贾金斯基等式及其推广

陈童

July 16, 2025

Contents

考虑一个与热库接触的有限经典系统。热力学的核心概念是当系统某些外部参数随时间变化时，对该系统所做的功。（这些参数可代表外场强度、系统的外部力学条件**x**比如受限空间体积，或更抽象地表示分子动力学模拟过程中开启/关闭的粒子间相互作用。）当参数沿参数空间中某路径γ 从初始点A无限缓慢地变化至终点B时，整个过程就是可逆的，根据第二章中引入正则系综时的相关分析，这时候系统获得的总功W[1] 等于初末构型间的自由能差$W = \Delta F = F_B - F_A$。相比之下，当参数以有限速率沿γ切换时(从而是不可逆过程)，W将取决于系统与热库的微观初始条件，其平均值将超过ΔF

$$\Delta F \leq \langle W \rangle. \tag{1}$$

这里$\langle \cdots \rangle$表示对所有可能的W测量结果的平均值，每次测量前需让系统与热库在温度T 下平衡，这时外参数固定于A。（从A 到B的路径γ及其参数

[1]注意，不是系统向外界做的总功，两者差个负号.

切换速率在各次测量间保持不变。）差值$\langle W\rangle - \Delta F$ 即为不可逆过程中熵增导致的耗散功。

然而，1996年，贾金斯基(C. Jarzynski)的一个重要结果表明，(1)式可以由不等式推广为一个等式。这个结果及其推广可以说是统计物理近数十年最重要的进展，也构成本章将要讲述的核心内容。

1 贾金斯基等式的推导

1.1 信息不守恒的经典力学

如果系统真的满足哈密顿正则方程，那么就有刘维尔定理成立，即$d\mu_t = d\mu_0$，从而必有

$$d\mu_{t+\epsilon} = d\mu_t, \tag{2}$$

式中ϵ为一个无穷小的时间间隔。根据第二章中的相关讨论，对于这样的系统，其精细化信息必定是守恒的。

但是现实中很多系统与外界之间是存在信息流耦合的，比如我们这里考察与外界热库之间存在热交换的系统就属于这种情形，因此就需要把哈密顿力学理论推广到这种信息不守恒情形。

推广到信息不守恒情形，那(2)式就不能成立了，根据熵与相空间体积的关系，这时候我们应该有

$$k_B \log\big(\frac{d\mu_{t+\epsilon}}{d\mu_t}\big) = \delta S\big(\mu_t\big), \tag{3}$$

式中$\delta S(\mu)$表示系统在ϵ时间之内熵的增加量，或者说系统增加的不确定性。不妨将这个熵的增量写成$\delta S(\mu_t) = k_B\kappa(\mu_t)\epsilon$，$\kappa(\mu_t)$**为单位时间之内增加的不确定性**，从而根据(3)式，即有

$$d\mu_{t+\epsilon} = d\mu_t \exp\big(\kappa(\mu_t)\epsilon\big). \tag{4}$$

将这个式子沿着相空间的演化轨迹积分，即有

$$d\mu_\tau = d\mu_0 \exp\big(\int_0^\tau \kappa(\mu_t)dt\big). \tag{5}$$

另一方面，信息不守恒的系统当然不再是哈密顿正则系统，所以我们需要对哈密顿正则方程进行修正。最简单的修正是给它加上一个反映与外界之间信息耦合的项$\kappa^i(\mathbf{x})$(下面μ^i表示相点μ的第i个坐标分量)

$$\frac{d\mu^i}{dt} = \{\mu^i, H\} + \kappa^i(\mu). \tag{6}$$

问题是，加上的这一项和前面的单位时间输入的熵$\kappa(\mu)$之间是什么关系呢？

为了看清这两者的联系，我们根据坐标变换可以写出

$$d\mu_{t+\epsilon} = d\mu_t \cdot \det\big(\frac{\partial \mu_{t+\epsilon}}{\partial \mu_t}\big) = d\mu_t\big(1 + \epsilon\partial_i(\dot{\mu}^i)\big). \tag{7}$$

