_L} - \frac{D(P(w; \tau) \| P(\bar{w}; 0))}{\beta_L Q_H}.\end{aligned}\quad (5.47)$$

### 5.3.4 確率的な場合

確率的マルコフ過程においては、軌道の確率とその時間反転軌道の確率との KL ダイバージェンスが、正味のエントロピー生成を表す。

**Thm. 平均エントロピー生成 (確率的)**

確率的マルコフ過程において、平均エントロピー生成は次のように KL ダイバージェンスの形で表される。

$$\langle \hat{\sigma} \rangle = \int d\Gamma P(\Gamma) \ln \frac{P(\Gamma)}{P^\dagger(\tilde{\Gamma})} = D(P(\Gamma) \| P^\dagger(\tilde{\Gamma})) \quad (5.48)$$

ここで、 $\Gamma$  は系の軌道、 $P(\Gamma)$  はその確率、 $\tilde{\Gamma}$  は時間反転された軌道を表す。

**Prf.**

式の対数項を熱項とエントロピー項に分解する。

$$\ln \frac{P(\Gamma)}{P^\dagger(\tilde{\Gamma})} = \ln \frac{P(\Gamma(\tau))P(w(0))}{P^\dagger(\Gamma^\dagger(\tau))P(w(\tau))} = \sum_\nu \beta_\nu \hat{Q}_\nu + \hat{s}(w(\tau); \tau) - \hat{s}(w(0); 0) \quad (5.49)$$

両辺のアンサンブル平均を取れば目的の式が得られる。  $\square$

Sect.3.3 で述べたように、エントロピー生成率は遷移確率の KL ダイバージェンスでも書ける。

$$\dot{\sigma} = \sum_{w, w'} p_w p_{w \rightarrow w'} \ln \frac{p_w p_{w \rightarrow w'}}{p_{w'} p_{w' \rightarrow w}^\dagger} = D(p_w p_{w \rightarrow w'} \| p_{w'} p_{w' \rightarrow w}^\dagger) \quad (5.50)$$

これを時間で積分することで、もう一つの平均エントロピー生成の表現が得られる。

$$\sigma = \int dt D(p_w(t) p_{w \rightarrow w'}(t) \| p_{w'}(t) p_{w' \rightarrow w}^\dagger(t - \tau)) \quad (5.51)$$

このようにして、エントロピー生成とその生成率は、それぞれ経路確率および遷移確率による KL ダイバージェンスとして記述できる。

### 5.3.5 絶対的不可逆性

IFT と KL ダイバージェンスによるエントロピー生成の表現は、多くの場合において同等な意味を持つが、特異な状況では一致しないことがある。その代表例が「絶対的不可逆過程」である。この現象は、位相空間密度に基づくエントロピー生成の理解とも関連する。



### 例：壁の除去

典型的な例として、中央に可動壁を持つ箱を考える。粒子を箱の左側に配置し、右側は真空とする。  $t = 0$  で壁を除去し、粒子の運動を観察する。

この操作はミクロな意味で絶対的に不可逆である。なぜなら、壁を除去した後のある状態は、物理的に到達可能な初期状態には存在しないからである。このため、IFT の証明において、逆過程を構成するときに、すべての終状態（またはすべての経路）を考慮できない。<sup>\*6</sup>

この特異性を明確にするため、以下のような 2 状態マルコフ遷移過程を考える。

状態  $w_1, w_2$  を持ち、エネルギー  $E_1 = 0, E_2$  の 2 状態系において、初期分布を  $p_1(0) = 1 - \varepsilon, p_2(0) = \varepsilon$  とする。

これが  $0 \leq t \leq \infty$  の間に平衡分布  $p^{\text{eq}}$  へ緩和する過程を考える。2 つのケースを比較する。

(1)  $\varepsilon \rightarrow 0$  の極限

(2)  $\varepsilon = 0$  と固定 ( $p_2(0) = 0$ )

このとき、シャノンエントロピー（および KL ダイバージェンス）は連続性を持つため、

$$\lim_{\varepsilon \rightarrow 0} (-\varepsilon \ln \varepsilon - (1 - \varepsilon) \ln(1 - \varepsilon)) = \ln 1 = 0 \quad (5.52)$$

となり、特異性を持たない。

一方で、IFT においては、軌道エントロピー（stochastic entropy）の寄与が以下のように特異性を持つ。

$$\lim_{\varepsilon \rightarrow 0} p_2(0) \ln \frac{p_2(0)}{p_2^{\text{eq}}} = p_2^{\text{eq}} \neq 0 \quad (5.53)$$

