

不卖，纯恶心人卷 3

命题人：@ 一碗星空汤

一. (30 分) 静电屏蔽

在空间中分布着电中性的等离子体，其数密度为 $2n_0$ （正离子密度为 n_0 ，电子密度为 n_0 ）。其中，正离子与电子带电量绝对值相同，整体呈电中性。现在我们在空间原点处放置一点电荷 q 。我们认为正离子由于质量较大，仍然保持均匀分布。但电子由于质量较小，受到点电荷影响，其在空间中的分布发生了改变，分布满足玻尔兹曼分布，电子的温度为 T ，质量为 m 。

1. 写出空间的总电势 $V(r)$ 满足的常微分方程。
2. 考察远离中心处的电势，有 $|eV(r)| \ll k_B T$ ，对微分方程作一阶近似并求解。

已知球坐标系中的拉普拉斯算子为

$$\nabla^2 f = \frac{1}{r} \frac{\partial^2(rf)}{\partial r^2} + \frac{1}{r^2 \sin \theta} \frac{\partial}{\partial \theta} \left(\sin \theta \frac{\partial f}{\partial \theta} \right) + \frac{1}{r^2 \sin^2 \theta} \frac{\partial^2 f}{\partial \phi^2}$$

[解答]

1. (15 分)

有泊松方程 (3 分)

$$\nabla^2 V = -\rho/\varepsilon_0$$

电子的数密度为 (3 分)

$$n_e \propto \exp\left(\frac{eV}{k_B T}\right)$$

当 $r \rightarrow \infty$ 时，有 $n_e \rightarrow n_0$ ， $V \rightarrow 0$ ，此时可知玻尔兹曼分布前的比例系数为 n_0 ，即 (3 分)

$$n_e = n_0 \exp\left(\frac{eV}{k_B T}\right)$$

于是有 (3 分)

$$\rho = q\delta^3(\mathbf{r}) + n_0 e \left[1 - \exp\left(\frac{eV}{k_B T}\right) \right]$$

可以得到 (3 分)

$$\frac{1}{r} \frac{d^2(rV)}{dr^2} = -\frac{1}{\varepsilon_0} \left\{ q\delta^3(\mathbf{r}) + n_0 e \left[1 - \exp\left(\frac{eV}{k_B T}\right) \right] \right\}$$

2. (15 分)

当 $|eV| \ll k_B T$ 时，有 (3 分)

$$\frac{1}{r} \frac{d^2(rV)}{dr^2} = -\frac{q}{\varepsilon_0} \delta^3(\mathbf{r}) + \frac{n_0 e^2 V}{\varepsilon_0 k_B T}$$

先考虑 $r > 0$ 的区域，有 (3 分)

$$\frac{d^2(rV)}{dr^2} = \frac{n_0 e^2 V r}{\varepsilon_0 k_B T}$$

所以可以解得 (3 分)

$$rV = A \exp\left(\sqrt{\frac{n_0 e^2}{\varepsilon_0 k_B T}} r\right) + B \exp\left(-\sqrt{\frac{n_0 e^2}{\varepsilon_0 k_B T}} r\right)$$

当 $r \rightarrow \infty$ 时， $rV \rightarrow 0$ ，因此有 $A = 0$ (1.5 分)；当 $r \rightarrow 0$ 时， $rV \rightarrow \frac{q}{4\pi\varepsilon_0}$ ，因此有 $B = \frac{q}{4\pi\varepsilon_0}$ (1.5 分)。因此有 (3 分)

$$V(r) = \frac{q}{4\pi\varepsilon_0 r} \exp\left(-\sqrt{\frac{n_0 e^2}{\varepsilon_0 k_B T}} r\right)$$

二. (60 分) 玻色爱因斯坦凝聚

玻色爱因斯坦凝聚是统计物理中一个十分重要的成果，本题中我们将会从玻色分布的基础，一步步计算玻色爱因斯坦凝聚。

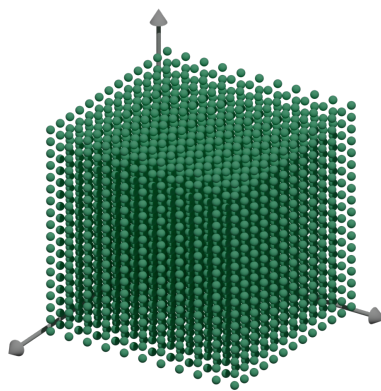
1. k-空间

i. 我们考察一个方形盒子中的自由粒子，粒子的物质波在方形盒子中形成驻波。定义物质波波矢为

$$\mathbf{k} = (k_x, k_y, k_z) = \left(\frac{2\pi}{\lambda_x}, \frac{2\pi}{\lambda_y}, \frac{2\pi}{\lambda_z}\right), \quad |\mathbf{k}| = \frac{2\pi}{\lambda}$$

假设方形盒子的三边边长为 l_x, l_y, l_z ，体积为 V ，其在三边上都形成驻波，给出 $\mathbf{k}$ 的可能取值。可以包含三个整数 n_x, n_y, n_z 。

ii. 可以看到自由粒子可以取得的 $\mathbf{k}$ 值在空间中形成一个方形格点。由此给出一个方形各点占据的 k-空间体积 V_k 。(你可以把这些格点看成晶格中的原子，给出一个小晶胞对应的体积。)



iii. 不同的 k 值对应了粒子不同的能量，给出能量与波矢之间的关系。由此给出能量在 $\varepsilon \sim \varepsilon + d\varepsilon$ 区间内一共有多少个可能的格点。

2. 对于玻色子，其分布满足玻色分布

$$f(\varepsilon) = \frac{1}{e^{\beta(\varepsilon - \mu)} - 1}$$

其中 μ 为体系的化学势， $\beta = \frac{1}{k_B T}$ 。其所指的是 k-空间中能量为 ε 一个格点对应的概率分布（即平均有 $f(\varepsilon)$ 个粒子处于 $\mathbf{k}$ 处）。

- i. 给出能量的概率分布函数（即 $\varepsilon \sim \varepsilon + d\varepsilon$ 区间内的粒子个数），如果这个分布函数在不断的降温过程中定义域始终为 $[0, +\infty)$ ，不存在奇点，给出 μ 的取值范围。
- ii. 当温度为定值 T 时，给出体系的总粒子数 N ，使用一个积分表示。试着说明总粒子数 N 与化学势 μ 的增减关系，并且计算 N 的上限。表达式使用 $\lambda = \frac{h}{\sqrt{2\pi m k_B T}}$ 代换。

数学小提示：

$$\int_0^\infty \frac{\sqrt{z}}{e^z - 1} dz = \frac{\sqrt{\pi}}{2} \zeta\left(\frac{3}{2}\right) \approx 2.31515$$

其中

$$\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s} = \frac{1}{\Gamma(s)} \int_0^\infty \frac{z^{s-1}}{e^z - 1} dz$$

- iii. 由上一问我们发现，玻色子的数密度 N/V 存在一个极大值。如果我们在此时继续塞入粒子，会发生什么？为何称其为凝聚？
- iv. 重复上文的计算，试着证明在一维与二维中，玻色爱因斯坦凝聚不会发生。（即证明不存在粒子数密度上限。）

[解答]

1. (20 分)

i. (4 分) 形成驻波时，有 (2 分)

$$l_x = \frac{n_x}{2} \lambda_x, l_y = \frac{n_y}{2} \lambda_y, l_z = \frac{n_z}{2} \lambda_z$$

因此有 (2 分)

$$\mathbf{k} = \left(\frac{n_x \pi}{l_x}, \frac{n_y \pi}{l_y}, \frac{n_z \pi}{l_z} \right), \quad n_x, n_y, n_z = 0, 1, 2, \dots$$

ii. (4 分) 利用前问可以得到 (2 分)

$$\Delta k_x = \frac{\pi}{l_x}, \Delta k_y = \frac{\pi}{l_y}, \Delta k_z = \frac{\pi}{l_z}$$

因此有 (2 分)

$$V_k = \Delta k_x \Delta k_y \Delta k_z = \frac{\pi^3}{l_x l_y l_z} = \frac{\pi^3}{V}$$

iii. (12 分) 根据德布罗意波关系 (2 分)

$$k = \frac{2\pi}{\lambda} = \frac{2\pi p}{h} = \frac{p}{\hbar}$$

因此有 (2 分)

$$\varepsilon = \frac{p^2}{2m} = \frac{\hbar^2 k^2}{2m}$$

对于 $k \sim k + dk$ 的 k -空间薄层，其对应的体积为 (2 分)

$$\frac{1}{2} \pi k^2 dk$$

由于一个格点占据的体积为 V_k ，薄层中所包含的格点数为 (2 分)

$$dN = \frac{1}{2} \pi k^2 dk / V_k = \frac{V k^2}{2\pi^2} dk$$

代入波矢与能量关系 (2 分)

$$k = \frac{\sqrt{2m\varepsilon}}{\hbar}, \quad dk = \frac{1}{\hbar} \sqrt{\frac{m}{2\varepsilon}} d\varepsilon$$

可以得到格点数为 (2 分)

$$dN = \frac{Vm^{3/2}}{\sqrt{2}\pi^2\hbar^3} \sqrt{\varepsilon} d\varepsilon$$

2. (40 分)

i. (8 分) 对于能量 ε 的态，每个格点有 $f(\varepsilon)$ 个格点，共有 dN 个态，因此可以得到 (2 分)

$$F(\varepsilon)d\varepsilon = f(\varepsilon)dN = \frac{Vm^{3/2}}{\sqrt{2}\pi^2\hbar^3} \frac{\sqrt{\varepsilon}}{e^{\beta(\varepsilon-\mu)} - 1} d\varepsilon$$

如果分母没有奇点，则 (2 分)

$$e^{\beta(\varepsilon-\mu)} \neq 1$$

因此有 (2 分)

$$\varepsilon \neq \mu$$

也就是 (2 分)

$$\mu < 0$$

ii. (6 分) 总粒子数可以通过积分得到 (2 分)

$$N = \frac{Vm^{3/2}}{\sqrt{2}\pi^2\hbar^3} \int_0^\infty \frac{\sqrt{\varepsilon} d\varepsilon}{e^{\beta(\varepsilon-\mu)} - 1}$$

当 μ 增加时，积分的分母减小，因此 N 增加。(2 分) 所以有 (2 分)

$$N < \frac{Vm^{3/2}}{\sqrt{2}\pi^2\hbar^3} \int_0^\infty \frac{\sqrt{\varepsilon} d\varepsilon}{e^{\beta\varepsilon} - 1} = \frac{Vm^{3/2}}{\sqrt{2}\pi^2\hbar^3 \beta^{3/2}} \int_0^\infty \frac{\sqrt{z} dz}{e^z - 1} = \frac{V}{\lambda^3} \zeta(3/2) \approx 2.612 \frac{V}{\lambda^3}$$

iii. (6 分) 当达到 N/V 的最大值时，此时 $\mu = 0$ ，因此能量为 0 的态能够存在的粒子数量为 (2 分)

$$f(0) \rightarrow \infty$$

所以添加的粒子将全部聚集于 k -空间能量为 0 的原点处，(2 分) 也就是“凝聚”在同一个态上。(2 分)

iv. (20 分) 在一维情况下，有 $k \sim k + dk$ 的短线段上，态的个数为 (2 分)

$$dN = \frac{dk}{\pi/l} = \frac{l dk}{\pi}$$

使用能量可以表示为 (2 分)

$$dN = \frac{l}{\pi\hbar} \sqrt{\frac{m}{2\varepsilon}} d\varepsilon$$

因此有 (2 分)

$$F_1(\varepsilon)d\varepsilon = \frac{lm^{1/2}}{\sqrt{2}\pi\hbar} \frac{\varepsilon^{-1/2} d\varepsilon}{e^{\beta(\varepsilon-\mu)} - 1}$$

