通过 MBAR 重加权合并多个界面集合路径系综

Rik S. Breebaart[1] 和 Peter G. Bolhuis*[1]

Van 't Hoff Institute for Molecular Sciences, Universiteit van Amsterdam, Science Park 904, 1098 XH Amsterdam,The Netherlands

我们提出了一种方法，通过合并以不同集体变量为条件的跃迁界面采样模拟，计算重加权路径系综。该方法基于将多态 Bennett 接受率（MBAR）方法应用于完整轨迹。我们通过简单的二维势能模型和较复杂的主客体系统展示了这一技术，并表明与直接合并相比，该方法能显著改善统计效果。

arXiv:2601.00458v1 [cond-mat.stat-mech] 1 Jan 2026

I. 引言

跃迁路径采样（TPS）是一种通用的模拟技术，可用于研究罕见的复杂分子过程，例如蛋白质构象变化、成核或化学反应。该技术最初由 Dellago 等人在一篇开创性论文中提出[1]。与沿集体变量引入偏置力、但以扰乱自然动力学为代价的增强采样方法不同[2]，，TPS 通过在轨迹空间中执行蒙特卡罗采样，收集连接初态与终态的无偏动力学路径集合 [1,3–9]。给定一条初始轨迹后，该算法会选取一个帧，并通过向前和向后积分运动方程，从该帧发射一条新路径，以生成新的试验路径。随后，根据满足细致平衡的 Metropolis 接受准则，以一定概率接受该试验路径。由此得到的跃迁路径系综能够为罕见跃迁机制提供无偏见解。TPS 还可通过计算跃迁路径系综与另一种轨迹系综之间的可逆功，获得罕见事件的速率常数；后一种系综中的轨迹均从初态出发，但终点不受限制。

尽管 TPS 已取得成功，过渡界面采样（TIS）[10] 的发展显著提高了效率。这种基于 TPS 的技术在稳定态之间定义了一组超曲面，即界面；轨迹必须穿越这些界面，才能成为该系综的一部分[6,11,12]。TIS 的主要目标是计算穿越概率，即一条直接从初始状态 A 出发并穿越某个界面的轨迹，也穿越下一个界面的概率。最后一个界面则给出跃迁至最终状态 B 的穿越概率。速率常数可由所有界面的穿越概率连乘，再乘以通过第一个界面的通量得到。这种分层方法将指数级问题转化为线性问题[13,14]。每个界面都由一个序参量或集体变量（CV）参数化，通常记作 λ。原则上，TIS 得到的速率常数是精确的，且独立于序参量的选择，不过这一选择确实会影响采样统计。引入副本交换[15–17]可以进一步改善统计效果，增强采样路径之间的去相关性，尤其是那些接近稳定态的路径。TIS 只需计算到下一个界面的穿越概率，而穿越概率可以针对任意界面计算；事实上，对于 λ 的任意取值也都可以计算。

文献 18 提出了重加权路径系综（RPE）这一概念框架，指出：正如构型样本可以通过重加权来校正增强采样的影响（例如采用伞形采样[19]或元动力学[20],），也可以对整个 TIS 轨迹进行重加权，以校正其属于选择偏置系综这一事实（注意，动力学仍然是无偏的）。重加权因子可通过对条件穿越概率应用加权直方图分析（WHAM）[21,22] 来获得。一个关键见解是，每条轨迹的权重由该轨迹达到的最大 λ 决定[18]。因此，利用 TIS 得到的穿越概率信息，可以在整体无偏路径系综中正确表示采样到的轨迹。其优势在于，与无偏路径系综（通过直接 MD 获得）相比，RPE 能提供更多关于能垒/过渡态附近路径的信息（在 MD 工作量相同的情况下）。具体而言，RPE 能够准确估计投影承诺概率[23]。重加权路径系综已在多项应用中得到常规计算，尤其是晶体成核[24–27]、膜渗透[28],、蛋白质折叠[29,30] 和缔合[31,32],，以及新型算法[33,34]。近期，随着分子机制发现人工智能（AIMMD）[35]的出现，人们重新燃起了对 RPE 的兴趣。具体而言，Lazzeri 等人将 AIMMD 扩展为可重建整个 RPE[36]。

通过加权直方图分析方法（WHAM）得到的 RPE 估计值可以通过采样更多界面来改进。此外，副本交换过渡界面采样有助于改善统计质量和收敛性[14–16]。然而，RPE 方法的一个局限在于，利用 WHAM 获取总穿越概率的过程取决于所选的集体变量（CV） λ。如果出于某种原因，界面 CV 不是最优的，需要进行更改（例如纳入更多自由度），那么要将这些信息与先前 CV 已有的路径集合结合起来并不容易。事实上，必须从头开始，重新计算 RPE。

在本研究中，我们展示了可以结合基于不同 CV 的多个路径系综，从而得到一个通用的类加权直方图分析方法重加权方程。虽然也可以采用原始方法推导这一加权直方图分析方法方程，但我们发现使用多

状态 Bennett 接受比率（MBAR）方法[37]更清晰。过去，MBAR 和加权直方图分析方法都曾用于重加权[22,37]。两者在视角上的主要区别是，加权直方图分析方法基于统计误差或方差的最小化[21,22]，而 MBAR 则可以通过似然最大化方法推导得出[37,38]。不过，两种方法得出的方程非常相似，甚至可能完全相同。我们通过一个简单的玩具模型，以及一个更复杂的主客体结合跃迁[34]，展示了这一新方法。

本文其余部分的组织如下。首先，我们介绍如何从使用单个集体变量界面函数 λ 构建的一组界面系综中获取重加权路径系综的 MBAR 重加权程序。随后，我们将该方法扩展到由两个不同的集体变量函数 λ 和 μ 构建的 TIS 系综。最后，我们将该方法推广到任意数量 M 的集体变量函数。我们在一个简单的双阱势上验证了所提出的框架，结果表明，与使用单个集体变量应用 TIS 或直接合并分别构建的两个重加权路径系综相比，该框架具有优势。

II. 理论

我们首先在似然框架下处理 TIS 轨迹数据，并展示重加权路径系综如何从有偏的 TIS 路径系综上的最大似然（MBAR）估计量[37–39] 中产生。具体而言，我们考虑单个界面集合 $\{\lambda_k\}$，并使用 TIS 生成条件于穿过 λ[10] 中每个界面 λ_k 的轨迹系综。该条件由指示函数 $\hat{h}^k_{A,\lambda}[\mathbf{x}]$, 编码：当路径 $\mathbf{x}$ 离开 A态并穿过界面 λ_k 时，该函数取值为 1；之后路径可以返回 A态，也可以继续前往状态 B。在这里，一条轨迹 $\mathbf{x} = \{x_0, x_1 \dots x_L\}$ 由 $L+1$ 个帧 x 组成，帧之间以时间间隔 $\Delta\tau$分隔，而每个帧又包含系统中所有粒子的位置和动量。

我们的目标是估计路径权重 $w[\mathbf{x}]$ 以及一组配分和 $\{Z^k_{A.\lambda}\}$，使得 (i) 能从有偏样本的并集中恢复离开 A 的无偏路径系综，$\mathcal{P}_A[\mathbf{x}]$, ；(ii) 能得到无偏观测量，例如穿越概率 $P_A(\lambda|\lambda_1)$，其中 λ_1 为第一个界面。我们通过最大化在条件系综下观测到这些采样轨迹的似然来获得这些因子，其中这些条件系综以其所依据的无偏完整路径系综为条件，从而得到熟知的、针对路径空间的 MBAR 不动点方程。这种单界面集合的表述确立了符号约定，阐明了假设，并揭示了后文所用的关键直觉：一条轨迹的权重仅取决于它穿越的最高界面。在接下来的章节中，我们将同样的似然构造扩展到由不同集体变量构成的两个界面集合，然后扩展到 M 个集合，从而得到统一的 MultiSet-MBAR 程序，用于组合以不同集体变量作为界面生成的轨迹。

A. 基于 MBAR 的轨迹 RPE

考虑一组 K 轨迹系综，对应界面集合 $\{\lambda_k\}$，每个系综包含 N_k 个样本，其中 $k = 1$、.....、 K。基于这组采样路径，观测到一条路径$\mathbf{x}$ 的无偏概率 $\mathcal{P}_A[\mathbf{x}]$ 表示为

$$\mathcal{P}_A[\mathbf{x}] = \frac{1}{\mathcal{Z}} \sum_k^K \sum_n^{N_k} w_{k,n} \delta[\mathbf{x} - \mathbf{x}_{k,n}]. \tag{1}$$

这里，索引 k 遍历 K 个 TIS 模拟，而 n 遍历每次界面模拟中的 N_k 个样本。因子 $\mathcal{Z} = \sum_k \sum_n w_{k,n}$ 用于归一化分布。函数 $\delta[\mathbf{x} - \mathbf{x}_{k,n}]$ 当轨迹 $\mathbf{x}$与 $\mathbf{x}_{k,n}$ 完全相同时返回 1，否则返回 0。每条路径 $w_{k,n} \equiv w[\mathbf{x}_{k,n}]$ 的权重未知，但也与有偏路径系综有关。例如，在 TIS 中，我们通过在每个界面施加一个指示函数 $\hat{h}^k_{A,\lambda}[\mathbf{x}]$, 来对路径系综进行偏置；只有当从 A 出发采样的路径穿过界面 λ_k时，该函数才为 1。综合所有 K 次模拟，得到此类偏置系综的分布 $\mathcal{P}^{k'}_{A\ ,\lambda}[\mathbf{x}]$

$$\mathcal{P}^{k'}_{A,\lambda}[\mathbf{x}] = \frac{1}{\mathcal{Z}^{k'}_{A,\lambda}} \sum_k^K \sum_n^{N_k} w_{k,n} \hat{h}^{k'}_{A,\lambda}[\mathbf{x}] \delta[\mathbf{x} - \mathbf{x}_{k,n}], \tag{2}$$

其中 $\mathcal{Z}^{k'}_{A\ ,\lambda} = \sum_k^K \sum_n^{N_k} w_{k,n} \hat{h}^{k'}_{A\ ,\lambda}[\mathbf{x}_{k,n}]$。注意，所有模拟中的所有路径都会对该分布作出贡献。

在 MBAR 中，未知权重 $w_{k,\ n}$ 由观测到采样路径属于其条件系综的似然确定。对数似然表示为

$$\ln L = \ln \prod_k^K \prod_n^{N_k} \mathcal{P}^k_{A,\lambda}[\mathbf{x}_{k,n}] = \sum_k^K \sum_n^{N_k} \ln \mathcal{P}^k_{A,\lambda}[\mathbf{x}_{k,n}]. \tag{3}$$

将方程2中的 $\mathcal{P}^k_{A,\ \lambda}[\mathbf{x}_{k,\ n}]$代入可得

$$\ln L = \sum_k^K \sum_n^{N_k} \ln \frac{w_{k,n} \hat{h}^k_{A,\lambda}[\mathbf{x}_{k,n}]}{\mathcal{Z}^k_{A,\lambda}}. \tag{4}$$

对 $w_{j,\ n}$ 进行优化可得

$$\frac{\partial \ln L}{\partial w_{j,n}} = 0 = \frac{1}{w_{j,n}} - \sum_k^K \frac{N_k \hat{h}^k_{A,\lambda}[\mathbf{x}_{j,n}]}{\mathcal{Z}^k_{A,\lambda}}, \tag{5}$$

