arXiv:2303.02115v1 [cs.LG] 3 Mar 2023

自然的代价函数：通过最小化作用量来仿真物理学

Tim StrangML Collective 马萨诸塞州霍利斯顿，01746 timdstrang@gmail.com

Isabella Caruso麻省理工学院 材料科学 马萨诸塞州剑桥，02142 icaruso@mit.edu

Sam GreydanusML Collective 俄勒冈州科瓦利斯，97330 greydanus.17@gmail.com

摘要

在物理学中，有一种称为作用量的标量函数，其作用类似于代价函数。将其最小化后，就会得到“最小作用量路径”，它代表物理系统在空间和时间中所经历的路径。这一函数在理论物理学中至关重要，通常通过解析方法将其最小化，从而求得各种问题的运动方程。本文提出一种不同的方法：我们不通过解析方法最小化作用量，而是将其离散化，然后直接采用梯度下降对其进行最小化。我们用这种方法求得六种不同物理系统的动力学，并表明结果与真实动力学几乎完全一致。我们还探讨了无约束能量效应等失效模式，并展示如何应对这些问题。最后，我们利用离散化的作用量构建了一种简单但新颖的量子模拟。代码：github.com/greydanus/ncf

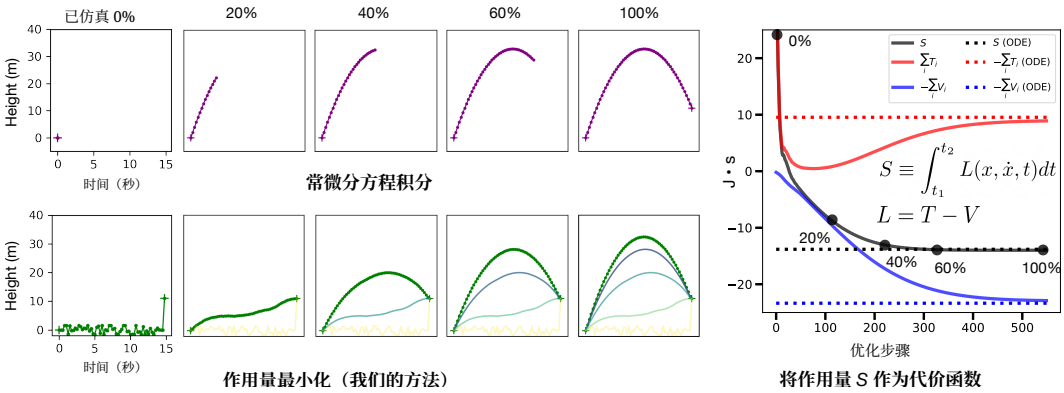

图1：使用梯度下降寻找最小作用量路径。**左侧**：我们将常微分方程积分这一常规方法与作用量最小化方法进行比较。作用量是路径上每个点的动能 T减去势能 V后求和。我们计算这一量关于路径坐标的梯度，然后使初始路径（黄色）变形为最小作用量路径（绿色）。最终得到的路径是一条抛物线，与通过常微分方程积分得到的路径相符。**右侧**：我们绘制了优化过程中路径的作用量 S、动能 T和势能 V的变化。三种量都渐近趋近于常微分方程轨迹的 S、 T和 V值。

1 引言

在过去两个世纪中，许多著作研究了各种物理和数学背景下的作用量（Goldstine，2012；Ferguson，2004）。在所有这些研究中，作者都以解析方式处理作用量。他们只能通过间接方式求解最小作用量路径：首先应用欧拉-拉格朗日方程，然后求解一组微分方程，最后使用常微分方程（ODE）求解器对这些方程进行时间积分。

在本文中，我们采用一种不同的方法：

我们不通过解析方法求解驻定作用量的路径，而是将作用量函数离散化，然后直接使用梯度下降将其最小化。

我们运用这种方法，为六种不同的物理系统推导其动力学：自由落体中的物体、一个摆、一个双摆、三体问题、具有 Lennard-Jones 势的气体，以及内太阳系的行星历表。我们记录了一种称为无约束能量效应 的失效模式，并讨论如何减轻这种效应。最后，我们指出，在量子力学中，作用量控制着路径的共振模态，并使用其离散化版本构建了一个简单的量子模拟。