当 $\mu = 0$ 时，粒子数密度达到最大值，发生凝聚，有 (2 分)

$$N < \frac{lm^{1/2}}{\sqrt{2\pi\hbar\beta^{1/2}}} \int_0^\infty \frac{z^{-1/2}dz}{e^z - 1}$$

可以看到积分中的值为 (2 分)

$$I = \Gamma(1/2)\zeta(1/2) = \sqrt{\pi} \sum_{n=1}^{\infty} \frac{1}{\sqrt{n}} > \sqrt{\pi} \int_1^\infty \frac{dx}{\sqrt{x}} \rightarrow \infty$$

因此粒子数上限的积分是发散的，不存在粒子数密度最大值，不会发生凝聚。

在二维情况下，有 $k \sim k + dk$ 的薄扇环上，态的个数为 (2 分)

$$dN = \frac{\pi k dk / 2}{\pi^2 / A} = \frac{A k dk}{2\pi}$$

使用能量代换可以得到 (2 分)

$$dN = \frac{A m d\varepsilon}{2\pi\hbar^2}$$

所以 (2 分)

$$F_2(\varepsilon)d\varepsilon = \frac{A m}{2\pi\hbar^2} \frac{d\varepsilon}{e^{\beta(\varepsilon-\mu)} - 1}$$

当 $\mu = 0$ 时有 (2 分)

$$N < \frac{A m}{2\pi\hbar^2\beta} \int_0^\infty \frac{dz}{e^z - 1}$$

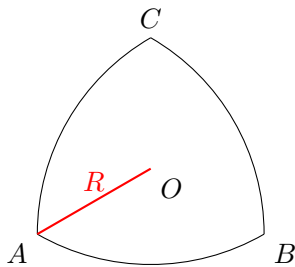
可以看到积分中的值为 (2 分)

$$I = \Gamma(1)\zeta(1) = \sum_{n=1}^{\infty} \frac{1}{n} > \int_1^\infty \frac{dx}{x} \rightarrow \infty$$

因此粒子数上限的积分是发散的，不存在粒子数密度最大值，不会发生凝聚。

三. (40 分) 有趣的莱洛三角形

莱洛三角形是一种等宽曲线，其由三个完全相同，圆心角为 $\pi/3$ 的圆弧组成。我们现在考察质量为 M 的一个莱洛三角形刚体，刚体的外接圆半径为 R 。



1. 计算莱洛三角形的转动惯量 I 。(提示：圆心角为 α ，半径为 R 的扇形的质心到圆心距离为 $r_c = \frac{4R \sin(\alpha/2)}{3\alpha}$ 。质量为 M 外接圆半径为 R 的正 n 边形的转动惯量为 $I = \frac{1}{3}MR^2 \left(\frac{1}{2} + \cos^2 \left(\frac{\pi}{n} \right) \right)$ 。)
2. 现在莱洛三角形在粗糙的地面上做无滑动的小振动，给出小振动周期。可以使用 OC 连线与竖直方向的小夹角 θ 描述莱洛三角形的运动。

3. 当 OC 与竖直方向夹角为 θ (不是小角) 时, 质心从离地高度为 h 处静止释放下落。其 AB 边 (不包含 A, B 两点) 与十分粗糙的地面做完全非弹性碰撞, 给出碰撞后莱洛三角形的角速度 Ω 。q

[解答]

1. (20 分)

将转动惯量分为两个部分计算: 三角形 ABC 与三个圆心角为 $\pi/3$ 的扇形。莱洛三角形的面积为 (2 分)

$$S = 3S_{\text{扇}} - 2S_{\Delta} = \frac{\pi}{2}(\sqrt{3}R)^2 - 2 \cdot \frac{\sqrt{3}}{4}(\sqrt{3}R)^2 = \left(-\frac{3\sqrt{3}}{2} + \frac{3\pi}{2}\right) R^2$$

因此莱洛三角形的面密度为 (2 分)

$$\sigma = \frac{M}{S} = \frac{M}{\left(-\frac{3\sqrt{3}}{2} + \frac{3\pi}{2}\right) R^2}$$

对于三角形 ABC , 其转动惯量为 (2 分)

$$I_{\Delta} = \frac{1}{3}M_{\Delta}R^2 \left(\frac{1}{2} + \cos^2 \frac{\pi}{3}\right) = \frac{1}{4}M_{\Delta}R^2$$

对于扇形 $\widehat{AB}C$, 其关于 C 点的转动惯量为 (2 分)

$$I_C = \frac{1}{2}M_{\text{扇}}(\sqrt{3}R)^2$$

则质心位置与扇形关于质心的转动惯量为 (3 分)

$$r_c = \frac{2\sqrt{3}}{\pi}R, \quad I_{\text{质心}} = \frac{3}{2}M_{\text{扇}}R^2 - M_{\text{扇}}r_c^2 = \left(\frac{3}{2} - \frac{12}{\pi^2}\right)M_{\text{扇}}R^2$$

因此扇形关于莱洛三角形的中心 O 的转动惯量为 (3 分)

$$r_{co} = \left(\frac{2\sqrt{3}}{\pi} - 1\right)R, \quad I_O = I_{\text{质心}} + M_{\text{扇}}r_{co}^2 = \left(\frac{5}{2} - \frac{4\sqrt{3}}{\pi}\right)M_{\text{扇}}R^2$$

代入 (4 分)

$$M_{\Delta} = \frac{\sqrt{3}}{4}(\sqrt{3}R)^2\sigma = \frac{\sqrt{3}M}{2(\pi - \sqrt{3})}, \quad M_{\text{扇}} = \frac{\pi}{6}(\sqrt{3}R)^2\sigma = \frac{\pi M}{3(\pi - \sqrt{3})}$$

因此有 (2 分)

$$I = 3I_{\text{扇}} - 2I_{\Delta} = \frac{10\pi - 17\sqrt{3}}{4(\pi - \sqrt{3})}MR^2 \approx 0.34959MR^2$$

2. (10 分)

在地面上振动时, 根据等宽曲线的性质, C 点到地面距离不变, 为 $\sqrt{3}R$ 。若转过小角 θ , 质心上升带来的重力势能变化为 (2 分)

$$E_p = MgR(1 - \cos \theta) \approx \frac{MgR}{2}\theta^2$$

在纯滚时, 其相当于绕地面接触点做定轴转动, 其动能为 (2 分)

$$E_k = \frac{1}{2}I_{\text{接触点}}\dot{\theta}^2 = \frac{1}{2}(I + M(\sqrt{3} - 1)^2R^2)\dot{\theta}^2$$

因此有 (4 分)

$$M_{\text{eff}} = I + MR^2(4 - 2\sqrt{3}), \quad K_{\text{eff}} = Mgr$$

所以有 (2 分)

$$T = 2\pi\sqrt{\frac{M_{\text{eff}}}{K_{\text{eff}}}} = 2\pi\sqrt{\frac{-33\sqrt{3} - 8\sqrt{3}\pi + 24 + 26\pi}{4(\pi - \sqrt{3})} \frac{R}{g}} \approx 5.9125\sqrt{\frac{R}{g}}$$

3. (10 分)

刚体在下落到地面时，质心一共下降的距离为 (2 分)

$$h' = h - (\sqrt{3} - \cos\theta)R$$

刚体落地时的速度为 (2 分)

$$v = \sqrt{2gh'}$$

刚体在落地后，由于发生了完全非弹性碰撞，与地面接触点的速度在垂直地面方向上速度为 0。由于地面十分粗糙，平行方向上速度也为 0。相当于碰撞之后刚体绕接触点转动。因此有角动量守恒 (2 分)

$$MvR\sin\theta = I_{\text{接触点}}\Omega$$

而 (2 分)

$$I_{\text{接触点}} = I + (\sqrt{3}^2 + 1^2 - 2\sqrt{3} \cdot 1 \cdot \cos\theta)MR^2 = \left(\frac{26\pi - 33\sqrt{3}}{4(\pi - \sqrt{3})} - 2\sqrt{3}\cos\theta\right)MR^2$$

因此有 (2 分)

$$\Omega = \frac{MvR\sin\theta}{\left(\frac{26\pi - 33\sqrt{3}}{4(\pi - \sqrt{3})} - 2\sqrt{3}\cos\theta\right)MR^2} = \frac{\sqrt{2g(h + (\cos\theta - \sqrt{3})R)}\sin\theta}{\left(\frac{26\pi - 33\sqrt{3}}{4(\pi - \sqrt{3})} - 2\sqrt{3}\cos\theta\right)R}$$

四. (30 分) 黑洞热力学

长期以来，天体物理学家一直在研究黑洞：这些是大质量恒星在引力作用下无法自我支撑而达到的最终状态。随着物质不断向中心坍缩，最终逃逸速度达到光速。在此之后，坍缩的物质无法再将信息传递到外部。一个质量为 M 的黑洞的半径为

$$R_s = \frac{2GM}{c^2},$$

其中 $G = 6.67 \times 10^{-8} \text{ cm}^3/\text{g} \cdot \text{s}^2$ 是引力常数， $c = 3 \times 10^{10} \text{ cm/s}$ 是光速。

霍金通过结合量子力学和广义相对论的方法，计算了黑洞的辐射功率。他得出了一个惊人的结果：黑洞会以一定的温度发射完美的黑体辐射：

$$T_{\text{bh}} = \frac{\hbar c^3}{8\pi GM k_B},$$

其中 $\hbar$ 是约化普朗克常数， k_B 是玻尔兹曼常数。

根据爱因斯坦的理论，黑洞的能量为 $E = Mc^2$ 。

1. 计算质量为 M 的黑洞的热容。

2. 计算黑洞的熵。(使用黑洞的表面积 A 与普朗克长度 $L^* = \sqrt{\frac{\hbar G}{c^3}}$ 表示结果, 当 M 为 0 时, 黑洞的熵可以认为是 0。)
3. 计算一个 1cm 的黑洞能存下的比特数, 使用 TB 为单位。

[解答]

1. (6 分)

黑洞的能量为 (3 分)

$$E = Mc^2 = \frac{\hbar c^5}{8\pi G k_B T_{\text{bh}}}$$

因此热容为 (3 分)

$$C = \frac{dE}{dT_{\text{bh}}} = -\frac{\hbar c^5}{8\pi G k_B T_{\text{bh}}^2} = -\frac{\hbar c^5}{8\pi G k_B} \cdot \left(\frac{8\pi G M k_B}{\hbar c^3}\right)^2 = -\frac{8\pi G M^2 k_B}{\hbar c}$$

2. (7 分)

利用熵的定义 (3 分)

$$dS = \frac{dQ}{T} = \frac{dMc^2}{T} = \frac{8\pi G M k_B dM}{\hbar c}$$

因此有 (4 分)

$$S = \frac{4\pi G M^2 k_B}{\hbar c} = \frac{4\pi G k_B}{\hbar c} \cdot \left(\frac{R_g c^2}{2G}\right)^2 = \frac{k_B}{4} \frac{A}{L^{*2}}$$