其中，因子 N_k 源自对 n的求和。重新整理并除以 $\mathcal{Z}$ 可得

$$\frac{w_{j,n}}{\mathcal{Z}} = \left[\sum_k^K \frac{N_k \hat{h}^k_{A,\lambda}[\mathbf{x}_{j,n}]}{\mathcal{Z}^k_{A,\lambda}/\mathcal{Z}} \right]^{-1}. \tag{6}$$

将此结果代入 $\mathcal{Z}^{k'}_{A\ ,\lambda}$ 的定义式（式 2），得到

$$\frac{Z^{k'}_{A,\lambda}}{\mathcal{Z}} = \sum_j^K \sum_n^{N_j} \frac{\hat{h}^{k'}_{A,\lambda}[\mathbf{x}_{j,n}]}{\sum_k^K \frac{N_k \hat{h}^k_{A,\lambda}[\mathbf{x}_{j,n}]}{\mathcal{Z}^k_{A,\lambda}/\mathcal{Z}}}, \tag{7}$$

为清晰起见，我们在此将外层 k 索引替换为 j。这是 MBAR[37]，中的核心隐函数，与加权直方图分析方法（WHAM）[21] 中的核心方程相同，并且可以通过迭代求解。

通过迭代求解式 7 得到权重后，我们也可以根据式 1 得到无偏概率

$$\mathcal{P}_A[\mathbf{x}] = \sum_j^K \sum_n^{N_j} \frac{\delta[\mathbf{x}-\mathbf{x}_{j,n}]}{\sum_k^K \frac{N_k \hat{h}^k_{A,\lambda}[\mathbf{x}_{j,n}]}{\mathcal{Z}^k_{A,\lambda}/\mathcal{Z}}}. \tag{8}$$

对于 TIS，我们知道，分母中的求和仅包含满足 $\hat{h}^k_A[\mathbf{x}_{k,n}]$非零的路径。将路径穿越的最远界面记为$k_{\max}$（由 $\lambda^{\max}[\mathbf{x}_{j,n}]$定义），分母可简化为

$$w_{j,n} = \left(\sum_k^{k_{\max}[\mathbf{x}_{j,n}]} \frac{N_k \mathcal{Z}}{\mathcal{Z}^k_{A,\lambda}} \right)^{-1}, \tag{9}$$

这看起来与参考文献[18]中用于计算路径权重的 WHAM 表达式非常相似。这意味着每条路径的权重由其穿越的最远界面决定，如 $\lambda^{\max}[\mathbf{x}_{j,n}]$所示。事实上，这是 RPE 方法的核心结果[18]。随后即可由式 1 得到无偏路径概率。

另一个特别值得关注的量是由所有重加权路径构造得到的总穿越概率：

$$P_A(\lambda|\lambda_1) = \sum_j^K \sum_n^{N_j} w_{j,n} \theta(\lambda^{\max}[\mathbf{x}_{j,n}] - \lambda). \tag{10}$$

最后一个界面 $\lambda_B = \lambda_K$ 处的穿越概率为

$$P_A(\lambda_B|\lambda_1) = n_{AB} w_{AB}, \tag{11}$$

其中，$w_{AB} = (\sum_k^K \frac{N_k \mathcal{Z}}{\mathcal{Z}^k_{A,\lambda}})^{-1}$ 是反应路径的权重，而 n_{AB} 是界面系综中采样到的反应路径累计数量。

除了 AB 转变，我们也可以使用 TIS 对反向的 BA 转变进行采样，从而得到其自身的 RPE，权重为 $w^B[\mathbf{x}]$。类似地，我们将 AB 转变的权重记为 $w^A[\mathbf{x}]$。

根据计算得到的权重 $w_{j,n}$，我们可以构建总重加权路径系综（RPE）。从 A 出发的采样路径系综中，每条轨迹 $\mathbf{x}_{j,n}$ 的权重为

$$w^A[\mathbf{x}_{j,n}] = c_A w^A_{j,n}. \tag{12}$$

同样，从 B 出发的采样路径系综中，每条轨迹 $\mathbf{x}_{j,n}$ 的权重为

$$w^B[\mathbf{x}_{j,n}] = c_B w^B_{j,n}, \tag{13}$$

其中，标签 A 和 B 分别区分正向和反向跃迁。未知常数 c_A 和 c_B 通过匹配正向和反向穿越反应路径的概率来确定（例如，使反应路径权重 $n_{AB}w_{AB} = n_{BA}w_{BA}$ 的（通量）贡献在正向和反向轨迹中相等。）

随后，每条采样路径的最终总 RPE 权重就是 $\mathbb{1}_A(x_0)w^A[\mathbf{x}] + \mathbb{1}_B(x_0)w^B[\mathbf{x}]$,之和，其中稳定态特征函数 $\mathbb{1}_A(x)$ 在构型处于状态 A 定义时返回 1，否则返回 0。然而，这个求和只考虑了离开 A（或 B）并穿越任一界面的路径，而未考虑那些始终留在稳定态 A（或 B）中的路径。不过，可以通过在稳定态中进行模拟轻松地纳入这些路径，再根据它们穿过第一个界面的通量进行加权，或采用负系综或正系综（在 RETIS[40,41]中）。

B. 基于 MBAR 的两个 TIS 界面集合的 RPE

现在设想我们有两组 TIS 模拟，分别包含 K_1 和 K_2 个界面。第一组界面集合 $\{\lambda_k\}$ 基于函数 $\lambda(x)$，第二组界面集合 $\{\mu\}$ 基于 $\mu(x)$，后者可能是一个不同的函数。我们假设稳定态定义相同，即 $\lambda_0 \equiv \mu_0$，但所有其他界面$i > 0$ 都由不同的函数 $\lambda(x) \neq \mu(x)$ 描述。通过这些 TIS 模拟，我们获得两组路径 $\mathbf{x}_{k,n}$ 和 $\tilde{\mathbf{x}}_{k,n}$，其中 $\mathbf{x}_{k,n}$ 是在第一组的k界面上采样的第 n条路径，并根据λ界面进行条件化。类似地，$\tilde{\mathbf{x}}_{k,n}$ 表示在第二组的 k界面上采样的第 n条路径，该界面根据 μ 界面进行偏置。这里，与之前一样，索引 k遍历K_1个界面模拟，而 n 遍历每次模拟中的N_k 个样本。其思想是，我们可以根据在无偏系综中观测到路径 $\mathbf{x}$ 的概率，将这两组路径合并为

$$\mathcal{P}_A[\mathbf{x}] = \frac{1}{\mathcal{Z}} \left(\sum_k^{K_1} \sum_n^{N_k} w_{k,n} \delta[\mathbf{x}-\mathbf{x}_{k,n}] + \sum_k^{K_2} \sum_n^{\tilde{N}_k} \tilde{w}_{k,n} \delta[\mathbf{x}-\tilde{\mathbf{x}}_{k,n}] \right), \tag{14}$$

其中因子

$$\mathcal{Z} = \sum_k^{K_1} \sum_n^{N_k} w_{k,n} + \sum_k^{K_2} \sum_n^{N_k} \tilde{w}_{k,n}, \tag{15}$$

对分布进行归一化。

每条路径 $w_{k,n}$ 和 $\tilde{w}_{k,n}$ 的权重未知，但与其采样自的有偏的过渡界面采样路径系综有关。与上一节一样，满足基于 λ穿越第 k 个界面函数这一条件的路径分布由指示函数给出

函数 $\hat{} h^k_{A\lambda}[\mathbf{x}]$,为:

$$\mathcal{P}^{k'}_{A,\lambda}[\mathbf{x}] = \frac{1}{\mathcal{Z}^{k'}_{A,\lambda}} \left(\sum_k^{K_1} \sum_n^{N_k} w_{k,n} \hat{h}^{k'}_{A,\lambda}[\mathbf{x}] \delta[\mathbf{x} - \mathbf{x}_{k,n}] + \sum_k^{K_2} \sum_n^{\tilde{N}_k} \tilde{w}_{k,n} \hat{h}^{k'}_{A,\lambda}[\mathbf{x}] \delta[\tilde{\mathbf{x}} - \tilde{\mathbf{x}}_{k,n}] \right). \quad (16)$$

请注意，所有 $K_1 + K_2$ TIS 模拟中的采样路径都可以对该分布作出贡献。条件配分函数为

$$\mathcal{Z}^{k'}_{A,\lambda} = \sum_k^{K_1} \sum_n^{N_k} w_{k,n} \hat{h}^{k'}_{A,\lambda}[\mathbf{x}_{k,n}] + \sum_k^{K_2} \sum_n^{\tilde{N}_k} \tilde{w}_{k,n} \hat{h}^{k'}_{A,\lambda}[\tilde{\mathbf{x}}_{k,n}]. \quad (17)$$

我们可以利用基于 μ 函数的界面，考察类似的路径分布，并使用 $\hat{} h^k_{A,\mu}[\mathbf{x}]$ 作为指示函数:

$$\mathcal{P}^{k'}_{A,\mu}[\mathbf{x}] = \frac{1}{\mathcal{Z}^{k'}_{A,\mu}} \left(\sum_k^{K} \sum_n^{N_k} w_{k,n} \hat{h}^{k'}_{A,\mu}[\mathbf{x}] \delta[\mathbf{x} - \mathbf{x}_{k,n}] + \sum_k^{K} \sum_n^{\tilde{N}_k} \tilde{w}_{k,n} \hat{h}^{k'}_{A,\mu}[\mathbf{x}] \delta[\tilde{\mathbf{x}} - \tilde{\mathbf{x}}_{k,n}] \right), \quad (18)$$

其中

$$\mathcal{Z}^{k'}_{A,\mu} = \sum_k^{K_1} \sum_n^{N_k} w_{k,n} \hat{h}^{k'}_{A,\mu}[\mathbf{x}_{k,n}] + \sum_k^{K_2} \sum_n^{\tilde{N}_k} \tilde{w}_{k,n} \hat{h}^{k'}_{A,\mu}[\tilde{\mathbf{x}}_{k,n}]. \quad (19)$$

由于路径 $\mathbf{x}_{k,n}$ （对应于 $n \in 1, \dots N_k$）是使用函数 $\lambda(x)$ 在 λ_k 处采样的，因此必须根据这些路径在条件分布 $\mathcal{P}^k_{A,\lambda}[\mathbf{x}]$下的似然来确定相应的权重$w_{k,n}$。类似地，路径 $\tilde{}\mathbf{x}_{k,n}$是按照 μ 在μ_k处采样的，因此来自条件分布 $\tilde{\mathcal{P}}^k_{A,\mu}[\mathbf{x}]$。然而，配分和 $Z^k_{A,\lambda}$ 和 $\tilde{Z}^k_{A,\mu}$ 是利用两组路径计算的。这种联合构造使我们能够从合并数据集中提取有关无偏分布的额外信息，从而改进对配分和的估计。

从第 k 个 λ 偏置系综中抽取的路径 $\mathbf{x}$ 的似然为

$$L_{k,\lambda}(\mathbf{x}|\{w_{k,n}\}, \{\tilde{w}_{k,n}\}) = \mathcal{P}^k_{A,\lambda}[\mathbf{x}], \quad (20)$$