2 背景

拉格朗日方法。 拉格朗日方法首先考虑物理系统从初始状态$\mathbf{x}(t_0)$到最终状态 $\mathbf{x}(t_1)$可能经过的所有路径。随后，它给出一条简单规则，用于选出自然实际采取的路径 $\hat{\mathbf{x}}$： 作用量 S（定义见公式 1）沿该路径必须取驻值。此处， T和 V分别是系统在任意给定时间 t 处的动能函数和势能函数，其中 $[t_0, t_1]$。

$$S := \int_{t_0}^{t_1} L(\mathbf{x},\, \dot{\mathbf{x}},\, t)\, dt \quad \text{where} \quad L = T - V \tag{1}$$

$$\hat{\mathbf{x}} \text{ has the property } \quad \frac{d}{dt}\left(\frac{\partial L}{\partial \dot{\hat{x}}(t)}\right) = \frac{\partial L}{\partial \hat{x}(t)} \quad \text{for} \quad t \in [t_0, t_1] \tag{2}$$

求出$\hat{\mathbf{x}}$ 与欧拉-拉格朗日方程。 当 S 处于驻定状态时，我们可以证明，在区间 $[t_0, t_1]$ 内欧拉-拉格朗日方程（方程 2）成立（Morin，2008年）。这一观察很有价值，因为它使我们能够求出$\hat{\mathbf{x}}$：首先，我们将欧拉-拉格朗日方程应用于拉格朗日量 L，并推导出一个偏微分方程组。然后，对这些方程积分，得到$\hat{\mathbf{x}}$。重要的是，这种方法适用于涵盖经典力学、电动力学、热力学和相对论的所有问题。它为整体研究经典物理学提供了一个连贯的理论框架。

求出$\hat{\mathbf{x}}$ 与作用量最小化（本文工作）。 一种更直接的求出$\hat{\mathbf{x}}$ 的方法，始于这样一个认识：驻定作用量路径几乎总是同时也是最小作用量路径（Morin，2008年）。因此，在不失一般性的前提下，我们可以用方程 3 第三部分所示的简单最小化目标，替代欧拉-拉格朗日方程。同时，如方程 3 第一部分所示，我们可以将 S 重新定义为对 N 个等间隔时间片的离散求和：

$$S := \sum_{i=0}^{N} L(\mathbf{x},\, \dot{\mathbf{x}},\, t_i)\Delta t \quad \text{where} \quad \dot{\mathbf{x}}(t_i) := \frac{\mathbf{x}(t_{i+1}) - \mathbf{x}(t_i)}{\Delta t} \quad \text{and} \quad \hat{\mathbf{x}} := \arg\min_{\mathbf{x}} S(\mathbf{x}) \tag{3}$$

还存在一个问题：对 $\mathbf{x}$ 进行离散化后，我们无法再对其求导以获得 $\dot{\mathbf{x}}(t_i)$ 的精确值。相反，我们必须使用公式 3 第二部分所示的有限差分近似。当然，对于求和中的最后一个 $\dot{\mathbf{x}}$，这种近似将无法进行，因为$\dot{\mathbf{x}}_{N+1}$ 不存在。对于这个值，我们假设当 N 较大时，在时间区间 Δt 内速度变化很小，因此令 $\dot{\mathbf{x}}_N = \dot{\mathbf{x}}_{N-1}$。完成这最后一个近似后，我们现在可以数值计算梯度 $\frac{\partial S}{\partial \mathbf{x}}$，并利用它来最小化 S。这可以通过 PyTorch（Paszke 等，2019）或任何支持自动微分的软件包来实现。

相关工作：最小化昂萨格-马赫卢普作用量。一类相关工作涉及最小化昂萨格-马赫卢普（OM）函数的作用量（Onsager & Machlup，1953年；Faccioli 等人，2006年）。该函数用于描述随机过程概率密度随时间的动力学变化。OM 函数的作用量与拉格朗日量的作用量具有不同的物理解释，但共享某些数学性质。Adib（2008年）对扩散轨迹进行离散化并计算其梯度（公式 22），但并未最小化 OM作用量。相反，Lee 等人（2017）；Zuckerman & Woolf（2000年）最小化 OM 作用量，以分析蛋白质折叠动力学，但使用无梯度优化方法：Metropolis 马尔可夫链蒙特卡罗和构象空间退火（分别参见 Metropolis 等人，1953年；Lee 等人，1997）。与我们的工作类似，这些研究固定初始状态和最终状态，并（有时）在该区间内最小化作用量。与我们的工作不同，它们使用 OM 函数，且不借助自动微分或梯度下降进行最小化。我们在附录 A.1 中讨论了物理学问题基于梯度的优化方面的其他相关工作。