3. (17 分)

对于 N 个比特, 其对应状态数为 (4 分)

$$\Omega = 2^N$$

其熵为 (4 分)

$$S = k_B \ln \Omega = N k_B \ln 2$$

这与黑洞的熵应当相等, 也就是 (4 分)

$$N = \frac{1}{4 \ln 2} \cdot \frac{A}{L^{*2}} \approx 1.735 \times 10^{66}$$

利用换算 (2 分)

$$1 \text{ TB} = 2^{10} \text{ GB} = 2^{20} \text{ MB} = 2^{30} \text{ KB} = 2^{40} \text{ B} = 2^{43} \text{ bit}$$

因此存储的信息量为 (3 分) $1.97 \times 10^{53} \text{ TB}$ 。

五. (40 分) 戴森球计划

1. 地球上的生命可以被视为一种热机, 从高温热源 (太阳, 温度为 $T_S = 6000 \text{ K}$) 吸收能量, 并将其排放到低温热源 (星际空间, 微波背景温度为 $T_{\text{MB}} = 2.725 \text{ K}$)。在地球轨道上的太阳能量通量为 $\Phi_S = 1370 \text{ W/m}^2$, 地球的半径约为 $r_E = 6.4 \times 10^6 \text{ m}$ 。

如果地球上的生命是完全高效的 (即使用卡诺循环, 高温热源温度为 T_S , 低温热源温度为 T_{MB}), 那么可以从这种能量流中提取多少有用功 (以瓦特为单位)? 并将其与世界市场能源消耗估计值 $4.5 \times 10^{20} \text{ J/year}$ 进行比较。(有

用常数：一年大约有 $\pi \times 10^7 \text{ s}$ 。)

- 如果我们建造一个半径为地球轨道半径 (约 1.5 亿公里, $R_{\text{ES}} \approx 1.5 \times 10^{11} \text{ m}$) 的球体将太阳包围, 我们文明可用的有用功将增加多少倍?
- 这个巨大的建设项目被称为戴森球 (Dyson sphere), 以提出这一概念的物理学家命名。他建议通过观测大型红外辐射源来寻找先进文明。然而, 地球并不以星际空间的温度辐射能量。它的辐射接近一个黑体, 温度约为 $T_E = 300 \text{ K} = 23^\circ\text{C}$ 。如果我们的热量必须有效地辐射到一个 300 K 的低温热源, 而不是 T_{MB} 的低温热源, 假设在这两种情况下我们都运行卡诺热机, 那么我们提取太阳能量通量的有用功的效率会降低多少?
- 还有一种替代观点是追踪熵而不是能量。生物通过将产生的熵排放到从太阳流向星际空间的能量流中, 来维持和复制它们的低熵状态。新的记忆存储本质上涉及熵的产生; 随着我们进入信息时代, 我们可能最终会更关心熵的排放, 而不是做功。类似于前问的“做功效率”(实际做功与给定高温热源和低温热源时卡诺上限的比值), 我们可以估计一种“熵排放效率”(实际添加到能量流中的熵与在相同高温热源和低温热源下可能添加的熵的比值)。每秒有多少熵以低熵、高能辐射的形式从太阳到达地球? 当太阳能量通量以温度 $T_E = 300 \text{ K}$ 辐射出去时, 每秒有多少熵离开地球? 若直接在太阳与外太空建立热机, 其对应的单位时间熵排放是多少? 两者比较计算熵排放效率 f 。从熵排放的角度来看, 高温热源温度 T_S 和低温热源温度 (分别为 T_E 或 T_{MB}) 哪个更重要?
- 在生成有用功方面, 太阳是关键, 而夜空几乎无足轻重。但对于排放文明产生的熵而言, 夜空却是生命的赋予者和机遇的领域。这两种视角并不矛盾。在某些情况下, 给定量的功能在低温下更有用。戴森后来推测生命如何通过更低的温度下运行来高效利用这一点。因此, 一个高度先进的信息文明将倾向于在微波波段而非红外波段辐射。为实现这一点, 需要扩大戴森球的面积: 更大的球体可以将太阳能量流以更低温度的黑体辐射形式重新辐射出去。

戴森球需要多大的半径 R_D 才能达到 50% 的熵排放效率? 该半径与冥王星轨道半径 ($R_{\text{PS}} \approx 6 \times 10^{12} \text{ m}$) 相比如何? 若以比特为单位测量熵 (使用 $k_S = \frac{1}{\ln 2}$ 而非 $k_B = 1.3807 \times 10^{-23} \text{ J/K}$), 我们的超级文明每秒可排放多少比特的熵? 评论你所得到的结论。

[解答]

1. (7 分)

太阳在地球处的辐射功率 (1 分)

$$P_S = \Phi_S \cdot \pi r_E^2 = 1.7629 \times 10^{17} \text{ W}$$

而若在太阳与宇宙之间建立卡诺热机, 其效率为 (2 分)

$$\eta = 1 - \frac{T_{\text{MB}}}{T_S} = 1 - \frac{2.725}{6000} = 0.99955$$

因此在太阳辐射的能量流中, 我们能提取的有用功为 (1 分)

$$P_{\text{有用}} = \eta P_S = 1.7621 \times 10^{17} \text{ W}$$

在一年之内, 可以利用的有用功为 (2 分)

$$W_{\text{有用}} = P_{\text{有用}} t = 5.5358 \times 10^{24} \text{ J/year}$$

与目前世界市场能源消耗估计值相比, (1 分)

$$\frac{W}{W_{\text{有用}}} = 8.13 \times 10^{-5}$$

目前人们利用的能量不到可利用的万分之一。

2. (4 分)

如果使用一个半径等于地球轨道半径的球体，我们可以截获所有来自太阳的能量，为 (2 分)

$$P_{S,\text{tot}} = \Phi_S \cdot 4\pi R_E^2$$

因此增加的倍数为 (2 分)

$$N = \frac{\eta P_{S,\text{tot}} - \eta P_S}{\eta P_S} = \frac{4R_E^2}{r_E^2} - 1 = 2.1973 \times 10^9$$

3. (4 分)

如果使用地球作为冷源，则卡诺热机的效率为 (2 分)

$$\eta_E = 1 - \frac{T_E}{T_S} = 0.95$$

因此效率的减小量为 (1 分)

$$\Delta\eta = \eta - \eta_E = 0.049546$$

相对减小量为 (1 分)

$$\frac{\Delta\eta}{\eta} = 0.049568$$

4. (11 分)

太阳单位时间给地球辐射的熵为 (2 分)

$$S_S = \frac{P_S}{T_S} = 2.9382 \times 10^{13} \text{ W/K}$$

地球单位时间向外辐射的功率等于太阳对地球照射的功率，因此地球单位时间向外辐射的熵为 (2 分)

$$S_E = \frac{P_S}{T_E} = 5.8764 \times 10^{14} \text{ W/K}$$

而若直接将太阳的热源排放到 T_{MB} 的冷源，单位时间辐射的熵为 (2 分)

$$S_{\text{MB}} = \frac{P_S}{T_{\text{MB}}} = 6.4694 \times 10^{16} \text{ W/K}$$

因子 f 指的是两种情况下对能量流添加的熵比值，因此有 (2 分)

$$f = \frac{S_E - S_S}{S_{\text{MB}} - S_S} = \frac{P_S/T_E - P_S/T_S}{P_S/T_{\text{MB}} - P_S/T_S} = \frac{1/T_E - 1/T_S}{1/T_{\text{MB}} - 1/T_S} = 0.0086331$$

为了分析重要性，可以计算 f 关于三个温度的偏导数，有 (2 分)

$$\begin{aligned} \frac{\partial f}{\partial T_S} &= 7.5 \times 10^{-8} \text{ K}^{-1} \\ \frac{\partial f}{\partial T_E} &= -3.0 \times 10^{-5} \text{ K}^{-1} \\ \frac{\partial f}{\partial T_{\text{MB}}} &= 3.2 \times 10^{-3} \text{ K}^{-1} \end{aligned}$$

可见低温热源温度更为重要，尤其是 T_{MB} 。(1 分)

5. (14 分)

当 $f = 50\%$ 时，要求戴森球向外辐射的温度 T_{DS} 满足 (2 分)

$$\frac{1/T_{DS} - 1/T_S}{1/T_{MB} - 1/T_S} = 0.5$$

解得 (2 分)

$$T_{DS} = 5.4475 \text{ K}$$

对于而其向外辐射的总功率等于太阳辐射的总功率，大小为 (2 分)

$$P_{S,\text{tot}} = 4\pi R_{DS}^2 \cdot \sigma T_{DS}^4 = \Phi_S \cdot 4\pi R_E^2$$

因此有 (2 分)

$$R_{DS} = R_E \sqrt{\frac{\Phi_S}{\sigma T_{DS}^4}} = 7.8568 \times 10^{14} \text{ m}$$

与冥王星的轨道半径相比，(2 分)

$$\frac{R_{DS}}{R_{PS}} \approx 130.95$$

若以比特作为单位，单个比特的熵大小为 $k_B \ln 2$ ，则我们的超级文明单位时间排放的熵为 (2 分)

$$S_{\text{bit}} = \frac{P_{S,\text{tot}}}{T_{DS}} \cdot \frac{1}{k_B \ln 2} = 7.43 \times 10^{48} \text{ s}^{-1}$$

可以看到，戴森球的半径比冥王星轨道还要大近 130 倍，是十分巨大的太空结构，说明这个超级文明具有强大的星际工程建设能力，并且能够创造出能承受极限强度的材料。再者，从信息熵的角度看，这一超级文明时时刻刻都在产生大量的信息，科技技术等在以难以想象的速度发展。(2 分)

六. (80 分) 抓住作弊赌徒：概率分布空间中的距离 [极难]

本题中积分都为全空间的积分，故不再重复书写积分区间。

你是一个赌场经理，你想要确保所有来到赌场的赌徒没有在赌博游戏中作弊。你知道每一个赌博游戏中出现的结果都会满足一定的概率分布，你希望去验证赌徒们在玩游戏时得到的结果时满足所预测的概率分布的，如果出现偏差，这意味着他们作弊了。一种方法是将赌徒玩游戏得到的分布与游戏真实的分布比对，计算两者之间的“距离”，如果两者偏差过大，那么意味着赌徒作弊了。

1. 为了研究概率分布的“距离”，我们不妨先从在生活中最普遍的分布开始说起：高斯分布。假设一个赌博游戏满足高斯分布，每次游戏会得到 N 个数 $\mathbf{x} = (x_1, x_2, \dots, x_N)$ 。当然，机器内部有一系列可调参数，我们用矢量 $\boldsymbol{\theta}$ 表示。 $\mathbf{x}$ 满足高斯分布

$$P(\mathbf{x}|\boldsymbol{\theta}) = \frac{e^{-\sum_i (y_i(\boldsymbol{\theta}) - x_i)^2 / (2\sigma^2)}}{(2\pi\sigma^2)^{N/2}}$$

这里函数 $\mathbf{y}(\boldsymbol{\theta}) = (y_1(\boldsymbol{\theta}), y_2(\boldsymbol{\theta}), \dots, y_N(\boldsymbol{\theta}))$ 可以理解成参数 $\boldsymbol{\theta}$ 影响结果分布的方式，而 σ 表示分布的标准差，是机器出厂就规定好的，无法更改。