对于 $\tilde{}\mathbf{x}$ 路径，则有

$$L_{k,\mu}(\tilde{\mathbf{x}}|\{w_{k,n}\}, \{\tilde{w}_{k,n}\}) = \mathcal{P}^k_{A,\mu}[\tilde{\mathbf{x}}], \quad (21)$$

基于 μ偏置数据。因此，两组观测数据的对数似然为

$$\begin{aligned} \ln L &= \ln \prod_k^{K_1} \prod_n^{N_k} \mathcal{P}^k_{A,\lambda}[\mathbf{x}_{k,n}] \prod_k^{K_2} \prod_n^{\tilde{N}_k} \mathcal{P}^k_{A,\mu}[\tilde{\mathbf{x}}_{k,n}] \\ &= \sum_k^{K_1} \sum_n^{N_k} \ln \mathcal{P}^k_{A,\lambda}[\mathbf{x}_{k,n}] + \sum_k^{K_2} \sum_n^{\tilde{N}_k} \ln \mathcal{P}^k_{A,\mu}[\tilde{\mathbf{x}}_{k,n}]. \quad (22) \end{aligned}$$

代入分布 16 和 18，得到

$$\begin{aligned} \ln L = &\sum_k^{K_1} \sum_n^{N_k} (\ln w_{k,n} \hat{h}^k_{A,\lambda}[\mathbf{x}_{k,n}] - \ln Z^k_{A,\lambda}) \\ &+ \sum_k^{K_2} \sum_n^{\tilde{N}_k} (\ln \tilde{w}_{k,n} \hat{h}^k_{A,\mu}[\tilde{\mathbf{x}}_{k,n}] - \ln Z^k_{A,\mu}). \quad (23) \end{aligned}$$

对 $w_{j,n}$ 和 $\tilde{w}_{j,n}$ 最大化此式，得到

$$\frac{\partial \ln L}{\partial w_{j,n}} = 0 = \frac{1}{w_{j,n}} - \sum_k^{K_1} \frac{N_k \hat{h}^k_{A,\lambda}[\mathbf{x}_{j,n}]}{\mathcal{Z}^k_{A,\lambda}} - \sum_k^{K_2} \frac{\tilde{N}_k \hat{h}^k_{A,\mu}[\mathbf{x}_{j,n}]}{\mathcal{Z}^k_{A,\mu}} \quad (24)$$

and

$$\frac{\partial \ln L}{\partial \tilde{w}_{j,n}} = 0 = \frac{1}{\tilde{w}_{j,n}} - \sum_k^{K_1} \frac{N_k \hat{h}^k_{A,\lambda}[\tilde{\mathbf{x}}_{j,n}]}{\mathcal{Z}^k_{A,\lambda}} - \sum_k^{K_2} \frac{\tilde{N}_k \hat{h}^k_{A,\mu}[\tilde{\mathbf{x}}_{j,n}]}{\mathcal{Z}^k_{A,\mu}} \quad (25)$$

重新整理式 (24) 和式 (25)，得到:

$$\frac{w_{j,n}}{\mathcal{Z}} = \left[\sum_k^{K_1} \frac{N_k \hat{h}^k_{A,\lambda}[\mathbf{x}_{j,n}]}{\mathcal{Z}^k_{A,\lambda}/\mathcal{Z}} + \sum_k^{K_2} \frac{\tilde{N}_k \hat{h}^k_{A,\mu}[\mathbf{x}_{j,n}]}{\mathcal{Z}^k_{A,\mu}/\mathcal{Z}} \right]^{-1} \quad (26)$$

and

$$\frac{\tilde{w}_{j,n}}{\mathcal{Z}} = \left[\sum_k^{K_1} \frac{N_k \hat{h}^k_{A,\lambda}[\tilde{\mathbf{x}}_{j,n}]}{\mathcal{Z}^k_{A,\lambda}/\mathcal{Z}} + \sum_k^{K_2} \frac{\tilde{N}_k \hat{h}^k_{A,\mu}[\tilde{\mathbf{x}}_{j,n}]}{\mathcal{Z}^k_{A,\mu}/\mathcal{Z}} \right]^{-1}, \quad (27)$$

这些权重实际上是相同的。将这些权重代入条件配分函数的表达式，得到

$$\begin{aligned} \frac{\mathcal{Z}^{k'}_{A,\lambda}}{\mathcal{Z}} = &\sum_j^{K_1} \sum_n^{N_j} \frac{h^{k'}_{A,\lambda}[\mathbf{x}_{j,n}]}{\sum_k^{K_1} \frac{N_k \hat{h}^k_{A,\lambda}[\mathbf{x}_{j,n}]}{\mathcal{Z}^k_{A,\lambda}/\mathcal{Z}} + \sum_k^{K_2} \frac{\tilde{N}_k \hat{h}^k_{A,\mu}[\mathbf{x}_{j,n}]}{\mathcal{Z}^k_{A,\mu}/\mathcal{Z}}} \\ &+ \sum_j^{K_2} \sum_n^{\tilde{N}_j} \frac{h^{k'}_{A,\lambda}[\tilde{\mathbf{x}}_{j,n}]}{\sum_k^{K_1} \frac{N_k \hat{h}^k_{A,\lambda}[\tilde{\mathbf{x}}_{j,n}]}{\mathcal{Z}^k_{A,\lambda}/\mathcal{Z}} + \sum_k^{K_2} \frac{\tilde{N}_k \hat{h}^k_{A,\mu}[\tilde{\mathbf{x}}_{j,n}]}{\mathcal{Z}^k_{A,\mu}/\mathcal{Z}}}, \end{aligned} \quad (28)$$

类似地

$$\begin{aligned} \frac{\mathcal{Z}^{k'}_{A,\mu}}{\mathcal{Z}} = &\sum_j^{K_1} \sum_n^{N_j} \frac{\hat{h}^{k'}_{A,\mu}[\mathbf{x}_{j,n}]}{\sum_k^{K_1} \frac{N_k \hat{h}^k_{A,\lambda}[\mathbf{x}_{j,n}]}{\mathcal{Z}^k_{A,\lambda}/\mathcal{Z}} + \sum_k^{K_2} \frac{\tilde{N}_k \hat{h}^k_{A,\mu}[\mathbf{x}_{j,n}]}{\mathcal{Z}^k_{A,\mu}/\mathcal{Z}}} \\ &+ \sum_j^{K_2} \sum_n^{\tilde{N}_j} \frac{\hat{h}^{k'}_{A,\mu}[\tilde{\mathbf{x}}_{j,n}]}{\sum_k^{K_1} \frac{N_k \hat{h}^k_{A,\lambda}[\tilde{\mathbf{x}}_{j,n}]}{\mathcal{Z}^k_{A,\lambda}/\mathcal{Z}} + \sum_k^{K_2} \frac{\tilde{N}_k \hat{h}^k_{A,\mu}[\tilde{\mathbf{x}}_{j,n}]}{\mathcal{Z}^k_{A,\mu}/\mathcal{Z}}}, \end{aligned} \quad (29)$$

通过迭代即可求解，从而得到一种针对通过不同界面函数获得的两个 TIS 集合的 MBAR 程序。当两个集合的界面函数相同时，该方法与单个集合的原始 MBAR 得到相同的解（见附录）。

迭代求解双集合 MBAR 方程

给定初始值 $\mathcal{Z}_A^{k'}{}_{,(\lambda/\mu)} = 1$，上述耦合方程组可以通过迭代求解 $\mathcal{Z}_A^{k'}{}_{,(\lambda/\mu)}/\mathcal{Z}$，从而得到每条路径的权重。与标准 MBAR（或加权直方图分析方法）类似，权重表达式中的分母仅包含满足 $\hat{} \; h^k_{A(\lambda/\mu)}[\mathbf{x}] \neq 0$ 的路径所作的贡献。分别将 λ 和 μ 个 CV 中穿越的最大界面记为 $k^{\lambda}_{\max}$ 和 $k^{\mu}_{\max}$，则式 26 中的路径权重为：

$$w[\mathbf{x}_{j,n}] = \left[\sum_k^{k^{\lambda}_{\max}[\mathbf{x}_{j,n}]} \frac{N_k}{\mathcal{Z}^k_{A,\lambda}/\mathcal{Z}} + \sum_k^{k^{\mu}_{\max}[\mathbf{x}_{j,n}]} \frac{\tilde{N}_k}{\mathcal{Z}^k_{A,\mu}/\mathcal{Z}} \right]^{-1}. \tag{30}$$

这意味着每条路径的权重由它在两个 λ 和 μ 界面上达到的最大值决定。因此，跟踪每条路径的最大值，就能计算λ 和 μ 采样轨迹中每条轨迹的权重。对于给定的 $\mathcal{Z}_A^{k'}{}_{,(\lambda/\mu)}/\mathcal{Z}$，使用式 30 中的权重，我们可以重新计算配分和：

$$\frac{\mathcal{Z}^{k'}_{A,\lambda}}{\mathcal{Z}} = \sum_j^{K_1}\sum_n^{N_j} \hat{h}^{k'}_{A,\lambda}[\mathbf{x}_{j,n}] w[\mathbf{x}_{j,n}] + \sum_j^{K_2}\sum_n^{\tilde{N}_j} \hat{h}^{k'}_{A,\lambda}[\tilde{\mathbf{x}}_{j,n}] w[\tilde{\mathbf{x}}_{j,n}]$$

$$\frac{\mathcal{Z}^{k'}_{A,\mu}}{\mathcal{Z}} = \sum_j^{K_1}\sum_n^{N_j} \hat{h}^{k'}_{A,\mu}[\mathbf{x}_{j,n}] w[\mathbf{x}_{j,n}] + \sum_j^{K_2}\sum_n^{\tilde{N}_j} \hat{h}^{k'}_{A,\mu}[\tilde{\mathbf{x}}_{j,n}] w[\tilde{\mathbf{x}}_{j,n}],$$

并对两个集合和界面重复此操作，直至收敛。

穿越概率

现在可以根据跨越 λ 和 μ 两个集合的完整重加权轨迹系综，构造得到的穿越概率 $P_A(\lambda|\lambda_1)$：

$$P_A(\lambda|\lambda_1) = \sum_k^{K_\lambda}\sum_n^{N_k} \theta(\lambda^{\max}[\mathbf{x}_{k,n}] - \lambda) w^A[\mathbf{x}_{k,n}] \hat{h}_{A,\lambda_1}[\mathbf{x}_{k,n}] + \sum_k^{K_\mu}\sum_n^{\tilde{N}_k} \theta(\lambda^{\max}[\tilde{\mathbf{x}}_{k,n}] - \lambda) w^A[\tilde{\mathbf{x}}_{k,n}] \hat{h}_{A,\lambda_1}[\tilde{\mathbf{x}}_{k,n}], \tag{31}$$

这里，指示函数 $h_{A,\lambda_1}[\mathbf{x}]$明确规定路径必须穿越参考界面 λ_1，从而确保对穿越概率进行正确的条件化。在基于单个界面集合的标准 TIS 中，由于所有路径都会越过 λ_1，这一条件总是隐式满足的；但在双集合 MBAR 公式中，必须将其显式纳入，尤其是在其中一个系综来自不同的 CV，或包含位于 λ_1之前的界面时。对于 $P_A(\mu|\mu_1)$，也有类似的表达式。