代入修改以后的演化方程(6)，注意到标准的泊松括号项是满足相体积守恒的(刘维尔定理的证明)，从而即有

$$d\mu_{t+\epsilon} = d\mu_t\big(1 + \epsilon\partial_i\kappa^i\big). \tag{8}$$

与(4)式相比，即有

$$\partial_i\kappa^i(\mu) = \kappa(\mu). \tag{9}$$

即是说，给定从系统擦除信息的比特率$\kappa(\mu)$, 通过求解这个方程，我们就能得出有信息擦除时哈密顿正则方程的修正项$\kappa^i(\mu)$。这样，我们就把经典力学推广到了信息不守恒情形。

1.2 信息不守恒与贾金斯基等式

回到我们要考察的置于温度为T的恒温热库中的系统，假设把控制系统的外参数笼统地记作$\lambda(t)$, 设参数是在0到τ的有限时间之内进行切换的。记系统的哈密顿量为$H_\lambda(\mu)$。通过切换系统的控制参数$\lambda(t)$，使得它在两个热平衡态之间演化。在演化的中间阶段，系统与热库之间有热量交换，因此这是一个信息不守恒的问题，在演化的中间阶段，其满足的动力学方程应该是(6)。

令系统在$0 \le t \le \tau$的时间段内从平衡态A开始经历参数切换，并最终稳定为平衡态B，注意平衡态B不需要是τ时刻确立的，而是可以在τ时刻系

统参数由λ_A切换为λ_B以后，再保持在这个参数上并经过足够的弛豫时间达成热平衡之后才确立的。在初末态，也就是A态和B态，系统可以用正则系综来描述，从而相空间的概率分布函数为(**本章我们以p而不是ρ表示相空间概率密度**)

$$p_{eq}(\mu_0, \lambda_A) = e^{\beta[F_A - H_A(\mu_0)]}, \quad p_{eq}(\mu_\tau, \lambda_B) = e^{\beta[F_B - H_B(\mu_\tau)]}, \tag{10}$$

式中F_A, F_B分别为A, B态的自由能，并且我们已经利用了$F = -k_B T \log \mathcal{Z}$。

则我们有

$$\begin{aligned} p_{eq}(\mu_\tau, \lambda_B) d\mu_\tau = e^{\beta[F_B - H_B(\mu_\tau)]} d\mu_\tau = e^{\beta[F_B - H_B(\mu_\tau)]} \exp\Big(\int_0^\tau \kappa(\mu_t) dt\Big) d\mu_0 \\ = e^{\beta\left[F_B - F_A - H_B(\mu_\tau) + H_A(\mu_0) + k_B T (\int_0^\tau \kappa(\mu_t) dt)\right]} p_{eq}(\mu_0, \lambda_A) d\mu_0. \end{aligned} \tag{11}$$

式中第二个等号我们代入了(5)式。由于$k_B \int_0^\tau \kappa(\mathbf{x}_t) dt$是整个演化过程热库向系统输入的熵，从而

$$Q = k_B T (\int_0^\tau \kappa(\mathbf{x}_t) dt) \tag{12}$$

就是整个过程热库向系统输入的热量。因此

$$W = H_B(\mu_\tau) - H_A(\mu_0) - k_B T (\int_0^\tau \kappa(\mu_t) dt) \tag{13}$$

就是切换参数的过程中，外界对系统所做的功。

从而我们就可以把(11)式改写成

$$p_{eq}(\mu_\tau, \lambda_B) d\mu_\tau = e^{\beta \Delta F} e^{-\beta W} p_{eq}(\mu_0, \lambda_A) d\mu_0, \tag{14}$$

式中$\Delta F = F_B - F_A$为初末态自由能的增量，这是一个纯粹的宏观状态参量，与微观过程无关。将上面这个式子两边对相空间积分，利用概率密度的归一化条件，即得

$$e^{-\beta \Delta F} = \langle e^{-\beta W} \rangle, \tag{15}$$

式中$\langle e^{-\beta W} \rangle$表示$e^{-\beta W}$的统计平均值。(15)式就是著名的贾金斯基等式, 它以定量的形式给出了封闭系统经历不可逆过程前后自由能的改变量与过程中外界对系统所做的功的统计平均值之间的联系。