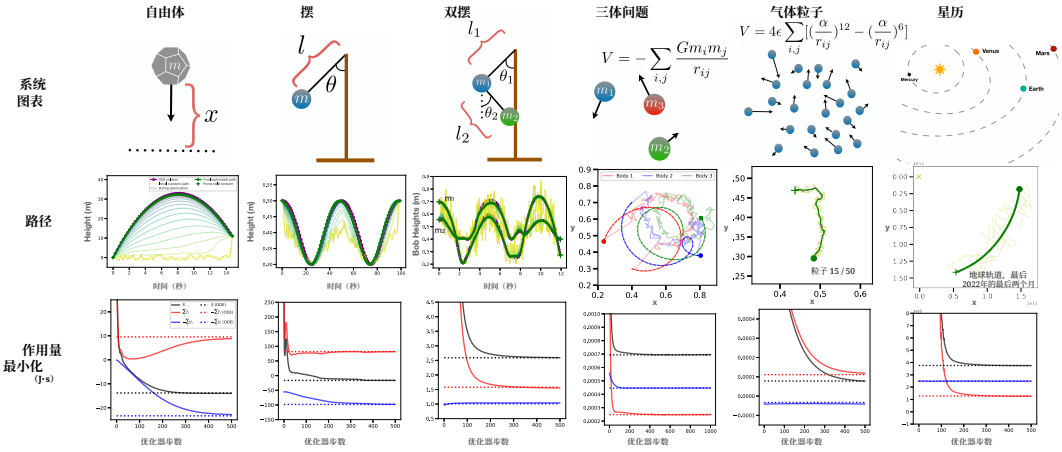

图 2：展示我们的方法应用于各种不同物理模拟的效果。**第一行**：展示所研究的六个系统的设置和变量图表。**第二行**：展示常微分方程路径（紫色）、初始半随机路径（黄色）以及作用量最小化后的最终路径（绿色）。这些图有助于直观理解路径在优化过程中的变化。**第三行**：以作用量、动能和势能的形式展示优化动力学。

3 六项实验

我们所描述的这种作用量最小化方法尚未得到详细研究。我们讨论过几项在 OM 函数背景下采用这种方法的相关工作，但它们的研究范围和具体细节都与本文以及彼此之间存在差异。因此，在实验中，我们优先考虑简洁性。除非另有说明，我们将质量、重力等所有常数都设为 1。在选择物理系统时，我们先从两个玩具问题（用于调试）入手：自由体和摆。随后，我们研究了四种更复杂的系统：双摆，

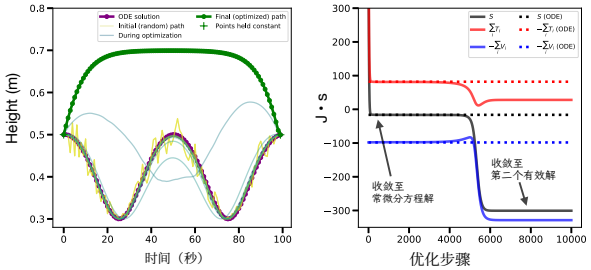

图 3：摆任务中的无约束能量效应。优化最初收敛到基线路径。经过更多优化器步数后，它会收敛到另一个同样有效的解：摆向上摆动、垂直悬挂，然后落入最终状态。这是因为我们当前的方法没有像常微分方程初始状态（$\mathbf{x}(t_0)$ $\dot{\mathbf{x}}(t_0)$）那样约束总能量$T+V$。

三体问题、简单气体以及真实的行星运动星历数据集。这些系统都具有非线性、混沌性和高维性，因此构成了有趣的挑战[1]。在每种情况下，我们都将结果与使用欧拉积分、通过简单常微分方程求解器获得的基准路径进行了比较。附录 A.2 提供了更多细节，包括拉格朗日量和运动方程。

无约束能量效应。 在实验早期，我们遇到了无约束能量效应。当优化器收敛到一条有效的物理路径，而其总能量与基准值不同时，就会发生这种情况。图 3 展示了一个示例。之所以会发生这种情况，是因为尽管我们固定了初始值和最终状态，却没有约束路径的总能量 $T+V$。尽管图 3 中的路径不一定无效，但它们使我们难以恢复基准常微分方程路径。因此，我们使用基准常微分方程路径来初始化路径，向其中加入高斯噪声，然后使用提前停止来筛选出与常微分方程基准路径相似（通常完全相同）的路径。这种方法符合“变分法”的数学设定，即研究真最小作用量路径附近的扰动路径。我们注意到，还有其他方法可以缓解这一效应，而不需要由常微分方程生成的初始路径。我们将在附录 A.2 中讨论这些方法，因为它们超出了本文的主要范围。

结果。 在全部六种物理系统中，我们得到的最小作用量路径与常微分方程求解器生成的基准路径几乎完全一致。图 2 展示了优化动力学和定性结果，表 2 则展示了定量结果。这些结果表明，即使对于双摆这样具有混沌特性且强耦合的系统，作用量最小化也能生成物理有效动力学