- i. 如果赌徒在玩了一次游戏后得到的值为 $\mathbf{x}$ ，我们可以由此反推此时机器最有可能的参数为 $\boldsymbol{\phi}$ 。试问一个参数为 $\boldsymbol{\theta}$ 的机器的一次游戏结果反推出参数 $\boldsymbol{\phi}$ 的概率密度。

- ii. 通过前问计算，我们发现概率密度只和矢量 $\mathbf{y}(\boldsymbol{\theta})$ 与 $\mathbf{y}(\boldsymbol{\phi})$ 之间的距离有关。由此我们不妨定义一个空间，参数 $\boldsymbol{\theta}$ 对应空间中的点 $\mathbf{y}(\boldsymbol{\theta})/\sigma$ 。由此，我们可以定义不同参数对应的概率分布的距离 $d(P(\mathbf{x}|\boldsymbol{\theta}), P(\mathbf{x}|\boldsymbol{\phi})) = |\mathbf{y}(\boldsymbol{\theta})/\sigma - \mathbf{y}(\boldsymbol{\phi})/\sigma|$ 。在这样的距离定义下， $\boldsymbol{\theta}$ 形成的参数空间（也可以理解为概率分布空间，因为参数 $\boldsymbol{\theta}$ 与概率分布 $P(\mathbf{x}|\boldsymbol{\theta})$ 一一对应）是怎样的？

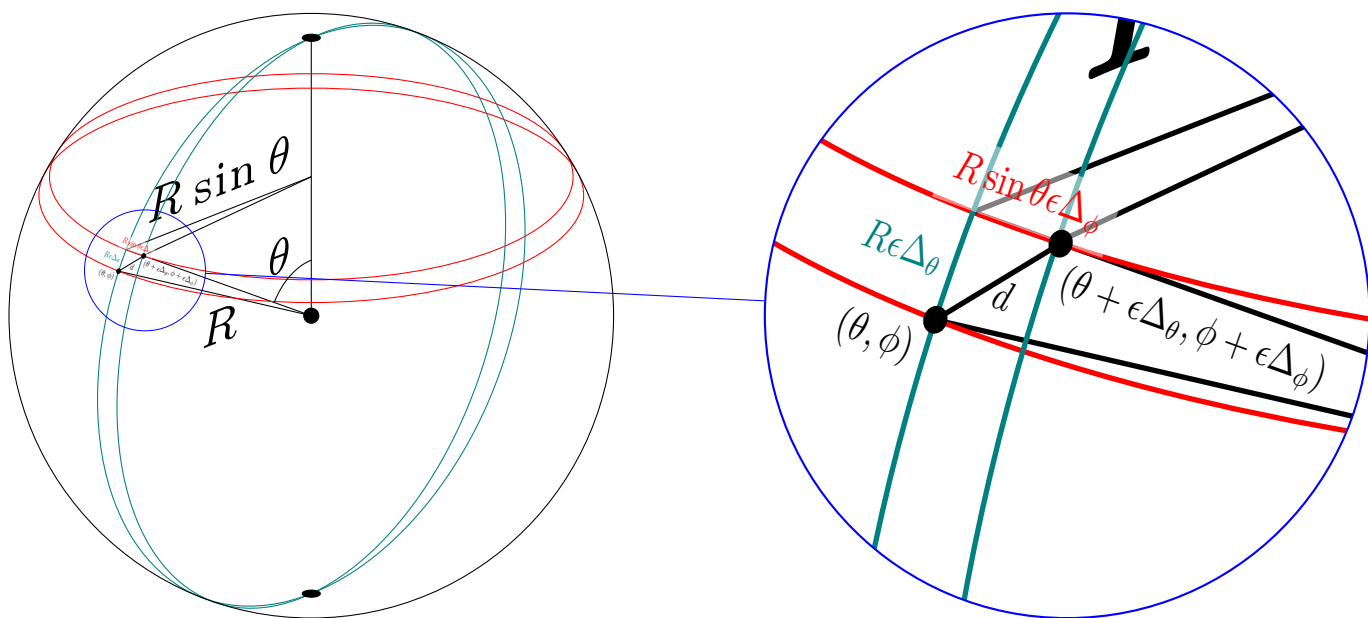
我们常常使用度规矩阵来表述空间的形状，其定义如下：

$$d^2(P(\mathbf{x}|\boldsymbol{\theta}), P(\mathbf{x}|\boldsymbol{\theta} + \epsilon\boldsymbol{\Delta})) = \epsilon^2 \sum_{\alpha, \beta} g_{\alpha\beta} \Delta_\alpha \Delta_\beta$$

其中 ϵ 是小量，而 α, β 指的是不同分量的下标。

我们不妨试着计算球面上的度规。对于点 (θ, ϕ) 与点 $(\theta + \epsilon\Delta_\theta, \phi + \epsilon\Delta_\phi)$ ，两点之间的距离为

$$d^2 = (R \sin \theta \epsilon \Delta_\phi)^2 + (R \epsilon \Delta_\theta)^2 = \epsilon^2 (R^2 \Delta_\theta^2 + R^2 \sin^2 \theta \Delta_\phi^2)$$



由此可以得到

$$g_{\theta\theta} = R^2, \quad g_{\phi\phi} = R^2 \sin^2 \theta, \quad g_{\theta\phi} = g_{\phi\theta} = 0$$

计算分布 $P(\mathbf{x}|\boldsymbol{\theta})$ 与 $P(\mathbf{x}|\boldsymbol{\theta} + \epsilon\boldsymbol{\Delta})$ 的距离，由此证明度规 $g_{\alpha\beta} = (J^T J)_{\alpha\beta} / \sigma^2$ ，其中 J 表示雅可比矩阵， J^T 表示矩阵的转置，也就是把行与列交换，也就是 $(J^T)_{\alpha i} = J_{i\alpha}$ 。有 $J_{i\alpha} = \frac{\partial y_i}{\partial \theta_\alpha}$ ，下标 $\alpha\beta$ 表示这个矩阵的第 α 行，第 β 列，一般而言这两者都是数字，不过也有时候使用其对应的有物理意义的字母替代。¹（提示：使用泰勒级数将距离平方展开到 ϵ 的二阶。）

- iii. 对于任意形式的概率分布，一种自然的描述临近分布距离的度规是 FIM 度规，其表达式如下：

$$g_{\alpha\beta}(\boldsymbol{\theta}) = - \left\langle \frac{\partial^2 \ln P(\mathbf{x}|\boldsymbol{\theta})}{\partial \theta_\alpha \partial \theta_\beta} \right\rangle_{\mathbf{x}}$$

重新计算度规 $g_{\alpha\beta}$ ，并证明你刚刚得到的与现在重新计算的结果相同。这里尖括号 $\langle \cdots \rangle_{\mathbf{x}}$ 表示关于 $\mathbf{x}$ 取平均，也就是 $\int \cdots P(\mathbf{x}|\boldsymbol{\theta}) d\mathbf{x}$

¹矩阵乘法的分量形式如下： $(AB)_{ij} = \sum_k A_{ik} B_{kj}$

iv. 利用 FIM 度规，我们能够更进一步，探讨一般的概率分布 $P(\mathbf{x}|\boldsymbol{\theta})$ 。试着计算参数 $\boldsymbol{\theta}$ 的机器的一次游戏给出临近分布 $P(\mathbf{x}|\boldsymbol{\theta} + \epsilon\boldsymbol{\Delta})$ 的概率密度与给出原本分布 $P(\mathbf{x}|\boldsymbol{\theta})$ 的概率密度的比值，结果利用 $g_{\alpha\beta}$ 表示（提示：试着将 $\ln P(\mathbf{x}|\boldsymbol{\theta} + \epsilon\boldsymbol{\Delta})$ 关于 ϵ 展开到二阶，然后对其取 e 指数。）

2. 除了上述的 FIM 度规之外，还有许多不同的距离测量方法。不过距离函数需要满足如几个性质：

- 正定性：对于分布 P 与 Q ，两者距离 $d(P, Q) \geq 0$ ，且当且仅当 $P = Q$ 时， $d(P, Q) = 0$ ；
- 对称性：对于分布 P 与 Q ，两者距离应满足 $d(P, Q) = d(Q, P)$ ；
- 三角不等式：对于分布 P, Q, R ，应当满足 $d(P, Q) \leq d(P, R) + d(R, Q)$ ；
- 当分布 P 与 Q 相差极大时，我们期望 $d(P, Q) \rightarrow \infty$ 。

现在赌场里有一个投骰子的游戏，一名赌徒带了偷偷掉包了一个骰子，使得其六面的概率分别为 0.1, 0.1, 0.1, 0.1, 0.1, 0.5，而正常的骰子六面的概率都为 1/6。

i. 一个很自然的方法是将骰子的概率分布看成一个单位球上的点。为什么？注意到对于骰子六个面的概率 P_i ，有 $\sum_i P_i = 1$ ，因此可以将其视为点 $(\sqrt{P_1}, \sqrt{P_2}, \dots, \sqrt{P_n})$ 。基于此我们可以定义概率分布的点乘 $P \cdot Q = \sum_i \sqrt{P_i} \sqrt{Q_i} = \int \sqrt{P(\mathbf{x})} \sqrt{Q(\mathbf{x})} d\mathbf{x}$ 。由此我们可以定义 Hellinger 距离（三个等价表述）

$$d_{\text{Hel}}^2(P, Q) = \sum_i \left(\sqrt{P_i} - \sqrt{Q_i} \right)^2 = \int \left(\sqrt{P(\mathbf{x})} - \sqrt{Q(\mathbf{x})} \right)^2 d\mathbf{x} = 2 - 2P \cdot Q$$

给出假骰子与真骰子分布之间的距离。对于两个临近分布，计算两者距离并且给出这一距离函数对应的度规，你可以假定分布有一个参数 $\boldsymbol{\theta}$ ，然后用关于 $\boldsymbol{\theta}$ 的偏导数表示，最后化简到类似于 FIM 度规的形式。证明这个距离存在上限 $\sqrt{2}$ 。

ii. 为了检测赌徒是否作弊，我们肯定需要观察其多次游戏的结果，由此我们可以定义多次游戏结果的概率分布 P_N 。将赌徒 N 游玩的结果做成一个新的随机变量 $\mathbf{j} = (j_1, j_2, \dots, j_N)$ ，我们考察其上的概率分布 $P_N(\mathbf{j})$ 。其中每一次结果 j_i 都是互相独立，并且遵循相同的分布 $P(j)$ 。试着用 $P(j)$ 表示 $P_N(\mathbf{j})$ ，并且由此证明 $P_N \cdot Q_N = (P \cdot Q)^N$ 。计算在赌徒玩了 100 次之后其对应分布与真实骰子玩 100 次分布的距离 $d_{\text{Hel}}(P_{100}, Q_{100})$ 。评述一下 Hellinger 距离在抓捕作弊赌徒上的效果。

iii. 为了解决 Hellinger 距离的有限性，我们不妨来考察由其延申出来的另一种距离：Bhattacharyya 距离。其定义如下：

$$d_{\text{Bhatt}}^2(P, Q) = \frac{1}{2} \lim_{N \rightarrow 0} \frac{d_{\text{Hel}}^2(P_N, Q_N)}{N} = -\ln(P \cdot Q)$$