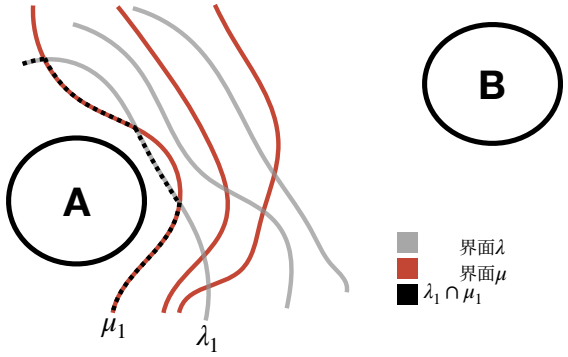

图 1. λ 和 μ 界面交集的示意图，以及联合初始界面的定义：轨迹穿越的第一个界面。

由两个集合构造 RPE

利用公式30计算出的正向（$A \to B$）和反向（$B \to A$）跃迁的双集合权重，我们可以采用类似单集合方案的方式构建总反应路径系综。每条从 A 出发的轨迹（来自 λ 和 μ 集合）都会被赋予一个权重，并像公式12中那样乘以常数因子 c_A，用于正向系综。同样，从 B 出发的轨迹则根据反向系综赋予权重，并乘以因子 c_B，如公式13所示。随后，我们调整未知常数 c_A 和 c_B，确保反应路径的加权通量贡献在正向和反向过程中相等。

通过 TIS 构造的 RPE 本身不包含稳定态布居信息。与单集合情形一样，可以通过模拟稳定态，并匹配穿过 TIS 系综进行条件采样所依据的第一个界面的通量，事后引入稳定态信息[18,33]。对于双集合情形，可以匹配穿过 λ 和 μ 集合的第一个界面交集的通量，具体来说，就是两个界面定义首次重叠的区域。这个匹配界面代表进入条件系综的入口，也是 TIS 系综中所包含的轨迹进行条件采样的区域边界。

匹配在第一个界面（$\lambda_1 \cap \mu_1$）的交集处进行，如图 1 中的黑色虚线所示。该交集定义了相空间中任何条件采样都未应用的区域。所有纳入 RPE 的轨迹都会穿过此界面，因此它为对齐稳定态通量提供了自然基准。

所有来自 A 且属于两个集合 RPE 系综的路径的配分函数 $\mathcal{Z}_A$，也就是那些经过相交界面的路径，代表进入条件采样区域的所有路径的总权重（即所有来自 TIS 模拟的路径）。通过将稳定态模拟的通量与 TIS RPE 的通量相匹配，我们确保

稳定态轨迹的通量与 TIS RPE 的通量相匹配$n_{A,\ \lambda_1\cap\mu_1} w^A_{\rm ss} = \mathcal{Z}_A$，其中 $w^A_{\rm ss}$ 是分配给稳定态轨迹的权重，用以确保通量匹配；而$n_{A,\ \lambda_1\cap\mu_1}$ 是通过联合界面的有效第一个界面穿越次数。为简单起见，移除稳定态轨迹中超过 $\lambda_1 \cap \mu_1$ 界面的部分，并移除 RPE 轨迹中穿越 $\lambda_1 \cap \mu_1$ 之前的部分。

C. 扩展到 M 个界面集合

针对两个集合情形给出的推导可直接推广到任意数量的界面系综集合 M。每个集合 i 均由其自身的界面函数 $\lambda^{(i)}$ 定义，其中包含 $K^{(i)}$ 个界面及相应的条件配分函数$\mathcal{Z}^k_{A,\ \lambda^{(i)}}$。假设所有集合共享相同的稳定态（$\lambda^{(i)}_0 = \lambda_0$），则可通过最大化所有采样路径的联合似然，得到无偏路径权重的自洽表达式。由此，每条路径$\mathbf{x}$ 的权重为

$$w^A[\mathbf{x}] = \left[\sum_{i=1}^{M}\sum_{k}^{K^{(i)}} \frac{N^{(i)}_k\, \hat{h}^k_{A,\lambda^{(i)}}[\mathbf{x}]}{\mathcal{Z}^k_{A,\lambda^{(i)}}/\mathcal{Z}}\right]^{-1}, \tag{32}$$

其中，$N^{(i)}_k$ 表示界面系综 k 中集合 i内采样路径的数量，而 $\hat{h}^k_{A,\ \lambda^{(i)}}[\mathbf{x}]$是指示函数；若路径满足相应的界面条件，则其值为 1。

由于对于沿 $\lambda^{(i)}$的$\mathbf{x}$ 所达到的最大交叉界面之后的界面，$\hat{h}^k_{A,\ \lambda^{(i)}}[\mathbf{x}] = 0$ 为零，因此权重仅取决于$k^{(i)}_{\max}$，即每个集合中达到的最高界面：

$$w^A[\mathbf{x}] = \left[\sum_{i=1}^{M}\sum_{k}^{k^{(i)}_{\max}[\mathbf{x}]} \frac{N^{(i)}_k}{\mathcal{Z}^k_{A,\lambda^{(i)}}/\mathcal{Z}}\right]^{-1}. \tag{33}$$

此表达式将两个集合的结果推广至所有 M 界面系综集合，同时保留了 MBAR 理想的归一化和重加权特性。MultiSet-MBAR 程序在 $M = 1$ 时可得回 Eq. 9，在 $M = 2$时可得回 Eq. 30。用于重构路径平衡分布的权重仅取决于每个集合中跨越的最大界面。配分函数的迭代求解$\mathcal{Z}^k_{A,\ \lambda^{(i)}}$ 随后为

$$\frac{\mathcal{Z}^{k'}_{A,\lambda^{(i)}}}{\mathcal{Z}} = \sum_{i=1}^{M}\sum_{j}^{K^{(i)}}\sum_{n}^{N^{(i)}_j} \hat{h}^{k'}_{A,\lambda^{(i)}}[\mathbf{x}^{(i)}_{j,n}]\, w^A[\mathbf{x}^{(i)}_{j,n}] \tag{34}$$

这些表达式的详细推导见附录 A2 节。

III. 结果

A. 二维双阱

我们在二维双势阱势能上测试了 MBAR 重加权程序，该势能由 x 中的一维双势阱势能和 y中的附加谐势能构成：

$$U_{\rm dw}(x,y) = A(x^2 - x_0^2)^2 + \omega y^2, \tag{35}$$

其中 $A = 1$、$x_0 = 1$ 和 $\omega = 1$，由此在 $x = 0$ 处形成一个鞍点，并在 $x = \pm 1$ 处形成两个极小值。对于 TIS 模拟，系统的稳定态定义为：A 态为 $x < -0.9$，B 态为 $x > 0.9$。动力学模拟采用 OpenPathSampling (OPS) 软件包[8] 中的 Langevin BOAOB 积分器[42]，温度设为 $T_0 = 0.1$，摩擦系数为 $\gamma = 10$，时间步长为 $dt = 0.05$。由此得到的能垒高度为 $1/T_0$ 个热能单位，其中k_B 为玻尔兹曼常数。

在稳定态中直接进行朗之万模拟，会使采样集中在极小值附近，而不会覆盖能垒。额外的 TIS 模拟将增强对极小值之间跃迁的采样。

为研究双集合和 MultiSet-MBAR 程序，我们定义了一个带有可调参数的简单界面函数，以构造多种不同类型的界面函数。具体而言，我们希望研究一个可在 x-y平面内旋转、旋转角为 θ的平面界面集合。此外，我们还希望能够加入一些偏离线性的扰动，例如使平面产生起伏。为同时实现这两点，我们为受正弦扰动的平面定义了如下函数，

$$\lambda(x,y) = x\cos\theta + y\sin\theta + A_{\rm mag}\sin(2\pi f y), \tag{36}$$

其中，θ 为旋转角，f 为正弦频率，$A_{\rm mag}$ 为扰动幅度。通过改变这些参数，可以生成不同的旋转和调制后的界面形状，以评估双集合和 MultiSet-MBAR 方法的稳健性。为确保 TIS 界面不与稳定态定义相交，在 OPS 中，必须将状态定义明确纳入界面系综定义，这样无论使用哪种具体的界面函数，只有当至少一帧位于稳定态区域之外时，轨迹才会被接受[8]。

B. 双集合 TIS MBAR 结果

作为初步测试，我们使用两个进行 TIS 的界面集合构建 RPE。第一个集合标记为$\{\lambda\}$，由沿 x 轴垂直定义的界面组成，编码为（$\theta = 0$、$f = 0$和 $A_{\rm mag} = 0$）。第二个集合标记为 $\{\mu\}$，由倾斜界面定义，编码为（$\theta = 5°$，$f = 0$，$A_{\rm mag} = 0$）。

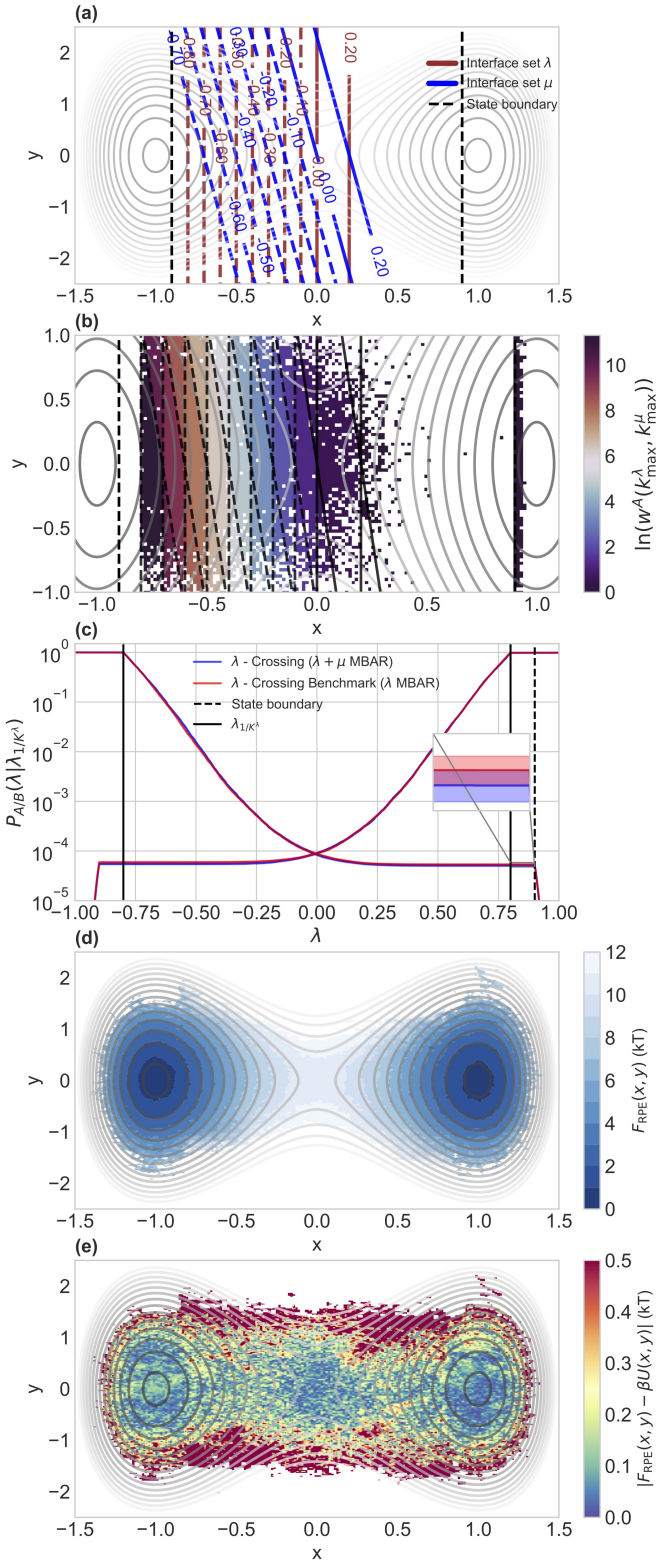

图 2。（a）双势阱势能，以及沿着 λ 和 μ 为正向系综设置的两组界面的示意图。（b）根据每条轨迹在 λ 和 μ 界面空间中的最大 k_λ^{max}和 k_μ^{max}所分配的权重。（c）通过双集合 MBAR 以及单集合基准值计算得到的正向和反向系综的穿越概率，结果表明双集合程序能够恢复相同的穿越概率。（d）RPE 的自由能与势能面叠加图。（e）相对于势能面的自由能误差。