[1]例如，简单气体的状态有一百个自由度。

以及三体问题。我们注意到一个有趣的规律：优化动力学主要受动能项支配 T（图 2 第三行）。这是因为 S 对 T 的敏感度往往更高（其随 $\dot{\mathbf{x}}^2$ 增长），而非 V。未来的方法应着重于稳定 T。

4 量子极限

拉格朗日在经典力学背景下提出的作用量方程，是少数几条在量子语境中无需重新定义的方程之一。在这里，系统用波函数描述，并允许同时取多条路径的叠加态（线性组合）。这一现象最著名的例子是著名的双缝实验：一个电子射向两条分开的狭缝，会同时穿过两条狭缝，并在另一侧与自身发生干涉。这种干涉的发生，是因为相位沿每条路径以不同方式演化，因此路径重新汇合时彼此不同步。公式4中的指数部分描述了系统沿给定路径运动时如何演化：它以与路径作用量成正比的速率振荡。因此，波函数的相位在沿着……运动后

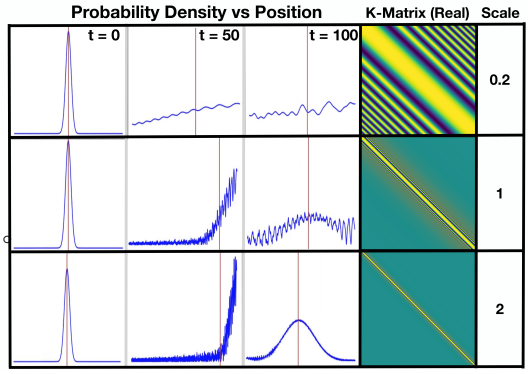

图4：在不同尺度（$\hbar \propto$ 尺度$^{-1}$）下可视化波包动力学。随着 $\hbar \to 0$，波能够更长时间保持相干性，并以更高频率与右侧墙面发生干涉，在尺度 = 2下趋近于经典粒子的行为。相干性随着相位-作用量矩阵 K非对角线上的振荡频率增加而增强。

在给定时间区间内的一条路径，其贡献也与作用量成正比。公式 4 的其余部分描述了经过一段时间后如何重新组合所有路径的贡献。首先，它根据每条路径对应的相位变化旋转波函数，然后将这些状态求和（积分），以获得该时间区间结束时的完整状态。在双缝实验中，这个求和会将来自两条狭缝的波重新组合成传感器处采样的完整波函数 $\mathbf{\Psi(x,\ t)}$。

$$\Psi(x,t) = \int_{\mathbf{x(0)}=\mathbf{x}} \Psi(\mathbf{x(t)},\mathbf{0}) * \mathbf{exp}\{\frac{\mathbf{i}}{\hbar}\int \mathbf{L}[\mathbf{x}(\tau),\ \dot{\mathbf{x}}(\tau),\tau]\mathbf{d}\tau\}\ \mathcal{D}\mathbf{x} \tag{4}$$

离散化路径积分。费曼等人（1965）以及许多其他研究者对公式 4 进行了解析处理，并由此推导出薛定谔方程和经典最小作用量原理。然而，与经典作用量的情况一样，很少有研究将路径离散化，并利用作用量构建数值仿真。这仍然是我们的目标。我们可以将公式 4 中的空间和时间维度划分为小而均匀的区间。然后，对于每个时间区间，我们可以将路径集合限制为仅包含连接各个 N 离散空间位置的 N^2 直线路径。我们可以计算这些路径的作用量，并借助公式 4 计算它们的相位变化。随后，我们可以将这些相位变化整理成一个 $N \times N$ 相位-作用量矩阵 K：行对应路径起点的位置，列对应路径终点的位置，而矩阵元素本身则表示路径的相位变化。该矩阵绘于图4中，直观地展示了在量子-经典极限下，彼此相邻的路径如何相互干涉。

一个简单的量子模拟。现在，要模拟一个离散时间片内的动力学，我们只需将波函数（由 N 个空间位置上的复振幅构成的离散向量）乘以 K。然后，通过反复乘以 K，就可以将其演化到任意遥远的未来（见图4）。我们得到的量子模拟与更为常见的基于哈密顿量的模拟（Hartree & Hartree, 1935；Richings 等人，2015）有着根本区别，其简洁性令人瞩目。它映照出我们在六个经典实验中采取的哲学方法：在这两种情况下，我们都尽可能直接地用作用量来表达计算。