注意不等式(1)可直接由以上贾金斯基等式，通过应用数学恒等式$\langle e^x \rangle \geq e^{\langle x \rangle}$ (也就是指数函数的凸性)推导得出。这直接从微观的哈密顿量基础确立了不等式(1)，而非通过熵增原理。

2 反馈控制下的贾金斯基等式

让我们回到不等式

$$\Delta F \leq \langle W \rangle. \tag{16}$$

当然，它等价于热力学第二定律。然而，正如我们在第二章讨论麦克斯韦妖时看到的，当系统受到一个“麦克斯韦妖”的反馈控制时，热力学第二定律需要推广。这里“反馈”指的是控制协议依赖于测量获得的结果，比如麦克斯韦妖对容器中间的活板门的控制是根据对粒子速度的测量结果。

当然，我们已经看到，(16)式是贾金斯基等式的自然推论。因此，很显然，在反馈控制存在的情况下，贾金斯基等式也需要进行推广。这就是本节我们想要讨论的问题。本节的结果最早是Takahiro Sagawa and Masahito Ueda, 在Generalized Jarzynski Equality under Nonequilibrium Feedback Control一文中得到的。本节将要给出的推导是基于对Takahiro Sagawa的Hamiltonian Derivations of the Generalized Jarzynski Equalities under Feedback Control 一文的改进。

现在，我们记$\lambda_t \equiv \lambda(t)$, 因此$\lambda_A = \lambda_0$, $\lambda_B = \lambda_\tau$, 相应的也有$F_A = F_0$, $F_B = F_\tau$。如前所述，初始时，系统处于如下热平衡系综

$$p_{eq}(\mu_0, \lambda_0) = e^{\beta[F_0 - H_{\lambda_0}(\mu_0)]}. \tag{17}$$

进而可以根据如下概率守恒定义相应的刘维尔绘景的概率密度$p_t(\mu)$

$$p_t(\mu_t)d\mu_t \equiv p_{eq}(\mu_0, \lambda_0)d\mu_0. \tag{18}$$

注意，由于系统与热库之间存在信息交换，所以$d\mu_t \neq d\mu_0$。

我们在中间时刻$t_m(0 < t_m < \tau)$对系统进行测量，获得结果ν。假设该测量存在由以下条件概率密度所描述的随机误差

$$p(\nu|\mu_m), \tag{19}$$

即在系统真实状态$\mu_m \equiv \mu_{t_m}$给定的前提下，t_m时刻测得结果ν的概率密度。当然，μ_m和ν的联合概率密度可以表示为

$$p(\mu_m, \nu) = p(\nu|\mu_m)p_{t_m}(\mu_m). \tag{20}$$

获得结果ν的总概率密度，由$p(\nu) = \int d\mu_m p(\mu_m, \nu)$给出。

为了表征状态μ_m与测量结果ν之间的信息关联，我们引入互信息$\langle I \rangle$, 根据第一章，其定义为

$$\langle I \rangle \equiv \int d\mu_m d\nu p(\mu_m, \nu) \log \frac{p(\mu_m, \nu)}{p_{t_m}(\mu_m) p(\nu)}. \tag{21}$$

其中我们记

$$I(\mu_m, \nu) = \log \frac{p(\mu_m, \nu)}{p_{t_m}(\mu_m) p(\nu)}. \tag{22}$$

如果通过测量，我们获取的信息越多，$\langle I \rangle$就越大，反之，如果误差越大，那么$\langle I \rangle$就越小。特别的，如果μ_m与ν相互独立，即满足$p(\mu_m, \nu) = p_{t_m}(\mu_m) p(\nu)$, 则有$\langle I \rangle = 0$, 它表示测量无法获得任何真正的信息。注意，$I(\mu_m, \nu)$依赖于$\mu_m$和$\nu$, 而$\mu_m$又由初始时的$\mu_0$决定，所以，$I(\mu_m, \nu)$依赖于$\mu_0$和$\nu$。