这些集合的界面值为$\{\lambda\} = [-0.8, -0.7, -0.6, -0.5, -0.4, -0.3, -0.2, -0.1, 0.0, 0.2]$和 $\{\mu\} = [-0.7, -0.6, -0.5, -0.4, -0.3, -0.2, -0.1, 0.0, 0.2]$。图 2a 展示了这两个集合的示意图。

我们对每个界面集合都采用双向射击执行了 TIS。使用宽度为 0.5 的高斯选择器，在界面附近选择射击点。每个界面的试验次数为 $N = 5000$。路径长度从低界面的 10 帧变化到最高界面的 900 帧。利用每条采样轨迹在 λ 和 μ 空间中达到的最大值，并通过对每个界面的配分和（式 31）进行迭代重新计算来求解式 30，我们获得了每条轨迹的权重。对于反向系综，也可以采用相同的程序，但此时使用达到的最小值。

图 2b 绘出了根据路径在 x-y 平面上达到最大值的位置所赋予的权重，突出显示了现在路径的权重取决于 λ 和 μ 界面空间中的最大值 k^{λ}_{max} 和 k^{μ}_{max}。

使用公式 31，沿着 λ 函数计算采用两组界面的正向和反向系综的穿越概率，如图 2c 所示。我们对每个界面包含 4900 条路径的数据集进行自助法程序处理（这些路径来自一次 5000 蒙特卡罗步的 TIS 模拟，其中已移除前 100 条路径）。自助法程序使用大小为 10 条路径的区块，以反映数据集中的相关性，并进行 100 次自助法重采样，从而得到穿越概率的统计误差和均值，如图所示。图 2c 中的穿越概率图显示，对于由每个界面包含 $N = 10000$ 条路径组成的 λ 界面集合上的基准 TIS 计算，双集合程序在统计误差范围内恢复了与单集合程序相同的 $\lambda(x) = x$ 的穿越概率 $P_A(\lambda|\lambda_1)$。

为评估 RPE自由能相对于真实自由能面的准确性，我们采用平均绝对误差（平均绝对误差）。然而，由于边缘处的一些分箱中点数很少，它们可能会对总体误差产生不成比例的影响。为减轻这一问题，我们计算加权平均绝对误差，其中每个分箱的权重由其局部点密度决定 $h(x, y)$，从而降低采样稀疏区域的贡献。图2d展示了投影到 x-y 平面上的 RPE 自由能，其中各状态按照通过共享第一界面的通量进行对齐（$\lambda_1 \cap \mu_1$）。与真实势能$\beta U_{\mathrm{dw}}(x, y)$的偏差见图2e。由于采样有限，外围区域的误差较高。为考虑这一点，我们定义加权平均绝对误差：

$$\mathrm{MAE} = \sum_{x,y} h(x, y) \left| F_{\mathrm{RPE}}(x, y) - \beta U_{dw}(x, y) \right|, \qquad (37)$$

其中 $h(x,\ y)$ 是 RPE 点按分箱统计得到的归一化直方图。我们使用下文的加权平均绝对误差来评估 RPE 的质量。

C. 多集合 TIS MBAR

上文展示了如何使用双集合 MBAR 程序，合并按照不同界面函数采样的两个 TIS 界面系综集合λ，μ。该程序可推广至 M 个集合，其中权重由界面函数联合空间中的最大取值决定（式 33）。MultiSet-MBAR 程序确保不同系综之间路径权重的相对缩放一致，因为无偏权重是根据耦合所有已采样路径系综、并明确考虑它们相互重叠情况的配分函数得到的。

1. 穿越概率分析

首先，我们关注重加权过程对穿越概率的影响

$$P_A(\lambda_B|\lambda_1) = \sum_{\mathbf{x}\in\{\mathbf{x}\}^{(M)}}^{RPE} w[\mathbf{x}]\theta(\lambda^{\max}[\mathbf{x}]-\lambda)h_{A,\lambda_1}[\mathbf{x}], \quad (38)$$

如公式 10 所示计算，但现在对合并的 RPE 中所有权重$\{w\}$和路径 $\{\mathbf{x}\}$ 求和，该 RPE 包含 M 组模拟。为清楚起见，这里使用 $\sum^{RPE}_{\mathbf{x}\in\{\mathbf{x}\}} = \sum_i^M \sum_k^{K^{(i)}} \sum_n^{N_k^{(i)}}$。为了评估 MultiSet-MBAR 重加权相较于对独立 RPE 进行选定的重新缩放组合所带来的改进，我们采用两种不同的方法构建集合 $\{w\}^{(M)}$ 和 $\{\mathbf{x}\}^{(M)}$。

在 MultiSet-MBAR 方法中，根据公式 33 $\in \{\ \}^{(M)} \equiv$为每条路径分配权重 $\mathbf{x}\,\mathbf{x}\mathbf{x},\mathbf{x}$，其中 $\mathbf{x}$是来自 M 次独立 TIS 模拟的路径合并集合，每次模拟均使用不同的界面集合。 $\{\{\ \}^1 \cdots \{\ \}^M\}$ $\{\ \}^{(M)}$

在另一种重缩放方法（称为反应路径匹配）中，使用公式 9 为每个子集 $\{\mathbf{x}\}^i$ 独立计算权重。由于每次单独加权都带有任意的归一化常数，我们将所有权重重缩放，使反应路径的权重为 1。随后，将每个子集中的所有路径合并为一个 RPE。另一种选择是，将每个集合中首个界面上的权重重缩放，使其与通过独立分子动力学模拟在稳定盆地中确定的通量相匹配。在这两种替代重缩放方法中，每个界面系综集合 $\lambda^{(i)}$ 的分区项 $\mathcal{Z}^k_{A\lambda^{(i)}}$ 都是独立确定的，且各自具有一个（未知的）相对偏移量。相比之下，MultiSet-MBAR 方法在重加权过程中同时纳入所有系综分布，自然地将这些分区项校准到同一尺度上。

我们比较了 $M = 10$ 个界面函数下的 MultiSet-MBAR 和反应匹配方法（示意图见附录中的图 6）。对于每个界面函数 $\lambda^{(i)}$，我们进行 10 次 TIS 模拟，覆盖从 $\lambda_1^{(i)} = -0.8$ 到 0.0 的范围，增量为 0.1，并在 0.2 处额外设置一个界面。因此，每个界面集 $i \in \{1,\ldots,M\}$ 都贡献了系综 $P^k_{A,\lambda^{(i)}}$，对应的界面索引为 $k \in \{1,\ldots,K^{(i)}\}$，这些系综通过在 $\lambda^{(i)}$ 上进行 TIS 条件路径采样得到。由于将新界面集加入加权过程的顺序会影响最终结果，我们将比较两种不同的顺序，分别标记为“由优到劣”和“由劣到优”。这两种顺序的区别取决于界面函数与理想界面集（此处为 x-界面）的差异程度，并通过一个简单的评分函数来衡量（见附录中的公式 B1）。

在自助法程序中，我们从每个 TIS 系综 $P^k_{A,\lambda^{(i)}}$中抽取 N 条路径，块大小为 1，从而为每个系综生成一组相关性很弱的路径，并重复此过程 $N_{\mathrm{bs}} = 100$ 次以获得统计误差。

在图 3a 中，我们展示了预测的穿越概率 $\log p_A(\lambda_B|\lambda_1)$ 如何随每个界面的采样路径数量变化，分别为 $N = 10$、50、100 和 1000。相对于使用沿 x-界面集合采样的 $N = 10000$ 条路径得到的基准值，这些结果的偏差如图 3b 所示。

随着纳入的界面集合增加，MultiSet-MBAR 估计值逐渐收敛至基准值。相比之下，当 N 较小时，反应匹配方法的结果偏离基准值，且随着集合数量增加，其准确性会降低。当 N 增大时，两种方法的结果开始更加接近；对于$N \geq 100$，反应匹配方法也开始与基准值高度吻合。

在图 3c 中，我们展示了相对误差，其定义为$\sigma_{\mathrm{bs}}/\mathbb{E}(\log P(\lambda_B|\lambda_1))_{\mathrm{bs}}$，并将其表示为集合数量 M 的函数。当 N 较小时，即使增加集合数量，反应匹配误差仍大致保持不变或趋于平稳。相比之下，MultiSet-MBAR 误差持续下降，其变化近似遵循$1/\sqrt{M}$ 标度关系，其中 M 为集合数量。当 N 足够大时，基于反应匹配的方法的误差也会下降，但准确性仍不及 MultiSet-MBAR。

这些结果表明，对于较小的样本量，随着纳入更多数据，MultiSet-MBAR 方法的性能会持续提升，而基于反应性的独立重标定方法则可能出现性能下降。在数据充足的情况下，两种方法都能从增加集合中受益，但 MultiSet-MBAR 始终保持较低的统计误差。

请注意，添加界面的顺序对结果准确性的影响似乎相对较小，因为两条橙色曲线和两条蓝色曲线彼此接近。不过，仔细观察会发现，对于单个集合（M=1），从最差到最佳的顺序下误差更大，这与预期一致；但误差很快就会下降，随后（自然地）收敛到相同的答案。界面集合选择顺序对结果相对不敏感，原因可能在于所有界面集合都能够对跃迁进行采样。事实上，这是 TIS[10] 的标志之一。