试着证明上述等式成立。提示：当 $N \ll 1$ 时 $a^N \approx 1 + N \ln a$

iv. 在机器学习中，我们还会遇到另一种概率分布的距离：Kullback-Liebler 散度。其定义如下

$$d_{\text{KL}}(Q|P) = -\sum_i P_i \ln \left(\frac{Q_i}{P_i} \right) = -\int P(\mathbf{x}) \ln \left(\frac{Q(\mathbf{x})}{P(\mathbf{x})} \right) d\mathbf{x}$$

证明 Kullback-Liebler 散度具有正定性，即 $d_{\text{KL}}(Q|P) \geq 0$ 且只有在 $P = Q$ 是取等，并给出 Kullback-Liebler 散度给出的度规。（提示： $\ln(x) \leq x - 1$ 当且仅当 $x = 1$ 时取等。）

v. 可以注意到 Kullback-Liebler 散度是不对称的，因此我们可以将其对称化

$$d_{\text{sKL}}(Q, P) = \frac{1}{2} (d_{\text{KL}}(Q|P) + d_{\text{KL}}(P|Q)) = \frac{1}{2} \int (P(\mathbf{x}) - Q(\mathbf{x})) \ln \left(\frac{P(\mathbf{x})}{Q(\mathbf{x})} \right) d\mathbf{x}$$

对于这个对称化的 Kullback-Liebler 散度和 Bhattacharyya 距离有许多共性，有好有坏。

试着证明两个距离都是“广延”的，即距离 d 与重复次数 N 成线性关系，也就是证明 $d_{\text{Bhatt}}^2(P_N, Q_N) \propto N, d_{\text{KL}}(Q_N, P_N) \propto N$ 。证明这两者不符合三角不等式，但满足其他距离函数需要满足的定律。最后对于两个方差都是 σ 的高斯分布 $P(\mathbf{x}|\theta)$ 与 $P(\mathbf{x}|\phi)$ ，计算两者的对称 Kullback-Liebler 散度和 Bhattacharyya 距离。
提示： $\left[\int f(x)g(x)dx\right]^2 \leq \int [f(x)]^2 dx \cdot \int [g(x)]^2 dx$ 当且仅当 $f(x) = c \cdot g(x)$ 时取等。

[解答]

1. (18 分)

i. (3 分) 当赌徒玩游戏得到的值为 $\mathbf{x}$ 时，其最有可能的参数 ϕ 显然满足 (1 分)

$$\mathbf{y}(\phi) = \mathbf{x}$$

因为此时 $\mathbf{x}$ 对应的概率密度最大。由此可以得到参数 ϕ 的概率密度为 (2 分)

$$P(\mathbf{y}(\phi)|\theta) = \frac{e^{-\sum_i (y_i(\theta) - y_i(\phi))^2 / (2\sigma^2)}}{(2\pi\sigma^2)^{N/2}} = \frac{e^{-|\mathbf{y}(\theta) - \mathbf{y}(\phi)|^2 / (2\sigma^2)}}{(2\pi\sigma^2)^{N/2}}$$

ii. (5 分) $P(\mathbf{x}|\theta)$ 与 $P(\mathbf{x}|\theta + \epsilon\Delta)$ 之间的距离平方为, (3 分)

$$\begin{aligned} d^2 &= \frac{1}{\sigma^2} (\mathbf{y}(\theta + \epsilon\Delta) - \mathbf{y}(\theta))^2 \\ &= \frac{1}{\sigma^2} \left(\mathbf{y}(\theta) + \epsilon \sum_{\alpha} \frac{\partial \mathbf{y}}{\partial \theta_{\alpha}} \Delta_{\alpha} - \mathbf{y}(\theta) \right)^2 \\ &= \frac{1}{\sigma^2} \left(\epsilon \sum_{\alpha} \sum_i \frac{\partial y_i}{\partial \theta_{\alpha}} \Delta_{\alpha} \hat{x}_i \right)^2 \\ &= \frac{\epsilon^2}{\sigma^2} \sum_i \sum_{\alpha} \frac{\partial y_i}{\partial \theta_{\alpha}} \Delta_{\alpha} \hat{x}_i \cdot \sum_j \sum_{\beta} \frac{\partial y_j}{\partial \theta_{\beta}} \Delta_{\beta} \hat{x}_j \\ &\stackrel{\hat{x}_i \cdot \hat{x}_j = \delta_{ij}}{=} \frac{\epsilon^2}{\sigma^2} \sum_i \sum_{\alpha, \beta} \frac{\partial y_i}{\partial \theta_{\alpha}} \frac{\partial y_i}{\partial \theta_{\beta}} \Delta_{\alpha} \Delta_{\beta} \end{aligned}$$

因此度规 (2 分)

$$g_{\alpha\beta} = \frac{1}{\sigma^2} \sum_i \frac{\partial y_i}{\partial \theta_{\alpha}} \frac{\partial y_i}{\partial \theta_{\beta}} = \frac{1}{\sigma^2} \sum_i J_{i\alpha} J_{i\beta} = \frac{1}{\sigma^2} \sum_i (J^T)_{\alpha i} J_{i\beta} = \frac{(J^T J)_{\alpha\beta}}{\sigma^2}$$

iii. (5 分) 这题就是一个简单的计算，不需要任何一点积分，请看操作, (3 分)

$$\begin{aligned} g_{\alpha\beta}(\theta) &= - \left\langle \frac{\partial^2 \ln P(\mathbf{x}|\theta)}{\partial \theta_{\alpha} \partial \theta_{\beta}} \right\rangle_{\mathbf{x}} \\ &= \left\langle \frac{\partial^2}{\partial \theta_{\alpha} \partial \theta_{\beta}} \left(\frac{1}{2\sigma^2} \sum_i (y_i(\theta) - x_i)^2 \right) \right\rangle_{\mathbf{x}} \\ &= \frac{1}{2\sigma^2} \left\langle \frac{\partial^2 (\mathbf{y}(\theta) - \mathbf{x})^2}{\partial \theta_{\alpha} \partial \theta_{\beta}} \right\rangle_{\mathbf{x}} \\ &= \frac{1}{2\sigma^2} \left\langle \frac{\partial}{\partial \theta_{\alpha}} \left(2(\mathbf{y}(\theta) - \mathbf{x}) \cdot \frac{\partial \mathbf{y}}{\partial \theta_{\beta}} \right) \right\rangle_{\mathbf{x}} \\ &= \frac{1}{2\sigma^2} \left\langle 2(\mathbf{y}(\theta) - \mathbf{x}) \cdot \frac{\partial^2 \mathbf{y}}{\partial \theta_{\alpha} \partial \theta_{\beta}} + 2 \frac{\partial \mathbf{y}}{\partial \theta_{\alpha}} \cdot \frac{\partial \mathbf{y}}{\partial \theta_{\beta}} \right\rangle_{\mathbf{x}} \end{aligned}$$

可以看到式子前半部分，由高斯分布可以知道 $\langle \mathbf{x} \rangle_{\mathbf{x}} = \mathbf{y}(\boldsymbol{\theta})$ ，因此前半部分是 0。后半部分与 $\mathbf{x}$ 无关，直接当常数拎出来，所以有 (2 分)

$$g_{\alpha\beta}(\boldsymbol{\theta}) = \frac{1}{\sigma^2} \frac{\partial \mathbf{y}}{\partial \theta_{\alpha}} \cdot \frac{\partial \mathbf{y}}{\partial \theta_{\beta}} = \frac{1}{\sigma^2} \sum_i \frac{\partial y_i}{\partial \theta_{\alpha}} \frac{\partial y_i}{\partial \theta_{\beta}} = \frac{(J^T J)_{\alpha\beta}}{\sigma^2}$$

iv. (5 分) 根据题意，我们将 $\ln P(\mathbf{x}|\boldsymbol{\theta} + \epsilon \boldsymbol{\Delta})$ 关于 ϵ 展开到二阶，有 (2 分)

$$\ln P(\mathbf{x}|\boldsymbol{\theta} + \epsilon \boldsymbol{\Delta}) = \ln P(\mathbf{x}|\boldsymbol{\theta}) + \epsilon \sum_{\alpha} \frac{\partial \ln P}{\partial \theta_{\alpha}} \Delta_{\alpha} + \frac{\epsilon^2}{2} \sum_{\alpha, \beta} \frac{\partial^2 \ln P}{\partial \theta_{\alpha} \partial \theta_{\beta}} \Delta_{\alpha} \Delta_{\beta}$$

因此有将 $\ln P$ 移到左侧后同取 e 指数得到 (2 分)

$$\frac{P(\mathbf{x}|\boldsymbol{\theta} + \epsilon \boldsymbol{\Delta})}{P(\mathbf{x}|\boldsymbol{\theta})} = \exp \left(\epsilon \sum_{\alpha} \frac{\partial \ln P}{\partial \theta_{\alpha}} \Delta_{\alpha} + \frac{\epsilon^2}{2} \sum_{\alpha, \beta} \frac{\partial^2 \ln P}{\partial \theta_{\alpha} \partial \theta_{\beta}} \Delta_{\alpha} \Delta_{\beta} \right)$$

值得注意的是，当 $\epsilon = 0$ 时，上式取极值 1，因为由一次游戏的预测得到参数 $\boldsymbol{\theta}$ 的可能性最大，因为其原本就是遵循参数 $\boldsymbol{\theta}$ 的分布的。这意味着表达式中无一次项，代入 $g_{\alpha\beta}$ 表达式可以得到 (1 分)

$$\frac{P(\mathbf{x}|\boldsymbol{\theta} + \epsilon \boldsymbol{\Delta})}{P(\mathbf{x}|\boldsymbol{\theta})} = \exp \left(-\frac{\epsilon^2}{2} \sum_{\alpha, \beta} g_{\alpha\beta} \Delta_{\alpha} \Delta_{\beta} \right)$$

2. (62 分)

i. (9 分) 假骰子和真骰子分布之间的 Hellinger 距离为 (1 分)

$$d_{\text{Hel}} = \sqrt{\left(\left(\sqrt{\frac{1}{6}} - \sqrt{0.1} \right)^2 \times 5 + \left(\sqrt{\frac{1}{6}} - \sqrt{0.5} \right)^2 \right)} = 0.3628$$

对于两个临近分布，有 (3 分)

$$\begin{aligned} d_{\text{Hel}}^2(P(\mathbf{x}|\boldsymbol{\theta} + \epsilon \boldsymbol{\Delta}), P(\mathbf{x}|\boldsymbol{\theta})) &= \int \left(\sqrt{P(\mathbf{x}|\boldsymbol{\theta} + \epsilon \boldsymbol{\Delta})} - \sqrt{P(\mathbf{x}|\boldsymbol{\theta})} \right)^2 d\mathbf{x} \\ &= \int \left(\sum_{\alpha} \frac{\partial \sqrt{P}}{\partial \theta_{\alpha}} \epsilon \Delta_{\alpha} \right)^2 d\mathbf{x} \\ &= \epsilon^2 \int \left(\sum_{\alpha} \frac{\partial \sqrt{P}}{\partial \theta_{\alpha}} \Delta_{\alpha} \right)^2 d\mathbf{x} \\ &= \epsilon^2 \int \left(\sum_{\alpha} \frac{1}{2\sqrt{P}} \frac{\partial P}{\partial \theta_{\alpha}} \Delta_{\alpha} \right)^2 d\mathbf{x} \\ &= \epsilon^2 \int \left(\sum_{\alpha} \frac{\sqrt{P}}{2} \frac{\partial \ln P}{\partial \theta_{\alpha}} \Delta_{\alpha} \right)^2 d\mathbf{x} \\ &= \frac{\epsilon^2}{4} \int P \left(\sum_{\alpha} \frac{\partial \ln P}{\partial \theta_{\alpha}} \Delta_{\alpha} \right)^2 d\mathbf{x} \end{aligned}$$