在t_m时刻进行测量以后，我们对系统实施反馈控制，使得t时刻$(t > t_m)$对系统参数λ_t的控制协议依赖于t_m时刻的测量结果ν, 记为$\lambda_t(\nu)(t > t_m)$, 因此，$t > t_m$时刻的系统状态μ_t也将依赖于ν, 记为$\mu_t(\nu)(t > t_m)$。特别的，最终的平衡系综B也将依赖于ν, 记为

$$p_{eq}\big(\mu_\tau(\nu), \lambda_\tau(\nu)\big) = \exp\big\{\beta[F_\tau(\nu) - H_{\lambda_\tau(\nu)}(\mu_\tau(\nu))]\big\}. \tag{23}$$

与上一节类似，记

$$\Delta F \equiv F_\tau(\nu) - F_0. \tag{24}$$

进而，过程中外界对系统所做的功为

$$W = H_{\lambda_\tau(\nu)}(\mu_\tau(\nu)) - H_{\lambda_0}(\mu_0) - k_B T(\int_0^\tau \kappa(\mu_t) dt). \tag{25}$$

式中$\kappa(\mu_t)$满足下式

$$d\mu_\tau(\nu) = d\mu_0 \cdot \exp\big(\int_0^\tau \kappa(\mu_t) dt\big). \tag{26}$$

注意，在$t > t_m$时刻，$\kappa(\mu_t)$依赖于ν。同样，无论ΔF也好，还是功W也好，均既依赖于μ_0，又依赖于ν。

我们将要证明的结论是下面的等式

$$\langle e^{-\beta(W-\Delta F)-I}\rangle = 1. \tag{27}$$

式中$\langle\cdots\rangle$表示对所有可能的μ_0和中间的测量结果ν进行统计平均。这个式子就是反馈控制下的贾金斯基等式，其关键特征是，等式左边包含测量获得的互信息项。特别的，如果$I=0$，那它就退回了上一节所讲的贾金斯基等式。

上式的证明如下：首先注意到

$$e^{-I} = \frac{p_{t_m}(\mu_m)p(\nu)}{p(\mu_m,\nu)} = \frac{p(\nu)}{p(\nu|\mu_m)}. \tag{28}$$

因此，

$$\begin{aligned}
\langle e^{-\beta(W-\Delta F)-I}\rangle &= \int d\mu_0 d\nu \Big[p_{eq}(\mu_0,\lambda_0)p(\nu|\mu_m)\times \\
&\quad e^{-\beta[H_{\lambda_\tau(\nu)}(\mu_\tau(\nu))-H_{\lambda_0}(\mu_0)-k_BT(\int_0^\tau \kappa(\mu_t)dt)]}e^{\beta(F_\tau(\nu)-F_0)}\frac{p(\nu)}{p(\nu|\mu_m)}\Big] \\
&= \int d\mu_0 d\nu \Big[e^{\beta(F_0-H_{\lambda_0}(\mu_0))}e^{-\beta[H_{\lambda_\tau(\nu)}(\mu_\tau(\nu))-H_{\lambda_0}(\mu_0)-k_BT(\int_0^\tau \kappa(\mu_t)dt)]} \\
&\quad \times e^{\beta(F_\tau(\nu)-F_0)}p(\nu)\Big] \\
&= \int d\mu_0 d\nu e^{\int_0^\tau \kappa(\mu_t)dt}e^{\beta[F_\tau(\nu)-H_{\lambda_\tau(\nu)}(\mu_\tau(\nu))]}p(\nu) \\
&= \int d\nu d\mu_\tau(\nu) e^{\beta[F_\tau(\nu)-H_{\lambda_\tau(\nu)}(\mu_\tau(\nu))]}p(\nu) = 1.
\end{aligned} \tag{29}$$

其中第一个等于号我们代入了功的定义式，最后一行代入了(26)式，以及热平衡系综的归一化条件。(27)式得证！

特别的，根据(27)式，利用指数函数的凸性，不难得到

$$\langle \Delta F - W\rangle \le k_BT\langle I\rangle. \tag{30}$$

这就是反馈控制下的广义热力学第二定律。这个不等式表明，通过反馈控制，系统的自由能增量可能超过外界对系统所做的功，其超出部分与测量所获互信息成正比。

西拉德热机

西拉德热机（Szilard engine）是由匈牙利物理学家利奥·西拉德（Leo Szilard）在1929年提出的一个思想实验，其核心目的是探讨**信息与热力学第二定律**的深层关系，尤其是可以作为(30)式的一个应用特例。