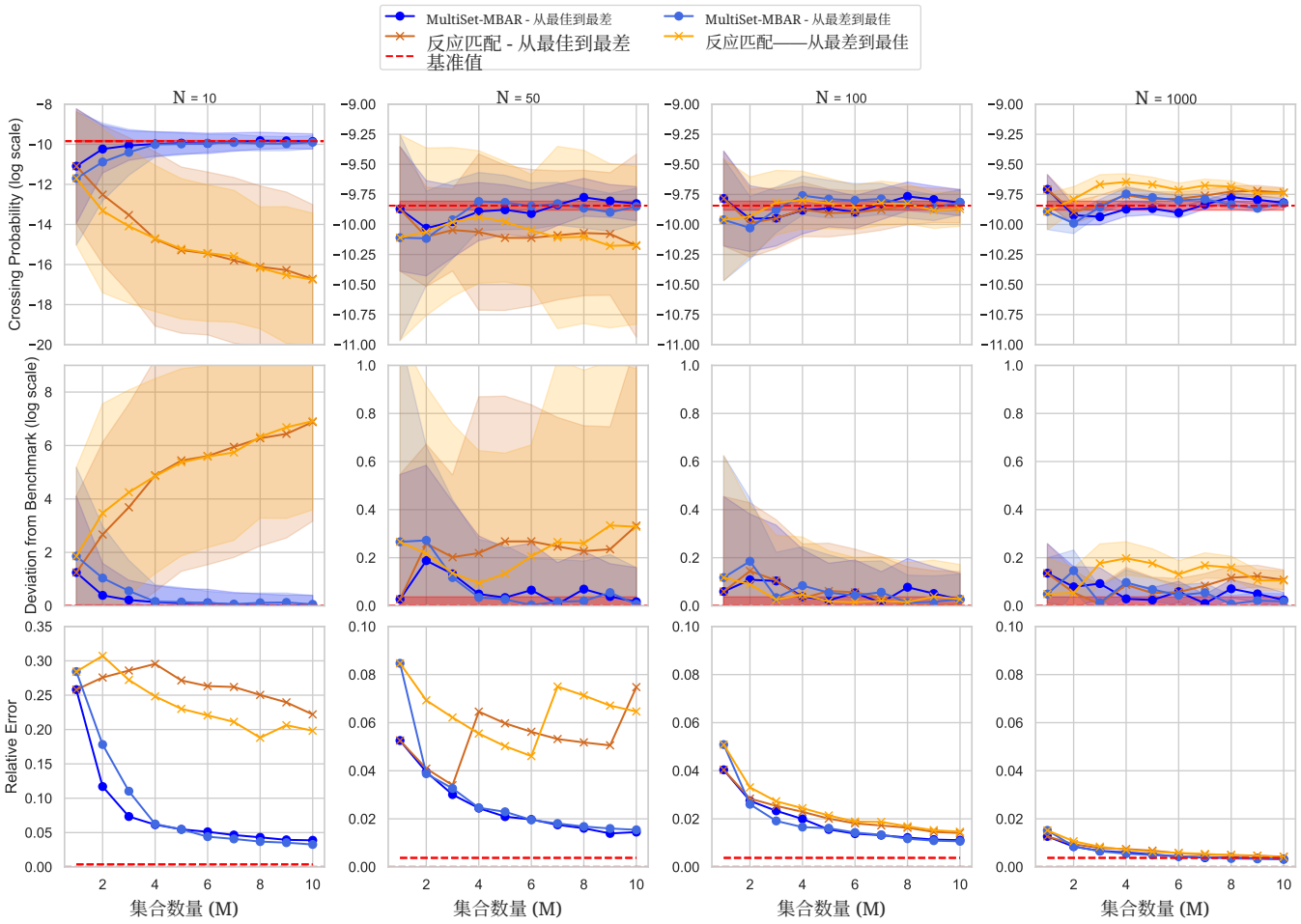

图 3。（上排）对于反应匹配方法和 MultiSet-MBAR 方法，展示了不同界面路径数 N 下，估计的穿越概率 $P_A(\lambda_B|\lambda_1)$ 随合并的集合数量 M 变化的情况，其中使用了 $\lambda(x) = x$ 。结果与从沿着 x 界面集合采样的 $N = 10000$ 条路径计算得到的基准值进行比较。随着加入更多集合，MultiSet-MBAR 估计值平稳收敛至基准值；而独立反应匹配估计值在 N 较小时起初会偏离，只有在样本量较大时才与基准值吻合。（中排）对于独立反应匹配方法和 MultiSet-MBAR 方法，展示了不同界面路径数 N 下，估计的穿越概率相对于基准值的对数偏差随合并的集合数量 M 变化的情况。随着纳入更多数据，MultiSet-MBAR 方法始终能减小偏差，表明其收敛性更佳。相比之下，反应匹配方法在 N 较小时偏差会增大，尤其是在加入更多集合时。（下排）对于独立反应匹配方法和 MultiSet-MBAR 方法，展示了不同界面路径数 N 下，穿越概率估计值的相对统计误差（定义为 $\sigma_{\text{bs}}/\mathbb{E}[\log P_A(\lambda_B|\lambda_1)]_{\text{bs}}$）随合并的集合数量$M$ 变化的情况。随着集合数量增加，MultiSet-MBAR 方法的相对误差持续减小，大致呈 $1/\sqrt{M}$ 趋势。相比之下，反应匹配方法在 N 较小时误差会趋于平稳，只有当样本量足够大时才会有所改善。

—

2. RPE分布与自由能

通过使用 MultiSet-MBAR 方法构建 RPE，我们还可以估计系统的构型分布和自由能。图 2b 展示了所得自由能的投影，其中包含了通过第 IIB 节所述程序匹配的稳定态数据。

为了量化纳入额外集合对 RPE 的改进效果，我们使用式 37 计算 MultiSet-MBAR 自由能估计值与势能面之间的加权平均绝对误差。具体而言，我们重点关注两个稳定态（此处定义为 $x > -0.9$ 和 $x < 0.9$）之间的过渡区域，因为该区域对穿越概率和自由能估计都尤为敏感。

在图4中，我们展示了加权平均绝对误差如何随着用于 MultiSet-MBAR 重加权的组合集合数量增加而降低。我们使用不同的样本量 $N = 10$, 50、100、1000 评估这一趋势，并采用自助抽样程序，从每个系综中采样 N 个样本，以估计统计不确定性。

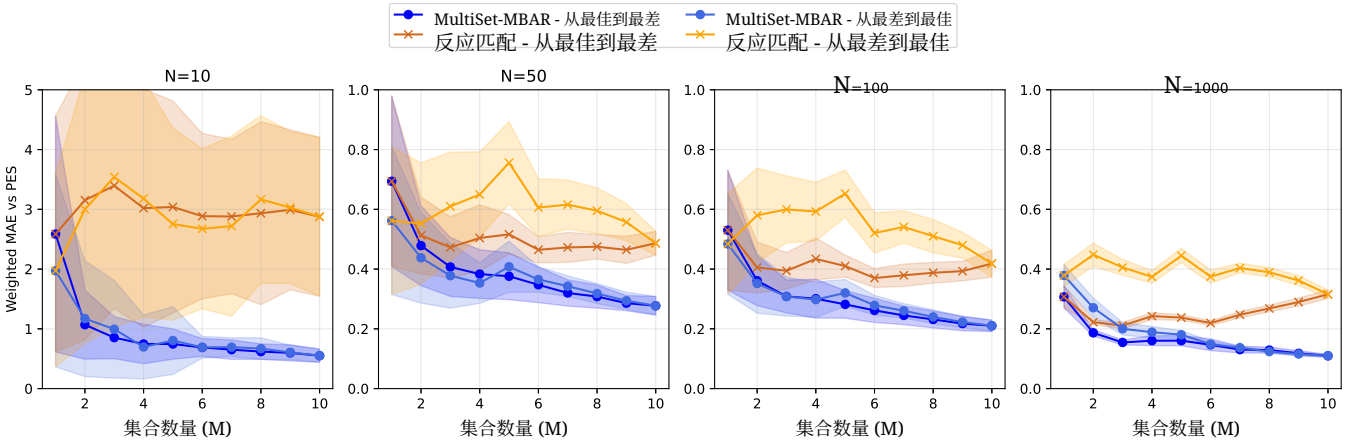

图4. 随集合数量增加，N=10、50、100、1000 条件下，RPE自由能与势能面之间的加权平均绝对误差（平均绝对误差），按每个分箱中的轨迹构型密度加权；通过对 N 条轨迹进行100次重复重采样的自助法，估算统计误差。

结果表明，纳入更多集合能够系统性地提高 RPE 的准确性，这体现在平均绝对误差不断降低。该结果凸显了 MultiSet-MBAR 方法整合多个 TIS 模拟信息、改进自由能景观的稳健性。相比之下，即使在样本量最大时，反应性匹配似乎仍未收敛。

这项分析补充了上一节关于穿越概率的讨论：两种观测量都取决于准确的路径权重估计，而它们的共同收敛进一步验证了该重加权方法的可靠性。

D. 基于 AIMMD-TIS 计算的主客体系统

为展示 MultiSet-MBAR 在简单模型系统之外的应用价值，我们将其应用于 AIMMD-TIS 研究[34]中的溶剂化主客体系统上的基于承诺概率的TIS。在这一迭代框架中，利用 AIMMD[35]（表示为 $q(x|\theta_{TPS})$）从 TPS射击数据中学习得到的初始承诺概率模型（定义为 $p_B(x) = 1/(1+e^{-q(x|\theta)})$ with$q(x|\theta)$ 神经网络模型）为首次 TIS计算定义等承诺概率界面，生成可通过 RPE 重新加权的路径系综。

随后使用得到的 RPE 数据训练改进后的承诺概率模型 $q(x|\theta_{RPE_1})$，该模型进而可为下一轮 TIS 迭代定义新的界面，如此循环往复。与标准方法不同，标准方法会因每轮迭代中的偏置不同而丢弃之前的路径数据；MultiSet-MBAR 则能够在统计一致的前提下，重复利用各轮迭代中的所有 TIS 路径系综。每个由不同承诺概率模型条件化的系综都被视为独立的偏置系综，并通过联合重新加权，得到共同的底层平衡路径分布。

图5展示了正向穿越概率 $P_A(q(x|\theta_{\mathrm{RPE_1}}) \mid q_A)$，使用不同的 RPE 构造方法计算得出：一种是利用 $q(x|\theta_{\mathrm{TPS}})$生成的 TIS 数据构造的 RPE，另一种是利用 $q(x|\theta_{\mathrm{RPE_1}})$生成的 TIS 数据构造的 RPE，还有一种是采用 MultiSet-MBAR 方法对两个数据集进行联合重加权后得到的合并 RPE。此外，我们比较了两种独立的缩放策略，即在重加权后合并 RPE。在第一种策略（称为反应性匹配）中，也用于双势阱结果的比较，MBAR 权重会经过缩放，使所有反应路径的权重均为 1。在第二种策略（称为通量匹配）中，权重会经过缩放，使通过第一个界面的通量与通过该界面的独立计算所得分子动力学通量相匹配。此处，q_A 定义为在受限于状态 A 的分子动力学模拟中采样到的$q(x|\theta_{\mathrm{RPE_1}})$的最大值，由此提供一个实用边界，超过该边界的构型即被视为处于该状态之外。

结果清楚地表明，仅进行第一次迭代是不够的，而第二次迭代则显著改善了穿越概率的估计值。使用 MultiSet-MBAR 合并两个数据集，可通过对所有可用路径系综进行联合重加权，进一步提升统计质量。相比之下，独立重标度策略高度依赖所选的匹配准则。反应性匹配得到的曲线更接近穿越概率较低的 RPE 所得曲线，因为相应的权重在合并后的估计值中占主导地位。通量匹配得到的平均穿越概率则更接近穿越概率较高的 RPE 所得结果，因为重标度后该数据集占主导地位。图5展示了通量匹配得到的平均穿越概率，但我们省略了其统计不确定性，因为重标度权重的变化幅度达到多个数量级，导致统计不确定性极大。