可以注意到，积分内部就是对 $(\sum_{\alpha} \frac{\partial \ln P}{\partial \theta_{\alpha}} \Delta_{\alpha})^2$ 关于 $\mathbf{x}$ 取平均，也就是 (2 分)

$$\begin{aligned} d_{\text{Hel}}^2 &= \frac{\epsilon^2}{4} \left\langle \left(\sum_{\alpha} \frac{\partial \ln P}{\partial \theta_{\alpha}} \Delta_{\alpha} \right)^2 \right\rangle_{\mathbf{x}} \\ &= \frac{\epsilon^2}{4} \left\langle \sum_{\alpha} \frac{\partial \ln P}{\partial \theta_{\alpha}} \Delta_{\alpha} \sum_{\beta} \frac{\partial \ln P}{\partial \theta_{\beta}} \Delta_{\beta} \right\rangle_{\mathbf{x}} \\ &\stackrel{\text{根据平均值的线性性}}{=} \frac{\epsilon^2}{4} \sum_{\alpha} \sum_{\beta} \left\langle \frac{\partial \ln P}{\partial \theta_{\alpha}} \frac{\partial \ln P}{\partial \theta_{\beta}} \right\rangle_{\mathbf{x}} \Delta_{\alpha} \Delta_{\beta} \end{aligned}$$

因此可以得到 (1 分)

$$g_{\alpha\beta} = \frac{1}{4} \left\langle \frac{\partial \ln P}{\partial \theta_{\alpha}} \frac{\partial \ln P}{\partial \theta_{\beta}} \right\rangle_{\mathbf{x}}$$

值得注意的是由于 $P, Q \geq 0$ ，因此 (1 分)

$$P \cdot Q = \int \sqrt{P(\mathbf{x})} \sqrt{Q(\mathbf{x})} d\mathbf{x} \geq 0$$

所以 Hellinger 距离存在上限 (1 分)

$$d_{\text{Hel}}^2(P, Q) = 2 - 2P \cdot Q \leq 2 \implies d_{\text{Hel}}(P, Q) \leq \sqrt{2}$$

ii. (8 分) 根据每次实验的独立性，我们可以知道 (1 分)

$$P_N(\mathbf{j}) = P(j_1)P(j_2) \cdots P(j_N) = \prod_{i=1}^N P(j_i)$$

于是有 (3 分)

$$\begin{aligned} P_N \cdot Q_N &= \sum_{\mathbf{j}} \sqrt{P(\mathbf{j})Q(\mathbf{j})} = \sum_{j_1, j_2, \dots, j_N} \sqrt{P(j_1)Q(j_1)P(j_2)Q(j_2) \cdots P(j_N)Q(j_N)} \\ &= \sum_{j_1} \sqrt{P(j_1)Q(j_1)} \sum_{j_2} \sqrt{P(j_2)Q(j_2)} \cdots \sum_{j_N} \sqrt{P(j_N)Q(j_N)} = (P \cdot Q)^N \end{aligned}$$

为了求出 $d_{\text{Hel}}(P_{100}, Q_{100})$ 可以先计算 (1 分)

$$P \cdot Q = 5\sqrt{\frac{1}{6}}\sqrt{0.1} + \sqrt{\frac{1}{6}}\sqrt{0.5} = 0.9341$$

所以 (1 分)

$$d_{\text{Hel}}(P_{100}, Q_{100}) = \sqrt{2}\sqrt{1 - (P \cdot Q)^{100}} = 0.9989\sqrt{2} = 1.4126$$

可以看到，随着重复次数的增加，Hellinger 距离越来越接近 $\sqrt{2}$ 。这意味着留给赌场“判罚”的区间是很小的，可能就是从比 $\sqrt{2}$ 小一点点到 $\sqrt{2}$ 的区间内认为是作弊。这导致赌场很有可能抓不到真正作弊的赌徒，还有可能抓到没有作弊的人。 (2 分)

iii. (2 分) 利用 $P_N \cdot Q_N = (P \cdot Q)^N$ 可以得到 (2 分)

$$d_{\text{Bhatt}}^2(P, Q) = \lim_{N \rightarrow 0} \frac{1 - (P \cdot Q)^N}{N} = \lim_{N \rightarrow 0} \frac{1 - 1 - N \ln(P \cdot Q)}{N} = -\ln(P \cdot Q)$$

iv. (9 分) 我们可以使用简单的不等式完成证明 (2 分)

$$d_{\text{KL}}(Q|P) = - \int P(\mathbf{x}) \ln \left(\frac{Q(\mathbf{x})}{P(\mathbf{x})} \right) d\mathbf{x} \geq - \int P(\mathbf{x}) \left(\frac{Q(\mathbf{x})}{P(\mathbf{x})} - 1 \right) d\mathbf{x} = \int (P(\mathbf{x}) - Q(\mathbf{x})) d\mathbf{x} = 1 - 1 = 0$$

不等号取等当且仅当 (1 分)

$$Q(\mathbf{x})/P(\mathbf{x}) = 1 \implies P(\mathbf{x}) = Q(\mathbf{x})$$

考察临近分布 $P(\mathbf{x}|\boldsymbol{\theta})$ 与 $P(\mathbf{x}|\boldsymbol{\theta} + \epsilon\boldsymbol{\Delta})$, 两者之间的距离为 (3 分)

$$\begin{aligned} d_{\text{KL}}(P(\mathbf{x}|\boldsymbol{\theta} + \epsilon\boldsymbol{\Delta})|P(\mathbf{x}|\boldsymbol{\theta})) &= - \int P(\mathbf{x}|\boldsymbol{\theta}) \ln \left(\frac{P(\mathbf{x}|\boldsymbol{\theta} + \epsilon\boldsymbol{\Delta})}{P(\mathbf{x}|\boldsymbol{\theta})} \right) d\mathbf{x} \\ &= - \int P \ln \left(\frac{P + \sum_{\alpha} \frac{\partial P}{\partial \theta_{\alpha}} \epsilon \Delta_{\alpha}}{P} \right) d\mathbf{x} \\ &= - \int P \ln \left(1 + \sum_{\alpha} \frac{\partial \ln P}{\partial \theta_{\alpha}} \epsilon \Delta_{\alpha} \right) d\mathbf{x} \\ &= -\epsilon \int P \sum_{\alpha} \frac{\partial \ln P}{\partial \theta_{\alpha}} \Delta_{\alpha} d\mathbf{x} \\ &= -\epsilon \sum_{\alpha} \left(\int P \frac{\partial \ln P}{\partial \theta_{\alpha}} d\mathbf{x} \right) \Delta_{\alpha} \\ &= -\epsilon \sum_{\alpha} \left\langle \frac{\partial \ln P}{\partial \theta_{\alpha}} \right\rangle_{\mathbf{x}} \Delta_{\alpha} \end{aligned}$$

所以我们可以将其平方得到 (2 分)

$$d_{\text{KL}}^2 = \epsilon^2 \sum_{\alpha, \beta} \left\langle \frac{\partial \ln P}{\partial \theta_{\alpha}} \right\rangle_{\mathbf{x}} \left\langle \frac{\partial \ln P}{\partial \theta_{\beta}} \right\rangle_{\mathbf{x}} \Delta_{\alpha} \Delta_{\beta}$$

所以有 (1 分)

$$g_{\alpha\beta} = \left\langle \frac{\partial \ln P}{\partial \theta_{\alpha}} \right\rangle_{\mathbf{x}} \left\langle \frac{\partial \ln P}{\partial \theta_{\beta}} \right\rangle_{\mathbf{x}}$$

v. (34 分) 对于两者的广延性的证明, 比较直接, (1 分)

$$d_{\text{Bhatt}}^2(P_N, Q_N) = -\ln(P_N \cdot Q_N) = -\ln((P \cdot Q)^N) = -N \ln(P \cdot Q) = N d_{\text{Bhatt}}^2(P, Q)$$

而对于连续分布 $P_N(\mathbf{x})$, 是一个建立在 $\mathbf{x} = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N)$ 上的概率分布, 可以写开来 $P_N(\mathbf{x}) = P(\mathbf{x}_1)P(\mathbf{x}_2) \cdots P(\mathbf{x}_N)$,

放到积分里，我们的 $d\mathbf{x}$ 可以展开为 $d\mathbf{x}_1 d\mathbf{x}_2 \cdots d\mathbf{x}_N$ 。(4 分)

$$\begin{aligned}
 d_{\text{KL}}(Q_N|P_N) &= - \int P(\mathbf{x}_1)P(\mathbf{x}_2) \cdots P(\mathbf{x}_N) \ln \left(\frac{Q(\mathbf{x}_1)Q(\mathbf{x}_2) \cdots Q(\mathbf{x}_N)}{P(\mathbf{x}_1)P(\mathbf{x}_2) \cdots P(\mathbf{x}_N)} \right) d\mathbf{x}_1 d\mathbf{x}_2 \cdots d\mathbf{x}_N \\
 &= - \sum_{i=1}^N \int P(\mathbf{x}_1)P(\mathbf{x}_2) \cdots P(\mathbf{x}_N) \ln \left(\frac{Q(\mathbf{x}_i)}{P(\mathbf{x}_i)} \right) d\mathbf{x}_1 d\mathbf{x}_2 \cdots d\mathbf{x}_N \\
 &= - \sum_{i=1}^N \underbrace{\int P(\mathbf{x}_1) d\mathbf{x}_1 \int P(\mathbf{x}_2) d\mathbf{x}_2 \cdots \int P(\mathbf{x}_N) d\mathbf{x}_N}_1 \cdot \int P(\mathbf{x}_i) \ln \left(\frac{Q(\mathbf{x}_i)}{P(\mathbf{x}_i)} \right) d\mathbf{x}_i \\
 &= - \sum_{i=1}^N \int P(\mathbf{x}_i) \ln \left(\frac{Q(\mathbf{x}_i)}{P(\mathbf{x}_i)} \right) d\mathbf{x}_i = -N \int P(\mathbf{x}) \ln \left(\frac{Q(\mathbf{x})}{P(\mathbf{x})} \right) d\mathbf{x} \\
 &= Nd_{\text{KL}}(Q|P) \\
 d_{\text{sKL}}(Q_N, P_N) &= \frac{1}{2}(d_{\text{KL}}(Q_N|P_N) + d_{\text{KL}}(P_N|Q_N)) = \frac{N}{2}(d_{\text{KL}}(Q|P) + d_{\text{KL}}(P|Q)) = Nd_{\text{sKL}}(Q, P)
 \end{aligned}$$