ここでは、熱や遷移確率の項は省略している。

したがって、絶対的不可逆性が存在する場合、IFT は元の形では成り立たない。

$$\langle e^{-\hat{\sigma}} \rangle \neq 1 \quad (5.54)$$

## 5.4 強結合系における熱力学量

決定論的な場合では、系と熱浴の相互作用ハミルトニアンが弱いと仮定してきた。ここでは、系が熱浴と強く相互作用する場合を考える。

系と熱浴の複合系の全ハミルトニアンは次のように表される。

$$H_{\text{tot}}(x, y) = H_S(x) + H_{\text{SB}}(x, y) + H_B(y) \quad (5.55)$$

---

<sup>\*6</sup> IFT の証明では、任意の順過程に対して対応する逆過程が存在することを仮定していた。

ここで、 $x, y$  はそれぞれ系と熱浴の状態を表す。 $H_S$  と  $H_B$  はそれぞれ系と熱浴のハミルトニアン、 $H_{SB}$  は相互作用ハミルトニアンである。 $H_S$  のみ時間依存性を持ち、その他の項は時間に依存しないと仮定する。相互作用を丸め込んだ有効ハミルトニアン  $H^{\text{eff}}(x)$  を定義する。複合系のカノニカル分布における周辺分布が、有効ハミルトニアン に対するカノニカル分布と一致するように定義する。すなわち、

$$p^{\text{eq}}(x) := \frac{1}{Z_{\text{tot}}} \int dy e^{-\beta(H_S(x) + H_{SB}(x,y) + H_B(y))} = \frac{1}{Z_S} e^{-\beta H^{\text{eff}}(x)} \quad (5.56)$$

対応する分配関数は以下になる。

$$Z_{\text{tot}} := \int dx dy e^{-\beta(H_S(x) + H_{SB}(x,y) + H_B(y))} \quad (5.57)$$

$$Z_S := \int dx e^{-\beta H^{\text{eff}}(x)} \quad (5.58)$$

また、熱浴の相互作用なしのカノニカル平均は次で定義される。

$$\langle A \rangle_B^{\text{eq}} := \frac{1}{Z_B} \int dy A e^{-\beta H_B(y)}, \quad Z_B := \int dy e^{-\beta H_B(y)} \quad (5.59)$$

このような性質を満たす有効ハミルトニアンを以下のように定義する。

**Def. 有効ハミルトニアン**

$$H^{\text{eff}}(x) := H_S(x) - \frac{1}{\beta} \ln \left( \left\langle e^{-\beta H_{SB}(x,y)} \right\rangle_B^{\text{eq}} \right) \quad (5.60)$$

この定義により、系の平衡分布  $p^{\text{eq}}(x)$  が先の式を満たすことが確認できる。<sup>\*7</sup>

次に、自由エネルギーとエントロピーを定義する。

$$F_X := -\frac{1}{\beta} \ln Z_X \quad (X = S, B, \text{tot}) \quad (5.61)$$

これらを用いて、エントロピーを以下のように定義する。

$$S_X := \beta^2 \frac{\partial F_X}{\partial \beta} \quad (5.62)$$

特に、系の有効エントロピーは次のようになる。<sup>\*8</sup>

$$S_S = - \int dx p^{\text{eq}}(x) \ln p^{\text{eq}}(x) + \beta^2 \int dx p^{\text{eq}}(x) \frac{\partial H^{\text{eff}}(x)}{\partial \beta} \quad (5.63)$$

<sup>\*7</sup> 初等的な計算。手書きのメモで確認する。

<sup>\*8</sup> これも手書きのメモで確認する。

第2項は、 $H^{\text{eff}}$  が  $\beta$  に依存することによる修正項である。

**Def. 有効エネルギー**

有効エネルギー  $E^{\text{eff}}(x)$  は次のように定義される。

$$E^{\text{eff}}(x) := H^{\text{eff}}(x) + \beta \frac{\partial H^{\text{eff}}(x)}{\partial \beta} \quad (5.64)$$

このもとで、熱力学関係式が成り立つ。<sup>a</sup>

$$F_S = E_S - \frac{S_S}{\beta} \quad (5.65)$$

---

<sup>a</sup> 手書きのメモで確認する。

エネルギーやエントロピーの加法性を確認するため、自由エネルギーの和を計算する。

$$F_S + F_B = -\frac{1}{\beta} \ln Z_S Z_B \quad (5.66)$$

$$= -\frac{1}{\beta} \ln \int dx e^{-\beta H^{\text{eff}}(x)} Z_B \quad (5.67)$$

$$= -\frac{1}{\beta} \ln \int dx e^{-\beta H_S(x)} \left\langle e^{-\beta H_{SB}(x,y)} \right\rangle_B^{\text{eq}} Z_B \quad (5.68)$$