在西拉德之前，麦克斯韦曾提出著名的“麦克斯韦妖”思想实验：假设一个绝热容器被隔板分为两部分，中间有一个“妖”可以控制隔板上的小门。“妖”能观察分子运动速度，只允许快分子从左到右、慢分子从右到左通过小门，最终使右侧温度升高、左侧温度降低——这似乎让系统自发熵减，直接挑战了“孤立系统熵永不减少”的热力学第二定律。西拉德热机的提出，正是为了分析“麦克斯韦妖”中“观察（获取信息）”这一行为的物理本质：信息是否具有热力学代价？

西拉德热机的核心是一个仅含单个分子的气缸（简化系统以聚焦微观状态），它与温度为T的恒温热库相接触，配合活塞、重物和一个“观察者”（类似麦克斯韦妖）。其工作循环可分为四步：

1.初始状态：分子随机分布。气缸内只有一个分子（气体分子），在无规则热运动中随机处于气缸左侧或右侧（两种可能状态，概率各50%）。此时系统的熵由分子位置的不确定性决定：根据玻尔兹曼熵公式$S = k_B \log \Omega$（Ω 为微观状态数），这里$\Omega = 2$，故初始熵$S_1 = k_B \log 2$。

2. 测量：获取分子位置信息。观察者（或“妖”）测量分子的位置，确定它具体在左侧还是右侧（排除不确定性）。这一步的关键是“获取信息”：原本分子有2种可能状态，测量后状态唯一确定（$\Omega = 1$），系统的“信息熵”减少了$k_B \log 2$（信息熵的减少等价于获取了1比特信息）。

西拉德假设：测量过程本身可以不消耗能量（理想情况下），但信息的“记录”会暂时储存在观察者的“记忆”中。

3. 做功：利用信息提取能量。根据测量结果（分子在左或右），观察者在气缸中间插入一个活塞，将气缸分为两部分（分子所在的一侧为“高压区”，另一侧为“低压区”）。由于分子的热运动，它会碰撞活塞并推动其移动（膨胀过程），并通过在活塞上挂上重物最终实现对外做功（例如提升重物）。

单分子膨胀做功的能量可通过热力学计算：对于等温过程，气体对外做功$W = \int_{V/2}^{V} P dV = \int_{V/2}^{V} k_B T \frac{dV}{V} = k_B T \log 2$(式中$V$为整个气缸的体积，并且我们利用了对于单个分子的理想气体物态方程$PV = k_B T$)。此时，

分子的运动空间从气缸的一半扩充到整个气缸（活塞移动到边缘），也就是和初始状态完全相同，所以与初始状态相比，自由能的增量为零，即$\Delta F = 0$。

因此，以上三步的总体效果就是，通过测量分子的位置信息，我们成功地从系统中提取了$W = k_B T \log 2$的功。这个结果完全吻合(30)式，因为我们通过测量提取的互信息为$\log 2$, 而系统向外做的功也正好是$k_B T \log 2$，而又有$\Delta F = 0$, 所以正好是使得不等式(30)的等号成立的情况。

4. 重置：擦除信息以完成循环为了让热机重复工作，必须重置系统：移除活塞，同时擦除观察者记忆中关于分子位置的信息（否则记忆会累积信息，无法形成循环）。

这一步是核心：根据后来由兰道尔（Rolf Landauer）在1961年提出的“兰道尔原理”，擦除1比特信息必然会向环境释放至少$k_B T \log 2$ 的热量（即产生至少$k_B \log 2$ 的熵增）。观察者的这部分熵增恰好补偿了步骤2中系统减少的熵（$k_B \log 2$），使整个孤立系统(把观察者包括在内)的总熵变≥ 0，完全符合通常的热力学第二定律。