由此，我们证明了广延性。

对于三角不等式的证伪，我们可以举一个反例：比如我们有三个不同的硬币，其抛出正反面的概率如下 (1 分)

$$P_i = \{0.01, 0.99\}, \quad Q_i = \{0.5, 0.5\}, \quad R_i = \{0.99, 0.01\}$$

我们可以分别计算三者之间的距离，(3 分)

$$\begin{aligned}
 d_{\text{Bhatt}}(P, Q) &= d_{\text{Bhatt}}(Q, R) = 0.5057, \quad d_{\text{Bhatt}}(P, R) = 1.2706 \\
 d_{\text{Bhatt}}(P, Q) + d_{\text{Bhatt}}(Q, R) &\leq d_{\text{Bhatt}}(P, R) \\
 d_{\text{sKL}}(Q, P) &= d_{\text{sKL}}(R, Q) = 1.1258, \quad d_{\text{sKL}}(R, P) = 4.5032 \\
 d_{\text{sKL}}(Q, P) + d_{\text{sKL}}(R, Q) &\leq d_{\text{sKL}}(R, P)
 \end{aligned}$$

可以看到这是一个三角不等式的反例。

对于正定性，我们已经完成了对 Kullback-Liebler 散度的证明，我们需要对 Bhattacharyya 距离进行证明。我们需要用到柯西-施瓦茨不等式 (2 分)

$$\left(\int \sqrt{P(\mathbf{x})Q(\mathbf{x})} d\mathbf{x} \right)^2 \leq \int \sqrt{P(\mathbf{x})}^2 d\mathbf{x} \int \sqrt{Q(\mathbf{x})}^2 d\mathbf{x} = 1$$

因此我们可以得到 (1 分)

$$P \cdot Q \leq 1$$

因此可以得到 (1 分)

$$-\ln(P \cdot Q) \geq 0$$

且柯西-施瓦茨不等式在 $P(\mathbf{x}) = cQ(\mathbf{x})$ 时取等，由 $\int P(\mathbf{x}) d\mathbf{x} = \int Q(\mathbf{x}) d\mathbf{x} = 1$ ，我们可以发现 c 一定得取 1，因此上述不等式的取等条件为 (1 分)

$$P(\mathbf{x}) = Q(\mathbf{x})$$

由此，我们证明了两者的正定性。

对于**对称性**，我们能够轻松证明 **(2 分)**

$$d_{\text{Bhatt}}^2(P, Q) = -\ln \left(\int \sqrt{P(\mathbf{x})Q(\mathbf{x})} d\mathbf{x} \right) \xrightarrow{\text{乘法交换律}} -\ln \left(\int \sqrt{Q(\mathbf{x})P(\mathbf{x})} d\mathbf{x} \right) = d_{\text{Bhatt}}^2(Q, P)$$

$$d_{\text{sKL}}(Q, P) = \frac{1}{2}(d_{\text{KL}}(Q|P) + d_{\text{KL}}(P|Q)) \xrightarrow{\text{加法交换律}} \frac{1}{2}(d_{\text{KL}}(P|Q) + d_{\text{KL}}(Q|P)) = d_{\text{sKL}}(P, Q)$$

由此，我们证明了两者的对称性。

对于高斯分布 $P(\mathbf{x}|\boldsymbol{\theta})$ 与 $P(\mathbf{x}|\boldsymbol{\phi})$ ，我们需要通过积分得到最后的结果。有 **(2 分)**

$$P(\mathbf{x}|\boldsymbol{\theta}) \cdot P(\mathbf{x}|\boldsymbol{\phi}) = \int \frac{1}{(2\pi\sigma^2)^{N/2}} \sqrt{e^{-|\mathbf{y}(\boldsymbol{\theta})-\mathbf{x}|^2/(2\sigma^2)} e^{-|\mathbf{y}(\boldsymbol{\phi})-\mathbf{x}|^2/(2\sigma^2)}} d\mathbf{x}$$

$$= \int \frac{e^{-(|\mathbf{y}(\boldsymbol{\theta})-\mathbf{x}|^2 + |\mathbf{y}(\boldsymbol{\phi})-\mathbf{x}|^2)/(4\sigma^2)}}{(2\pi\sigma^2)^{N/2}} d\mathbf{x}$$

现在做一个巧思换元 $\mathbf{x} = \mathbf{x}' + \frac{\mathbf{y}(\boldsymbol{\theta}) + \mathbf{y}(\boldsymbol{\phi})}{2}$ ，有 $d\mathbf{x} = d\mathbf{x}'$ ，代入可以得到 **(2 分)**

$$P(\mathbf{x}|\boldsymbol{\theta}) \cdot P(\mathbf{x}|\boldsymbol{\phi}) = \int \frac{e^{-\left(\left|\frac{\mathbf{y}(\boldsymbol{\theta})-\mathbf{y}(\boldsymbol{\phi})}{2}-\mathbf{x}'\right|^2 + \left|\frac{\mathbf{y}(\boldsymbol{\phi})-\mathbf{y}(\boldsymbol{\theta})}{2}-\mathbf{x}'\right|^2\right)/(4\sigma^2)}}{(2\pi\sigma^2)^{N/2}} d\mathbf{x}'$$

$$= \int \frac{e^{-\left(\left(\frac{\mathbf{y}(\boldsymbol{\theta})-\mathbf{y}(\boldsymbol{\phi})}{2}\right)^2 + \mathbf{x}'^2\right)/(2\sigma^2)}}{(2\pi\sigma^2)^{N/2}} d\mathbf{x}'$$

$$= e^{-\left(\frac{\mathbf{y}(\boldsymbol{\theta})-\mathbf{y}(\boldsymbol{\phi})}{2}\right)^2/(2\sigma^2)} \int \frac{e^{-\mathbf{x}'^2/(2\sigma^2)}}{(2\pi\sigma^2)^{N/2}} d\mathbf{x}'$$

注意到积分号里面的就是一个 N 维高斯分布，积分结果应当是 1，所以可以得到 **(1 分)**

$$P(\mathbf{x}|\boldsymbol{\theta}) \cdot P(\mathbf{x}|\boldsymbol{\phi}) = e^{-\left(\frac{\mathbf{y}(\boldsymbol{\theta})-\mathbf{y}(\boldsymbol{\phi})}{2}\right)^2/(2\sigma^2)}$$

因此两者的 Bhattacharyya 距离为 **(1 分)**

$$d_{\text{Bhatt}}^2(P(\mathbf{x}|\boldsymbol{\theta}), P(\mathbf{x}|\boldsymbol{\phi})) = -\ln(P(\mathbf{x}|\boldsymbol{\theta}) \cdot P(\mathbf{x}|\boldsymbol{\phi})) = \frac{1}{8\sigma^2} |\mathbf{y}(\boldsymbol{\theta}) - \mathbf{y}(\boldsymbol{\phi})|^2$$

而两个高斯分布的对称 Kullback-Liebler 散度为 **(2 分)**

$$d_{\text{sKL}}(P(\mathbf{x}|\boldsymbol{\phi}), P(\mathbf{x}|\boldsymbol{\theta})) = \frac{1}{2} \int \frac{e^{-|\mathbf{y}(\boldsymbol{\theta})-\mathbf{x}|^2/(2\sigma^2)} - e^{-|\mathbf{y}(\boldsymbol{\phi})-\mathbf{x}|^2/(2\sigma^2)}}{(2\pi\sigma^2)^{N/2}} \cdot \frac{|\mathbf{y}(\boldsymbol{\phi}) - \mathbf{x}|^2 - |\mathbf{y}(\boldsymbol{\theta}) - \mathbf{x}|^2}{2\sigma^2} d\mathbf{x}$$

为了计算这个积分，我们需要将问题拆成四个小积分 **(2 分)**

$$I_1 = \int \frac{e^{-|\mathbf{y}(\boldsymbol{\theta})-\mathbf{x}|^2/(2\sigma^2)}}{(2\pi\sigma^2)^{N/2}} |\mathbf{y}(\boldsymbol{\theta}) - \mathbf{x}|^2 d\mathbf{x}, I_2 = \int \frac{e^{-|\mathbf{y}(\boldsymbol{\phi})-\mathbf{x}|^2/(2\sigma^2)}}{(2\pi\sigma^2)^{N/2}} |\mathbf{y}(\boldsymbol{\theta}) - \mathbf{x}|^2 d\mathbf{x}$$

$$I_3 = \int \frac{e^{-|\mathbf{y}(\boldsymbol{\theta})-\mathbf{x}|^2/(2\sigma^2)}}{(2\pi\sigma^2)^{N/2}} |\mathbf{y}(\boldsymbol{\phi}) - \mathbf{x}|^2 d\mathbf{x}, I_4 = \int \frac{e^{-|\mathbf{y}(\boldsymbol{\phi})-\mathbf{x}|^2/(2\sigma^2)}}{(2\pi\sigma^2)^{N/2}} |\mathbf{y}(\boldsymbol{\phi}) - \mathbf{x}|^2 d\mathbf{x}$$

我们先来计算 I_1 与 I_4 ，可以注意到 (2 分)

$$\begin{aligned} I_1 &= \int \frac{e^{-|\mathbf{y}(\boldsymbol{\theta})-\mathbf{x}|^2/(2\sigma^2)}}{(2\pi\sigma^2)^{N/2}} \sum_{i=1}^N (y_i(\boldsymbol{\theta}) - x_i)^2 d\mathbf{x} \\ &= \left\langle \sum_{i=1}^N (y_i(\boldsymbol{\theta}) - x_i)^2 \right\rangle_{P(\mathbf{x}|\boldsymbol{\theta})} \\ &= \sum_{i=1}^N \langle (y_i(\boldsymbol{\theta}) - x_i)^2 \rangle_{P(\mathbf{x}|\boldsymbol{\theta})} \\ &= N\sigma^2 \end{aligned}$$

这里最后一步的操作是因为我们计算的实际上就是分布在 $\hat{x}_i$ 方向上的方差，这正是高斯分布中的 σ^2 ，这在题干中也有提到。同理我们可以得到 (1 分)

$$I_4 = N\sigma^2$$

对于 I_2 我们可以使用一个巧妙转换变成 I_4 ，具体如下 (2 分)

$$\begin{aligned} I_2 &= \int \frac{e^{-|\mathbf{y}(\boldsymbol{\phi})-\mathbf{x}|^2/(2\sigma^2)}}{(2\pi\sigma^2)^{N/2}} \sum_{i=1}^N (y_i(\boldsymbol{\theta}) - x_i)^2 d\mathbf{x} \\ &= \sum_{i=1}^N \langle (y_i(\boldsymbol{\theta}) - x_i)^2 \rangle_{P(\mathbf{x}|\boldsymbol{\phi})} \\ &= \sum_{i=1}^N \langle (y_i(\boldsymbol{\theta}) - y_i(\boldsymbol{\phi}) + y_i(\boldsymbol{\phi}) - x_i)^2 \rangle_{P(\mathbf{x}|\boldsymbol{\phi})} \\ &= \sum_{i=1}^N \langle (y_i(\boldsymbol{\theta}) - y_i(\boldsymbol{\phi}))^2 + 2(y_i(\boldsymbol{\theta}) - y_i(\boldsymbol{\phi}))(y_i(\boldsymbol{\phi}) - x_i) + (y_i(\boldsymbol{\phi}) - x_i)^2 \rangle_{P(\mathbf{x}|\boldsymbol{\phi})} \\ &= \sum_{i=1}^N (y_i(\boldsymbol{\theta}) - y_i(\boldsymbol{\phi}))^2 + \sum_{i=1}^N 2(y_i(\boldsymbol{\theta}) - y_i(\boldsymbol{\phi})) \underbrace{\langle y_i(\boldsymbol{\phi}) - x_i \rangle_{P(\mathbf{x}|\boldsymbol{\phi})}}_0 + \sum_{i=1}^N \langle (y_i(\boldsymbol{\phi}) - x_i)^2 \rangle_{P(\mathbf{x}|\boldsymbol{\phi})} \\ &= |\mathbf{y}(\boldsymbol{\theta}) - \mathbf{y}(\boldsymbol{\phi})|^2 + N\sigma^2 \end{aligned}$$