5 结论

基本的物理原理及其背后的优雅推理，往往会被数值近似和领域特定记号所掩盖。在这项工作中，我们试图通过尽可能直接地用作用量来表达计算，避免这种情况发生。

致谢

全体作者感谢 Jason Yosinski、Rosanne Liu 以及 MLCollective 研究组的其他成员，就实验提供了有益的讨论、想法和反馈。

TS 感谢 Central Rock Gym 提供了适宜的环境，鼓励人们进行学术思考。IC 感谢麻省理工学院材料科学系，无意间让她在等待论文申请反馈期间得以投入这个副项目。SG 感谢 TS 和 IC 与他共同踏上这段研究之旅；这条路既不是最短的，也不是作用量最小的。

参考文献

Artur B Adib。扩散动力学的随机作用量：重加权、采样与最小化。《物理化学杂志*B*》，112(19):5910–5916，2008年。

Berni J Alder 和 Thomas Everett Wainwright。分子动力学研究。一、一般方法。《化学物理学杂志》，31(2):459–466，1959年。

Yohai Bar-Sinai、Stephan Hoyer、Jason Hickey 和 Michael P Brenner。为偏微分方程学习数据驱动的离散化方法。《美国国家科学院院刊》，116(31):15344–15349，2019。

Ricky TQ Chen、Yulia Rubanova、Jesse Bettencourt 和 David K Duvenaud。神经常微分方程。《神经信息处理系统进展》，31，2018。

Miles Cranmer、Sam Greydanus、Stephan Hoyer、Peter Battaglia、David Spergel 和 Shirley Ho。拉格朗日神经网络。*arXiv*预印本 *arXiv:2003.04630*，2020。

P Faccioli、M Sega、F Pederiva 和 H Orland。蛋白质折叠中的主导路径。《物理评论快报》，97(10):108101，2006年。

James Ferguson。变分法史及其应用概述。*arXiv* 预印本 *math/0402357*，2004年。

Richard P Feynman、Robert B Leighton 和 Matthew Sands。《费曼物理学讲义》；第 II 卷。《美国物理学杂志》，33(9):750–752，1965。

C Daniel Freeman、ErikFrey、Anton Raichuk、Sertan Girgin、Igor Mordatch 和 Olivier Bachem。Brax——一款用于大规模刚体模拟的可微分物理引擎。*arXiv* 预印本 *arXiv:2106.13281*，2021。

Herman Heine Goldstine。《从*17*世纪到*19*世纪的变分法史》，第5卷。Springer Science & Business Media，2012。

Douglas Rayner Hartree 和 William Hartree。铍的含交换作用自洽场。《伦敦皇家学会会刊。*A*辑：数学与物理科学》，150 (869):9–33，1935。

Scott A Hollingsworth 和 Ron O Dror. 面向所有人的分子动力学模拟。*Neuron*, 99(6): 1129–1143, 2018.

Diederik P Kingma 和 Jimmy Ba. Adam：一种随机优化方法。*arXiv* 预印本 *arXiv:1412.6980*, 2014.

Jooyoung Lee、Harold A Scheraga 和 Shelly Rackovsky. 多肽构象能计算的新优化方法：构象空间退火。《计算化学杂志》, 18(9):1222–1232, 1997.

Juyong Lee、In-Ho Lee、InSuk Joung、Jooyoung Lee 和 Bernard R Brooks. 通过作用量的全局优化寻找多条反应路径。《自然·通讯》,8(1):15443,2017.

Michael Lutter、Christian Ritter 和 Jan Peters。《深度拉格朗日网络：将物理学作为深度学习的模型先验》。*arXiv* 预印本 *arXiv:1907.04490*，2019。

Nicholas Metropolis、Arianna W Rosenbluth、Marshall N Rosenbluth、Augusta H Teller 和 Edward Teller。《利用快速计算机进行状态方程计算》。《化学物理学杂志》，21(6)：1087–1092，1953年。David Morin。《经典力学导论：含问题与解答》。剑桥大学出版社，2008年。Lars Onsager 和 Stefan Machlup。《涨落与不可逆过程》。《物理评论》，91(6)：1505，1953年。Adam Paszke、Sam Gross、Francisco Massa、Adam Lerer、James Bradbury、Gregory Chanan、Trevor Killeen、Zeming Lin、Natalia Gimelshein、Luca Antiga 等。《PyTorch：一种命令式风格的高性能深度学习库》。《神经信息处理系统进展》，32，2019。Gareth W Richings、Iakov Polyak、Kaite E Spinlove、Graham A Worth、Irene Burghardt 和 Benjamin Lasorne。《使用高斯波包进行量子动力学模拟：VM CG 方法》。《国际物理化学评论》，34(2)：269–308，2015。Samuel Schoenholz 和 Ekin Dogus Cubuk。《JAX MD：可微物理框架》。《神经信息处理系统进展》，33：11428–11441，2020。Daniel V Schroeder。《交互式分子动力学》。《美国物理学杂志》，83(3)：210–218，2015。Chelsea Sweet、Oyewumi Akinfenwa 和 Jonathan J Foley IV。《通过基于发现的方法和分子动力学模拟，促进学生与真实气体性质的互动》。《化学教育杂志》，95(3)：384–392，2018。BL Zhang、J Wang、M Li 和 Q Hou。《钛表面附近氦气泡形成与气体释放的分子动力学研究》。《核材料杂志》，438(1-3)：178–182，2013。Daniel M Zuckerman 和 Thomas B Woolf。《一维稀有事件的高效动态重要性采样》。《物理评论 *E*》，63(1)：016702，2000年。