使用块大小10进行自助法分析后，我们发现，平均穿越

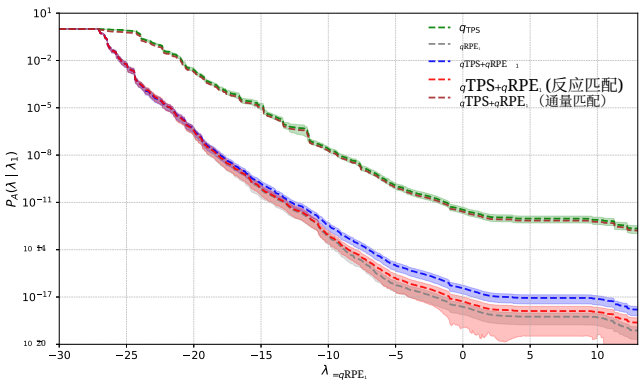

图5. 使用主客体系统的承诺概率模型 $q(x|\theta)$ 上连续 TIS 迭代所得的重加权路径系综，计算正向穿越概率 $P_A(q(x|\theta_{\mathrm{RPE}_1}) \mid q_A)$。将使用 $q(x|\theta_{\mathrm{TPS}})$（绿色）和 $q(x|\theta_{\mathrm{RPE}_1})$（灰色）的 TIS 模拟构建的独立重加权路径系综，与采用 MultiSet-MBAR（蓝色）进行联合重加权的结果进行了比较。图中还展示了基于反应性匹配（红色）和通量匹配（棕色）的独立重加权路径系综重缩放策略。由于通量匹配的统计不确定性极大，为清晰起见省略了其误差条，相对误差见正文。

概率的对数的相对误差为： $q(x|\theta_{\mathrm{TPS}})$ 数据集为 1.68%， $q(x|\theta_{\mathrm{RPE}_1})$ 为 1.66%，反应匹配组合为 2.36%，通量匹配组合为 14.18%，MultiSet-MBAR 合并的 RPE 为 1.47%。这些结果表明，MultiSet-MBAR 在避免独立重标度策略引入的主导偏差的同时，能够将统计不确定性降至最低。

IV. 结论

我们开发了 MultiSet-MBAR 方法，用于合并基于不同集体变量函数的 TIS 模拟。该方法能够在统一框架下重用采用不同界面定义生成的轨迹，从而得到更准确、更具代表性的重加权路径集合，并纳入来自多个集合的轨迹。对于（重加权）路径系综，MultiSet-MBAR 方法的准确性显著高于独立重缩放组合。与简单重缩放方法不同，后者实际上为每个界面系综指定独立的配分和，且各自具有未知的相对偏移；MultiSet-MBAR 方法则通过在重加权过程中同时纳入所有系综分布，自然地将这些配分和对齐到同一标度上。

我们的新方法能够促进 TIS 中所用界面函数的迭代优化，即使是基于神经网络承诺概率模型的界面函数也不例外，同时保留重用先前采样轨迹的能力。这使得一种高效、由反馈驱动的流程成为可能，在该流程中，界面函数的选择可以在连续的 TIS 迭代中不断优化，从而改善采样，并获得更可靠的机制性见解。

数据可用性声明

支持本研究结论的数据可在本文中获取。用于生成和分析数据的代码和脚本可在线获取。

附录 A：MBAR 方程的推导

1. 方程24和方程25的一致性检验

这里我们检验双集合 MBAR 程序与标准单集合 MBAR 的一致性。从方程24和方程25出发，若假设 $\mu = \lambda$ 和界面集合 $K_1 = K_2 \equiv K$，那么所得结果应与使用单个界面集合的 MBAR 相同，但此时 ˜ 遍历每 k 次模拟中的 $N_k + N_k$ 条路径：

$$\frac{\partial \ln L}{\partial w_{j,n}} = 0 = \frac{1}{w_{j,n}} - \sum_k^K \frac{(N_k + \tilde{N}_k)\hat{h}^k_{A,\lambda}[\mathbf{x}_{j,n}]}{\mathcal{Z}^k_{A,\lambda}} \tag{A1}$$

$$\frac{\partial \ln L}{\partial \tilde{w}_{j,n}} = 0 = \frac{1}{\tilde{w}_{j,n}} - \sum_k^K \frac{(N_k + \tilde{N}_k)\hat{h}^k_{A,\lambda}[\tilde{\mathbf{x}}_{j,n}]}{\mathcal{Z}^k_{A,\lambda}}, \tag{A2}$$

这与原始的单集合 MBAR 对数˜似然表达式（方程 5）相同，但现在使用了 $N_k + N_k$ 条路径。这表明，在集合相同的极限情况下，两集合 MBAR 程序会得出一致的结果。

表 I. MBAR/RPE 推导中使用的符号。

符号	含义
$\mathbf{x}$	路径空间中的轨迹（路径）
$P_0[\mathbf{x}]$	无偏路径分布
$\hat{h}^k_{A,\lambda^{(i)}}[\mathbf{x}]$	指示函数：$\mathbf{x}$ 穿过集合 i 的界面 k leav-入 A
$\lambda^{(i)}$	集合 i 的界面（集体变量）函数，包含界面-面 $\{\lambda^{(i)}_k\}_{k=0}^{K^{(i)}}$
$K^{(i)}$	集合 i 中的界面数量
$N^{(i)}_k$	界面系综中的采样路径数量-合 k，属于集合 i
$\mathcal{Z}$	全局归一化常数
$\mathcal{Z}^k_{A,\lambda^{(i)}}$	界面 k in 的条件配分函数和集合 i （离开 A）
$w^A[\mathbf{x}]$	分配给离开$\mathbf{x}$ 的路径的 MBAR 权重状态 A
$K^{(i)}_{\max}[\mathbf{x}]$	i 集合中由 $\mathbf{x}$ 穿越的最高界面
$P_A(\lambda \mid \lambda_1)$	以参考界面为条件的穿越概率-界面 λ_1
c_A, c_B	用于匹配正向/反向反应直方图的常数和通量

2. MultiSet-MBAR 的推导

本节将第二节 B 部分针对两个集合给出的推导推广到 M 个界面系综集合。考虑 M 界面函数$\lambda^{(1)},\ldots,\lambda^{(M)}$，其对应于 $i\in 1,\ldots,M$，每个函数都有一个 $N_k^{(i)}$，对应于 $k\in 1,\ldots K^{(i)}$，其中 $K^{(i)}$是集合 i 中的界面数量，而 $N_k^{(i)}$ 表示界面系综 k 中的路径数量（在 $\lambda^{(i)}$ 中）。每个集合 i 还具有相应的条件配分函数集合$\mathcal{Z}_{A,\lambda^{(i)}}^k$，分别对应离开 A 和 $\mathcal{Z}_{B,\lambda^{(i)}}^k$ 的路径。

我们假设，对于每个集合，稳定态都相同（对于所有 $i\in 1,\ldots,M$ 均为 $\lambda_A\equiv\lambda_0^{(i)}$）。对于每个界面系综 i, k，我们得到一个包含 $N_k^{(i)}$ 条路径的集合 $\{\mathbf{x}_n\}_k^{(i)}$，其中 $\mathbf{x}_{k,n}^{(i)}$ 是界面函数 $\lambda^{(i)}$的系综 k 中第 n 条路径。

我们可以将这些 M 集合合并，得到在无偏系综中观测到它们的概率:

$$\mathcal{P}_A[\mathbf{x}]=\frac{1}{\mathcal{Z}}\left(\sum_i^M\sum_k^{K^{(i)}}\sum_n^{N_k^{(i)}}w_{k,n}^{(i)}\delta[\mathbf{x}-\mathbf{x}_{k,n}^{(i)}]\right),\tag{A3}$$

其中因子

$$\mathcal{Z}_A=\sum_i^M\sum_k^{K^{(i)}}\sum_n^{N_k^{(i)}}w_{k,n}^{(i)},\tag{A4}$$

其中权重 $w_{k,n}^{(i)}$未知。

如果从理论角度考虑我们最初用于获得条件分布的（TIS）系综，那么我们可以从平衡路径分布 $\mathcal{P}[\mathbf{x}]$中抽取所有路径，此时

$$\mathcal{P}_{A,\lambda^{(i)}}^{k'}[\mathbf{x}]=\frac{1}{Z_{A,\lambda^{(i)}}^{k'}}\int\mathcal{D}\mathbf{x}'\hat{h}_{A,\lambda^{(i)}}^{k'}[\mathbf{x}']\mathcal{P}[\mathbf{x}']\delta[\mathbf{x}-\mathbf{x}']\tag{A5}$$

其中$Z_{A,\lambda^{(i)}}^{k'}=\int\mathcal{D}_{\mathbf{x}'}\hat{h}_{A,\lambda^{(i)}}^{k'}[\mathbf{x}']\mathcal{P}[\mathbf{x}']$ 是理论条件配分函数。

在离散情形下，满足偏置函数 $\hat{h}_{A\lambda^{(i)}}^k[\mathbf{x}]$的路径分布为:

$$\mathcal{P}_{A,\lambda^{(m)}}^{k'}[\mathbf{x}]=\frac{1}{\mathcal{Z}_{A,\lambda^{(m)}}^{k'}}\sum_i^M\sum_k^{K^{(i)}}\sum_n^{N_k^{(i)}}\hat{h}_{A,\lambda^{(m)}}^{k'}[\mathbf{x}]\mathcal{P}[\mathbf{x}]\delta[\mathbf{x}-\mathbf{x}_{k,n}]\tag{A6}$$

$$=\frac{1}{\mathcal{Z}_{A,\lambda^{(m)}}^{k'}}\sum_i^M\sum_k^{K^{(i)}}\sum_n^{N_k^{(i)}}\hat{h}_{A,\lambda^{(m)}}^{k'}[\mathbf{x}]w_{k,n}^{(i)}\delta[\mathbf{x}-\mathbf{x}_{k,n}^{(i)}],\tag{A7}$$

其中$\mathcal{Z}_{A,\lambda^{(m)}}^{k'}=\sum_i^M\sum_k^{K^{(i)}}\sum_n^{N_k^{(i)}}\hat{h}_{A,\lambda^{(m)}}^{k'}[\mathbf{x}]w_{k,n}^{(i)}$。

路径 $\{\mathbf{x}\}_k^{(i)}$ 从分布$\mathcal{P}_{A\lambda^{(i)}}^k[\mathbf{x}]$中采样，使得路径 $\mathbf{x}_{k,n}^{(i)}$ 其由第 k 个界面系综抽取，且界面函数为 $\lambda^{(i)}$ 的路径似然为

$$L_{k,\lambda^{(i)}}(\mathbf{x}_{k,n}^{(i)}|\{w_{k,n}\}^1,\ldots,\{w_{k,n}\}^M)=\mathcal{P}_{A\lambda^{(i)}}^k[\mathbf{x}_{k,n}^{(i)}].\tag{A8}$$

M 个集合的观测数据的对数似然因此为

$$\begin{aligned}\ln L&=\ln\prod_i^M\prod_k^{K^{(i)}}\prod_n^{N_k^{(i)}}\mathcal{P}_{A,\lambda^{(i)}}^k[\mathbf{x}_{k,n}^{(i)}]\\&=\sum_i^M\sum_k^{K^{(i)}}\sum_n^{N_k^{(i)}}\ln\mathcal{P}_{A,\lambda^{(i)}}^k[\mathbf{x}_{k,n}^{(i)}].\end{aligned}\tag{A9}$$

将分布代入可得

$$\ln L=\sum_i^M\sum_k^{K^{(i)}}\sum_n^{N_k^{(i)}}(\ln w_{k,n}^{(i)}\hat{h}_{A,\lambda^{(i)}}^k[\mathbf{x}_{k,n}]-\ln Z_{A,\lambda^{(i)}}^k).\tag{A10}$$