同理，我们可以得到 (1 分)

$$I_3 = |\mathbf{y}(\boldsymbol{\theta}) - \mathbf{y}(\boldsymbol{\phi})|^2 + N\sigma^2$$

所以我们可以计算积分 (2 分)

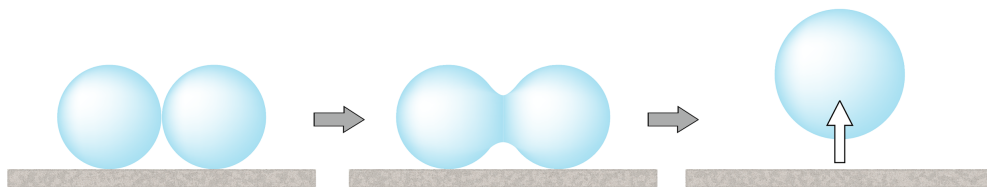
$$d_{\text{KL}}(P(\mathbf{x}|\boldsymbol{\phi}), P(\mathbf{x}|\boldsymbol{\theta})) = \frac{1}{4\sigma^2} (I_2 + I_3 - I_1 - I_4) = \frac{1}{2\sigma^2} |\mathbf{y}(\boldsymbol{\theta}) - \mathbf{y}(\boldsymbol{\phi})|^2$$

七. (40 分) 表面张力系数的有趣问题

1. 在平面上有两个半径为 $a = 100 \mu\text{m}$ 的静止水滴，现在两个水滴突然聚合形成一个大水珠并腾空跳起。在这一过程中，水滴表面张力的能量变化有 k 部分转化为水滴的质心动能，试着计算水滴跳起时的速度 v 。

- 水滴的密度 $\rho = 1.00 \times 10^3 \text{ kg/m}^3$
- 水滴的表面张力系数为 $\gamma = 7.27 \times 10^{-2} \text{ J/m}^2$

- $k = 0.06$
- 在融合前后，水的总体积不变。

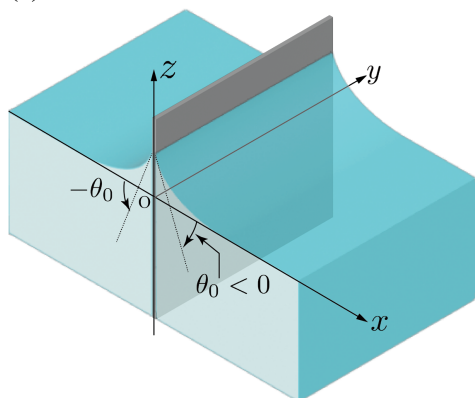


2. 将一个竖直放置的平板插入水中。两图分别展示水面在亲水与疏水平板周围形成的曲面。本题中我们忽略平板的厚度。

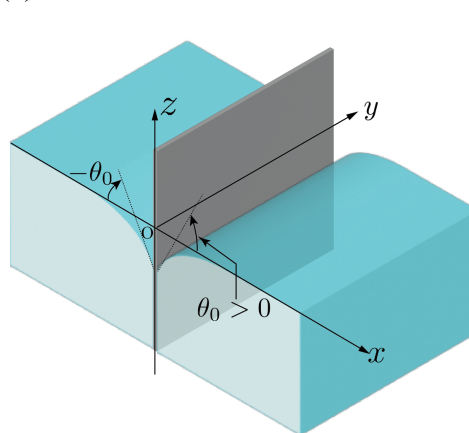
我们不妨建立如下描述水面的坐标系。竖直放置的平板即为 yz 平面，而远离平板的水面位于 xy 平面上，其 $z = 0$ 。水面的形状与 y 坐标无关。设 $\theta(x)$ 为横坐标 x 处水面切平面与 xy 平面的夹角，这里我们以 x 轴正方向为起始方向，逆时针旋转到切平面上为正方向。（反方向则为负）设 θ_0 为水面与平板接触的地方的切平面角度，这一角度只与平板材质有关。

本题中，水的密度为 ρ 而表面张力系数为 γ 。重力加速度常数为 g ，大气压强均匀且为 P_0 。

(a)

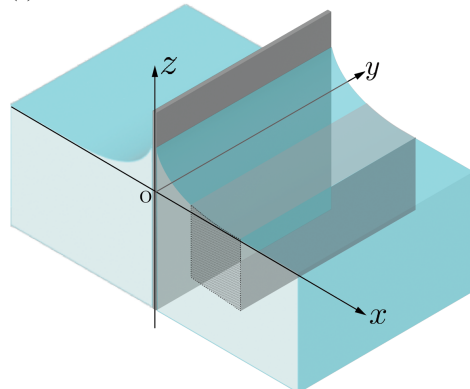


(b)

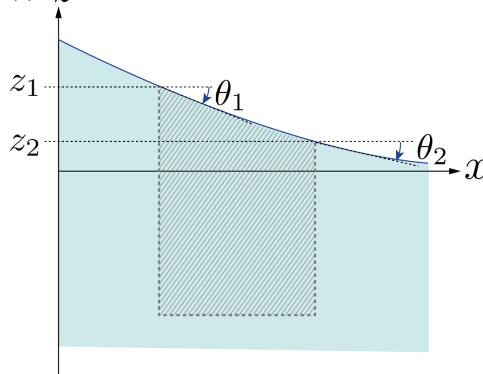


- 为了简单起见，我们不妨先考察亲水平板的情况。我们注意到，水中的压强 P 在 $z > 0$ 时，满足 $P < P_0$ 。而在 $z = 0$ 时满足 $P = P_0$ 。利用 z, ρ, g, P_0 给出 P 在 $z > 0$ 中与 z 的关系式。
- 现在，我们考察液体中的一个水块，由下图阴影部分所示。 z_1, z_2 分别是水块左右边界水面的高度。由于水的内部与外界大气有压强差，这会水块在水平方向上产生压强差力。压强差对水块在 y 方向单位长度施加的水平（ x 方向）作用力为 f_x 。使用 ρ, g, z_1, z_2 表示 f_x 。注意到 P_0 会在最后的表达式中消去。

(a)



(b)



iii. 表面张力与前一小问的压差力之间相互平衡，使得水块最后静止在原地。列出受力平衡方程，重新用 $\gamma, \theta_1, \theta_2$ 表示 f_x 。这里 θ_1, θ_2 为左右边界处水面的倾角，如图所示。

iv. 由此，我们可以得到如下等式：对于水表面处的任意一点 (x, z) ，有

$$\frac{1}{2} \left(\frac{z}{\ell} \right)^a + \cos \theta(x) = \text{常数}$$

确定指数 a 的值，并且用 γ, g, ρ 表示 ℓ 。

v. 现在，我们假设水面高度的变化很小，即 $|z'(x)| \ll 1$ ，由此我们能将 $\cos \theta(x)$ 展开到关于 $z'(x)$ 的二阶表达式。然后两边关于 x 求导得到关于 $z(x)$ 的一个二阶常微分方程。对 $x \geq 0$ 的部分求解微分方程，并且用 $\ell, \tan \theta_0$ 表示。

[解答]

1. (12 分)

融合前，每一个液滴的表面张力能为 (2 分)

$$E_\gamma = 4\pi a^2 \gamma$$

而融合后，由于体积不变，液滴的半径变为 (2 分)

$$a' = 2^{1/3} a$$

所以融合后的总表面张力能为 (2 分)

$$E'_\gamma = 2^{2/3} \cdot 4\pi a^2 \gamma$$

故有表面张力能减少了 (2 分)

$$\Delta E = 2E_\gamma - E'_\gamma = (2 - 2^{2/3}) \cdot 4\pi a^2 \gamma$$

融合后液体的动能为 (2 分)

$$E_k = k \Delta E = k (2 - 2^{2/3}) \cdot 4\pi a^2 \gamma$$

所以液体最后的跳起速度为 (2 分)

$$v = \sqrt{\frac{2E_k}{m}} = \sqrt{\frac{2k(2 - 2^{2/3}) \cdot 4\pi a^2 \gamma}{\frac{8}{3}\pi a^3 \rho}} \approx 0.23 \text{ m/s}$$

2. (28 分)

i. (2 分) 根据流体静力学知识，有 (2 分)

$$P = P_0 - \rho g z$$

ii. (2 分) 流体受到的压差力来源于高度 $z_2 < z < z_1$ 的区间内，水内部对水块的压强 P 小于外界大气压 P_0 ，产生水平向左的压差力。所以单位长度的水平压差力为 (2 分)

$$f_x = \int_{z_2}^{z_1} (P_0 - \rho g z) dz - \int_{z_2}^{z_1} P_0 dz = -\frac{1}{2} \rho g (z_1^2 - z_2^2)$$

iii. (4 分) 水块受到的表面张力的水平分量为 $\gamma \cos \theta_2 - \gamma \cos \theta_1$ ，因此可以列出受力平衡方程 (2 分)

$$f_x + \gamma \cos \theta_2 - \gamma \cos \theta_1 = 0$$

所以 (2 分)

$$f_x = \gamma(\cos \theta_1 - \cos \theta_2)$$

iv. (5 分) 由 (ii) 与 (iii) 的结果，我们可以得到 (2 分)

$$\frac{1}{2}\rho g z_1^2 + \gamma \cos \theta_1 = \frac{1}{2}\rho g z_2^2 + \gamma \cos \theta_2$$

由于上述表达式对任意表面处的点 (x, z) 都成立，我们可以得知 (2 分)

$$\frac{1}{2}\rho g z^2 + \gamma \cos \theta = \text{常量}$$

所以变形并对比可得 (2 分)

$$a = 2, \ell = \sqrt{\frac{\gamma}{\rho g}}$$

v. (14 分) 使用导数表示三角函数，有 (2 分)

$$\cos \theta(x) = \frac{1}{\sqrt{1 + (z')^2}}$$

当 $|z'(x)| \ll 1$ 时，有 (2 分)

$$\cos \theta(x) = 1 - \frac{(z')^2}{2}$$

代入原方程可以得到 (2 分)

$$\frac{1}{2}\left(\frac{z}{\ell}\right)^2 - \frac{1}{2}(z')^2 = \text{常数}$$

两边关于 x 求导可以得到 (2 分)

$$z' \cdot \frac{z}{\ell^2} - z'' \cdot z' = 0$$

$$z'' = \frac{z}{\ell^2}$$

方程的通解为 (2 分)

$$z = Ae^{x/\ell} + Be^{-x/\ell}$$

当 $x \rightarrow \infty$ 时， $z \rightarrow 0$ ，所以 (1 分)

$$A = 0$$

而当 $x = 0$ 时， $z' = \tan \theta_0$ 代入得到 (1 分)

$$B = -\ell \tan \theta_0$$

故 (2 分)

$$z(x) = -\ell \tan \theta_0 e^{-x/\ell}$$