A 附录

A.1 附加背景

作用量的历史与解释。物理学家在 19th 世纪试图以纯解析方式推导固体和流体的力学时，发现了作用量 Ferguson (2004)。当时，人们希望建立一个连贯且数学上自洽的框架来描述物理世界。彼时的物理学观念高度依赖直觉，缺乏严谨的解析处理。因此，作用量问世之初，人们关注的重点是其实用性，而非其物理解释。即便如此，作用量的解释在物理学中仍具有根本意义。在大多数情况下（相对论中存在例外），作用量写作沿系统的（闭合）时间路径上动能总和与势能总和之差（见公式 1）。我们的直觉是，这两项大致分别对应于熵力和惯性力所引起的能量总流动。因此，最小作用量路径体现了这样一种观点：宇宙动力学源于与熵相关的力和与惯性相关的力之间的微妙平衡。

牛顿物理学观与拉格朗日物理学观。研究物理学时，人们通常从最符合人类直觉、最容易理解的概念入手，然后逐步转向能够以一般性方式描述自然的更抽象概念。例如，这正是基础物理课程强调牛顿物理学观的原因。在这种观念下，人们研究某一特定时间快照中的一组物体，列出它们所受的力，并更新其位置，从而得到系统的下一状态。这种观念很直观，因为它符合我们对时空的自然体验。相比之下，拉格朗日方法将物理系统的完整路径视为一个独立的数学实体。它是一种更抽象、也更违背直觉的方法，但同时具有更强的普适性，并且在不同领域中数学上更加一致。

相关工作：基于梯度的物理学优化。我们已经讨论过其他研究如何在不使用基于梯度的方法的情况下，研究 OM 作用量的最小化。此外，还有许多研究探索了在物理科学中使用自动微分和基于梯度的优化，但并未专门研究作用量。Schoenholz & Cubuk (2020) 在分子动力学背景下使用这些工具研究相变和晶体堆积等现象。Freeman et al. (2021) 引入了一种通用的可微分物理引擎，用于策略优化研究。近期的另一条研究路线将基于梯度的优化与微分方程（DE）相结合。例如，Chen et al. (2018) 提议使用神经网络对常微分方程进行参数化，而 Lutter et al. (2019)；Cranmer et al. (2020) 展示了如何对拉格朗日微分方程进行同样的处理。与此同时，Bar-Sinai et al. (2019) 等人的工作则着重于改进自适应积分方法。

A.2 附加方法

优化。 我们尝试了标准梯度下降、带动量的梯度下降，以及 Adam 优化器 Kingma & Ba (2014)，后者可以自适应地缩放动量。加入动量——无论是直接加入还是通过 Adam 实现——都使收敛时间至少缩短了一半。我们选择 Adam 优化器，以避免单独设置动量超参数。除了三体问题外，所有实验均使用 500 个优化步骤；三体问题在 1000 步后以近乎精确的精度恢复了常微分方程路径。

构建基准路径。 要用传统方法仿真经典系统的动力学，我们必须首先根据物理原理推导出一组二阶运动微分方程。通常通过应用牛顿第二定律或欧拉-拉格朗日方程（公式 2）来完成这项工作。本文讨论的所有系统，其运动方程都可以直接推导出来，并表示为表1中的一组耦合常微分方程。为了建立传统基准，以便与新的作用量优化方法进行对比，在优化器试验之前，我们使用已知能够产生复杂行为的边界条件，通过欧拉法对这些方程逐一进行数值求解。欧拉法是最基础且广为人知的常微分方程数值求解技术；选择它是为了使基准方法与同样基础、但尚未经过检验的基于优化的方法尽可能具有可比性。