$$= -\frac{1}{\beta} \ln \int dx dy e^{-\beta(H_S(x) + H_{SB}(x,y) + H_B(y))} \quad (5.69)$$

$$= F_{\text{tot}} \quad (5.70)$$

よって、自由エネルギーの加法性  $F_S + F_B = F_{\text{tot}}$  が成り立つ。

$0 \leq t \leq \tau$  の過渡過程を考える。始状態と終状態の差を  $\Delta$  で表すことにし、単一の軌道上でのエネルギー保存則に基づき仕事を定義する。

$$W := -\Delta H_{\text{tot}} := -H_{\text{tot}}(x(\tau), y(\tau)) + H_{\text{tot}}(x(0), y(0)) \quad (5.71)$$

有効エネルギー（式 5.62）の定義に基づき、系から放出される熱は次のように定義される。

$$Q := -W - \Delta E^{\text{eff}} = \Delta[H_{\text{tot}} - E^{\text{eff}}] \quad (5.72)$$

この定義により、有効エネルギー  $E^{\text{eff}}$  を内部エネルギーとみなすことで、第1法則が正しく回復される。

さらに、第2法則の帰結としてのゆらぎ定理も回復できる。

初期時刻における複合系の分布を以下のようにとる。

$$p^0(x, y) := p_S^0(x) p_B^{\text{eq}}(y|x) \quad (5.73)$$

ここで  $p_S^0(x)$  は任意の系の分布、 $p_B^{\text{eq}}(y|x)$  は与えられた  $x$  に対するカノニカル分布であり、

$$p_B^{\text{eq}}(y|x) := \frac{1}{Z_B \langle e^{-\beta H_{SB}(x,y)} \rangle_B^{\text{eq}}} e^{-\beta H_{SB}(x,y)} e^{-\beta H_B(y)} \quad (5.74)$$

( $Z_B$  は熱浴の分配関数)。TODO 式のお気持ち確認  
また、有効エントロピーを次のように定義する。

$$s^{\text{eff}}(x) := -\ln p_S(x) + \beta^2 \frac{\partial H^{\text{eff}}(x)}{\partial \beta} \quad (5.75)$$

ここで  $p_S(x)$  は系の周辺分布である。

**Thm. 強結合系における IFT**

過渡過程において、初期分布を  $p^0(x, y) = p_S^0(x) p_B^{\text{eq}}(y|x)$  としたとき、以下の IFT が成り立つ：

$$\langle e^{-(\Delta s^{\text{eff}} + \beta Q)} \rangle = 1 \quad (5.76)$$

ここで、 $\Delta s^{\text{eff}} = s^{\text{eff}}(x(\tau)) - s^{\text{eff}}(x(0))$  は有効エントロピーの変化を表す。

この式は、見た目は通常の IFT と同じ形式だが、エントロピーや熱の定義が強結合の効果を反映して修正されている点に注意が必要である。

**Prf.**

時間  $t = \tau$  における系の周辺分布を  $p_S^\tau(x) := \int dy p^\tau(x, y)$  と定義する。

次の関係式に注意する。

$$\begin{aligned} \ln p_B^{\text{eq}}(y|x) &= -\beta (H_{SB}(x, y) + H_B(y)) - \ln Z_B - \ln \langle e^{-\beta H_{SB}(x,y)} \rangle_B^{\text{eq}} \\ &= -\beta (H_{\text{tot}}(x, y) - H^{\text{eff}}(x)) - \ln Z_B, \end{aligned} \quad (5.77)$$

これより、以下が得られる。

$$-\Delta s^{\text{eff}} - \beta Q = -\Delta [s^{\text{eff}} + E^{\text{eff}} - H^{\text{eff}}] = \ln \frac{p_S^\tau(x(\tau)) p_B^{\text{eq}}(y(\tau)|x(\tau))}{p_S^0(x(0)) p_B^{\text{eq}}(y(0)|x(0))} \quad (5.78)$$

これは任意の単一軌道  $\Gamma$  に対して成り立つ。

DFT の式 (5.9) と同様に、以下が得られる。

$$\left\langle e^{-(\Delta s^{\text{eff}} + \beta Q)} \right\rangle = \int dx(0) dy(0) p_S^0(x(0)) p_B^{\text{eq}}(y(0)|x(0)) e^{-(\Delta s^{\text{eff}} + \beta Q)} \quad (5.79)$$

$$= \int dx(0) dy(0) p_S^{\tau}(x(\tau)) p_B^{\text{eq}}(y(\tau)|x(\tau)) \quad (5.80)$$

$$= 1, \quad (5.81)$$

ここで、 $dx(0)dy(0) = dx(\tau)dy(\tau)$  を用いた。 □