对 $w_{j\,n}^{(m)}$求最大值

$$\frac{\partial\ln L}{\partial w_{j,n}^{(m)}}=0=\frac{1}{w_{j,n}^{(m)}}-\sum_{i=1}^M\sum_k^{K^{(i)}}\frac{N_k^{(i)}\hat{h}_{A,\lambda^{(i)}}^k[\mathbf{x}_{j,n}^{(m)}]}{\mathcal{Z}_{A,\lambda^{(i)}}^k}\tag{A11}$$

$$\frac{w_{j,n}^{(m)}}{\mathcal{Z}}=\left[\sum_{i=1}^M\sum_k^{K^{(i)}}\frac{N_k^{(i)}\hat{h}_{A,\lambda^{(i)}}^k[\mathbf{x}_{j,n}^{(m)}]}{\mathcal{Z}_{A,\lambda^{(i)}}^k/\mathcal{Z}}\right]^{-1}\tag{A12}$$

$$\frac{\mathcal{Z}_{A,\lambda^{(i)}}^{k'}}{\mathcal{Z}}=\sum_{i=1}^M\sum_j^{K^{(i)}}\sum_n^{N_j^{(i)}}\hat{h}_{A,\lambda^{(i)}}^{k'}[\mathbf{x}_{j,n}^{(i)}]\left[\sum_{i=1}^M\sum_k^{K^{(i)}}\frac{N_k^{(i)}\hat{h}_{A,\lambda^{(i)}}^k[\mathbf{x}_{j,n}^{(i)}]}{\mathcal{Z}_{A,\lambda^{(i)}}^k/\mathcal{Z}}\right]^{-1},\tag{A13}$$

可通过迭代求解 $\sum_{i=1}^M K^{(i)}$ 分区项来求解。路径 $\mathbf{x}$ 获得一个权重

$$w^A[\mathbf{x}]=\left[\sum_{i=1}^M\sum_k^{K^{(i)}}\frac{N_k^{(i)}\hat{h}_{A,\lambda^{(i)}}^k[\mathbf{x}]}{\mathcal{Z}_{A,\lambda^{(i)}}^k/\mathcal{Z}}\right]^{-1},\tag{A14}$$

它仅基于路径 $\mathbf{x}$ 在每个 M 单调递增的 λ 函数中的最大值（因为总和中所有后续项对于 $k>K_{\max}^{(i)}[\mathbf{x}]$ 都有 $h_{A\lambda^{(i)}}^k=0$ for）。因此

$$w^A[\mathbf{x}]=\left[\sum_{i=1}^M\sum_k^{K_{\max}^{(i)}[\mathbf{x}]}\frac{N_k^{(i)}}{\mathcal{Z}_{A,\lambda^{(i)}}^k/\mathcal{Z}}\right]^{-1},\tag{A15}$$

对于所有路径，无论其最初是从哪个分布中采样得到的。

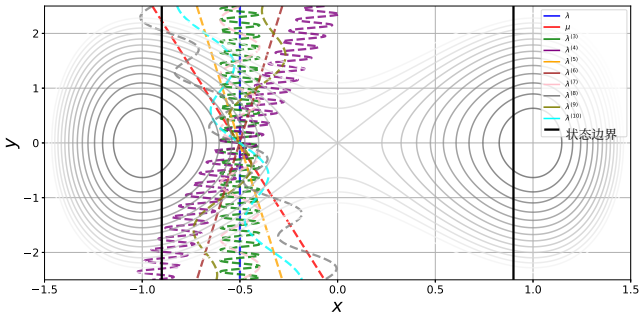

图6. 双势阱势能，其中示意了 10 个不同界面函数各自对应的界面。

附录B：界面函数与偏差评分

为测试 MultiSet-MBAR程序，我们定义了一族界面函数 $\lambda(x,\ y)$，其中界面的形状在空间中经过调制，并由旋转角 θ、正弦频率 f 和振荡幅度 A_{mag} 参数化。其一般形式见式 36。我们使用以下评分函数评估每个界面偏离沿 x 轴对齐的平坦界面的程度：

$$\mathrm{score}(\theta, f, A_{\mathrm{mag}}) = \begin{cases} |\theta| + b \cdot A_{\mathrm{mag}}, & \text{if } f = 0 \\ |\theta| + a \cdot \left(\frac{1}{f}\right) + b \cdot A_{\mathrm{mag}}, & \text{if } f \neq 0 \end{cases} \tag{B1}$$

它会惩罚旋转，并使用 $a = 2$ 和 $b = 100$ 作为加权参数，分别惩罚低频和较大的振荡幅度。

表 II 列出了所考虑的十个界面函数、其参数值及计算得到的分数。函数按偏差分数递增排列。前两个分别记为 λ 和 μ，因为它们已用于双集合方法；其余函数标记为 $\lambda^{(i)}$ 至 $i \geq 3$。图6展示了这十种不同界面函数各自的一个界面。

符号	θ [度]	f [赫兹]	A_{mag}	评分
λ	0.0	0.0	0.0	0.00
μ	5.0	0.0	0.0	5.00
$\lambda^{(3)}$	-5.0	0.0	0.0	5.00
$\lambda^{(4)}$	10.0	0.0	0.0	10.00
$\lambda^{(5)}$	0.0	5.0	0.1	10.40
$\lambda^{(6)}$	0.0	1.0	0.1	12.00
$\lambda^{(7)}$	-5.0	0.5	0.1	19.00
$\lambda^{(8)}$	5.0	0.5	0.1	19.00
$\lambda^{(9)}$	-10.0	5.0	0.1	20.40
$\lambda^{(10)}$	10.0	1.0	0.1	22.00

表 II. 界面函数的定义及其偏差分数。

[1]C. Dellago、P. G. Bolhuis、F. S. Csajka 和 D. Chandler，J. Chem. Phys.**108**，1964（1998）。[2]J. H´enin、T. Leli`evre、M. R.Shirts、O. Valsson 和 L. Dele- motte，Living J. Comput. Mol. Sci. **4** （2022），10.33011/live- coms.4.1.1583。

[3]P. G. Bolhuis, D. Chandler, C. Dellago, and P. Geissler, Ann. Rev. Phys. Chem. 53, 291 (2002).[4]C. Dellago, P. G. Bolhuis, and P. L. Geissler, Adv. Chem. Phys. 123, 1 (2002).[5]C. Dellago and P. G. Bolhuis, Adv Polym Sci 221, 167 (2009).[6]P. G. Bolhuis and C. Dellago, in *Reviews of ComputationalChemistry* (Wiley-VCH, 霍博肯, 2009).[7]P. G. Bolhuis and C. Dellago, Eur. Phys. J. Spec. Top. 224, 2409 (2015).[8]D. W. H. Swenson, J.-H. Prinz, F. Noe, J. D. Chodera, and P. G. Bolhuis, J. Chem. Theor. Comp. 15, 813 (2019).[9]P. G. Bolhuis and D. W. H. Swenson, Adc. Theor. Simul. 4, 2000237 (2021).[10]T. S. van Erp, D. Moroni, and P. G. Bolhuis, J. Chem. Phys. 118, 7762 (2003).[11]T. S. van Erp, Adv. Chem. Phys. 151, 27 (2012).[12]R. Cabriolu, K. M. S. Refsnes, P. G. Bolhuis, and T. S. van Erp, J. Chem. Phys. 147, 152722 (2017).[13]T. S. van Erp and P. G. Bolhuis, J. Comput. Phys. 205, 157 (2005).[14]T. S. van Erp, Adv. Chem. Phys. 151, 27 (2012).[15]T. van Erp, Phys. Rev. Lett. 98, 268301 (2007).[16]P. G. Bolhuis, J. Chem. Phys. 129, 114108 (2008).[17]W.-N. Du and P. G. Bolhuis, J. Chem. Phys. 139, 044105 (2013).[18]J. Rogal, W. Lechner, J. Juraszek, B. Ensing, and P. G. Bolhuis, J. Chem. Phys. 133, 174109 (2010).[19]G. M. Torrieand J. P. Valleau, Chem. Phys. Lett. 28, 578 (1974).[20]A. Laio and M. Parrinello, Proc. Nat. Acad. Sci. USA 99, 12562 (2002).[21]A. M. Ferrenberg and R. H. Swendsen, Phys. Rev. Let. 63, 1195 (1989).[22]S. Kumar, J. M. Rosenberg, D. Bouzida, R. H. Swendsen, and P. A. Kollman, J. Comput. Chem. 13, 1011–1021 (1992).[23]W. Lechner, J. Rogal, J. Juraszek, B. Ensing, and P. G. Bolhuis, J. Chem. Phys. 133, 174110 (2010).[24]W. Lechner, C. Dellago, and P. G. Bolhuis, J. Chem. Phys. 135 (2011), 10.1063/1.3651367.[25]G. D´ıaz Leines, R. Drautz, and J. Rogal, J. Chem. Phys. 146 (2017), 10.1063/1.4980082.[26]Y. Liang, G. D´ıaz Leines, R. Drautz, and J. Rogal, J. Chem. Phys. 152 (2020), 10.1063/5.0010074.[27]A. Arjun and P. G. Bolhuis, J. Phys. Chem. B 124, 8099–8109 (2020).[28]S. Safaei, L. Baldauf, T. S. van Erp, and A. Ghysels, J. Phys. Chem. B 129, 10019–10034 (2025).[29]W. Du and P. G. Bolhuis, J. Chem. Phys. 140, 195102 (2014).[30]D. T. Zhang, L. Baldauf, S. Roet, A. Lervik, and T. S. van Erp, Proc. Natl. Acad. Sci 121 (2024), 10.1073/pnas.2318731121.[31]A. C. Newton, J. Groenewold, W. K. Kegel, and P. G. Bolhuis, Proc. Nat. Acad. Sci. USA 112, 15308 (2015).[32]A. C. Newton, R. Kools, D. W. H. Swenson, and P. G. Bolhuis, J. Chem. Phys. 147 (2017), 10.1063/1.5006485.[33]T. S. van Erp, M. Moqadam, E. Riccardi, and A. Lervik, J. Chem. Theor. Comput. 12, 5398–5410 (2016).[34] R. S. Breebaart, G. Lazzeri, R. Covino, and P. G. Bolhuis, (2025), arXiv:2507.04052 [physics.chem-ph].[35]H. Jung, R. Covino, A. Arjun, C. Leitold, C. Dellago, P. G. Bolhuis, and G. Hummer, Nat. Comput. Sci. 3, 334–345 (2023).[36]G. Lazzeri, H. Jung, P. G. Bolhuis, and R. Covino, J. Chem. Theor. Comput. 19, 9060–9076 (2023).[37]M. R. Shirts and J. D. Chodera, J. Chem. Phys. 129 (2008), 10.1063/1.2978177.[38]C. Bartels, Chem. Phys. Lett. 331, 446–454 (2000).[39]D. Frenkel and B. Smit, *Understanding Molecular Simulation: From algorithms to applications* (Elsevier, 阿姆斯特丹, Netherlands, 2023).[40]T. S. Van Erp, Phys. Rev. Let. 98, 268301 (2007).[41] P. G. Bolhuis, J. Chem. Phys. 129, 114108 (2008).[42]B. Lei mkuhler and C. Matthews, Appl. Math. Res. X , 34–56 (2012).