解决无约束能量效应的其他方法。在最终选定正文中讨论的路径扰动方法之前，我们曾实施了几种替代方法，试图解决无约束能量效应

。起初，我们意识到问题在于缺乏足够的约束，因此尝试向优化器提供与常微分方程求解器用于解析系统总能量的相同的 $\mathbf{x}(t_0)$、 $\dot{\mathbf{x}}(t_0)$ 信息。在我们的离散路径模型中，最简单的做法是向优化器提供基线常微分方程模拟中的更多参考点。实际操作上，这意味着在优化过程中不仅冻结路径的初始值和最终值点，还要冻结其相邻的几个点。虽然这种方法确实约束了优化后路径的初始值和最终值速度，但除非冻结大范围的点，否则并不会显著影响整个路径的行为。由于这会大幅缩小潜在应用空间，我们转而关注其他可能的解决方案。

接下来，我们实现了两种不同的正则化方案，二者都基于在优化器损失函数中加入能量守恒项。第一种加入了全局能量损失项，该项会根据路径上各点的系统总能量与预期值之间的差异，对优化器施加相应的惩罚；预期值根据常微分方程$\mathbf{x}(t_0)$ 和 $\dot{\mathbf{x}}(t_0)$ 输入条件计算得出。第二种正则化项不依赖常微分方程的初始条件，而是加入了一个与系统总能量随时间的导数成比例的损失项，即局部能量守恒项。第一种方法在一定程度上有效，但在所测试的大多数系统中，路径扰动方法得出的结果要清晰得多。此外，路径扰动还展示了优化器方法根据较粗略的模型改进能量预测的潜力。尤其是在化学领域，精确的能量预测是计算模型的常见目标。另一方面，局部能量守恒方法并未带来明显优势。事实上，它有时会对优化器的表现产生显著的负面影响，这可能是因为它对可能的中间路径空间施加了额外限制。

就我们的探索性工作而言，路径扰动方法最为合适。我们认为，或许存在一种更好的方法，能够恰当地约束总能量，并将指向作用量最小化导向一条非退化路径。未来在这一领域的工作重点之一，应当是对此类方法进行界定。

系统	拉格朗日量 $\mathcal{L} = T - V$	运动方程	噪声 σ	Δt	LR
自由体	$\frac{1}{2}m\dot{x}^2 - mgx$	$\ddot{x} = g$	1.5	0.25	1
摆	$\frac{1}{2}ml^2\dot{\theta}^2 - mgl(1 - cos(\theta))$	$\ddot{\theta} = -\frac{g}{l}sin(\theta)$	$2e$-1	1	$5e$-2
双摆	附录	附录	$6e$-1	$6e$-2	$1e$-2
三体	$\sum_{i,j} \frac{1}{2}m_i\dot{\mathbf{x}}_\mathbf{i}^2 + \frac{\mathbf{G}\mathbf{m_i}\mathbf{m_j}}{\lvert\mathbf{r_{ij}}\rvert}$	$\ddot{\mathbf{x}}_\mathbf{i} = -\mathbf{G}\sum_\mathbf{j} \frac{\mathbf{m_j}}{\mathbf{r_{ij}^2}}\hat{\mathbf{r}}_\mathbf{ij}$	$3e$-2	0.5	$2e$-4
Gas	$\sum_{i,j} \frac{1}{2}m_i\dot{\mathbf{x}}_\mathbf{i}^2 + \mathbf{V_{LJ}}(\mathbf{r_{ij}})$	$\ddot{\mathbf{x}}_\mathbf{i} = \frac{1}{\mathbf{m_i}}\sum_\mathbf{j} \mathbf{F_{LJ}}(\mathbf{r_{ij}})$	$1e$-2	0.5	$1e$-4
星历	$\sum_{i,j} \frac{1}{2}m_i\dot{\mathbf{x}}_\mathbf{i}^2 + \frac{\mathbf{G}\mathbf{m_i}\mathbf{m_j}}{\mathbf{r_{ij}}}$	$\ddot{\mathbf{x}}_\mathbf{i} = -\mathbf{G}\sum_\mathbf{j} \frac{\mathbf{m_j}\,\mathbf{r}}{\mathbf{r_{ij}^2}}\hat{\mathbf{r}}_\mathbf{ij}$	$2e$+10	1天	$1e$+9

表1：所有优化均运行了500步。学习率 =学习率。

双摆拉格朗日量：

$$\mathbf{L} = \frac{\mathbf{1}}{\mathbf{2}}(\mathbf{m_1}+\mathbf{m_2})\mathbf{l_1^2}\dot{\theta}_\mathbf{1}^\mathbf{2} + \frac{\mathbf{1}}{\mathbf{2}}\mathbf{m_2}\mathbf{l_2^2}\dot{\theta}_\mathbf{2}^\mathbf{2} + \mathbf{m_2}\mathbf{l_1}\mathbf{l_2}\ \dot{\theta}_\mathbf{1}\dot{\theta}_\mathbf{2}\mathbf{cos}(\theta_1-\theta_2)\ + (\mathbf{m_1}+\mathbf{m_2})\mathbf{g}\mathbf{l_1}\mathbf{cos}(\theta_1)\ + \mathbf{m_2}\mathbf{g}\mathbf{l_2}\mathbf{cos}(\theta_2)$$

双摆运动方程：

$$\ddot{\theta}_i = \frac{f_i - \alpha_i f_j}{1 - \alpha_i\alpha_j} \quad \text{where} \quad f_1 \equiv -\frac{l_2}{l_1}[\frac{m_2}{m_1+m_2}\dot{\theta}_2^2 sin(\theta_1-\theta_2) - \frac{g}{l_1}sin(\theta_1)]$$

$$f_2 \equiv \frac{l_1}{l_2}\dot{\theta}_1^2 sin(\theta_1-\theta_2) - \frac{g}{l_2}sin(\theta_2)$$

Lennard-Jones 势：

$$V_{LJ}(\mathbf{r}) = \mathbf{4}\epsilon\left[\left(\frac{\sigma}{\mathbf{r}}\right)^{\mathbf{12}} - \left(\frac{\sigma}{\mathbf{r}}\right)^{\mathbf{6}}\right]$$

其中，$\mathbf{r}$ 表示粒子间距，ϵ 表示粒子间相互作用势能阱的深度，σ 与势能达到最小值时的 $\mathbf{r}$ 值有关。利用公式 $F = -\nabla U$，我们还可以推导出伦纳德-琼斯力：

$$F_{LJ}(\mathbf{r}) = \mathbf{4}\epsilon\left[\left(\frac{\mathbf{12}\sigma^{\mathbf{12}}}{\mathbf{r^{13}}}\right) - \left(\frac{\mathbf{6}\sigma^{\mathbf{6}}}{\mathbf{r^{7}}}\right)\right]$$

	作用量总和 S（Js）		动能总和 T（Js）		势能总和 V（Js）		均方误差	
系统	ODE	Sim	ODE	Sim	ODE	Sim	初始值	最终值
自由体	-13.8	-14.0	9.54	8.84	23.4	22.8	630	0.352
摆	-16.3	-16.3	81.9	81.7	98.2	98.0	0.118	1.20e-2
双摆	2.59	2.59	1.58	1.55	-1.01	-1.04	0.170	3.22e-3
三体	6.94e-4	6.96e-4	2.48e-4	2.49e-4	-4.46e-4	-4.47e-4	9.74e-3	5.74e-6
Gas	7.81e-5	7.71e-5	1.11e-4	1.18e-4	3.30e-5	4.13e-5	3.09e-5	2.80e-6
星历	3.76e33	3.75e33	1.28e33	1.26e33	-2.48e33	-2.49e33	1.50e20	7.12e17

表 2：第3节所述六个经典系统的定量结果。

气体模拟。分子动力学（MD）模拟能够研究那些难以通过实验观察的动态系统（Hollingsworth & Dror，2018）。这类模拟可追溯至20世纪50年代，最早的工作研究的是硬球气体（Alder & Wainwright，1959）。如今，MD 被用于理解复杂得多的系统，从钛中的氦扩散到蛋白质折叠（Zhang 等人，2013；Lee 等人，2017）。这些研究方向随着我们计算能力的提升而同步发展。在最基础和传统的形式中，MD 模拟通过获取系统中所有原子的位置，利用某种势来模拟原子之间的相互作用，计算每个两两相互作用，并相应地更新原子的速度和位置来运行（Schroeder，2015）。

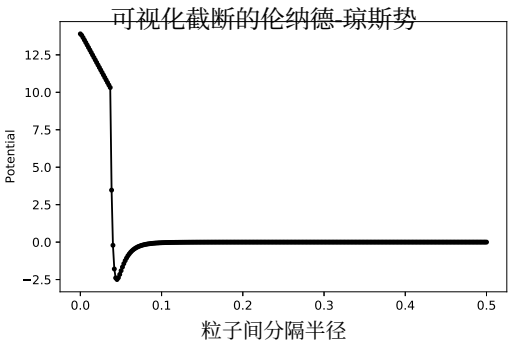

图5：截断的伦纳德-琼斯势的可视化。势的截断部分采用线性化斜率，以防止与较小半径相关的力变得极大并造成数值不稳定性。

我们的仿真与早期的分子动力学仿真以及简单的分子动力学仿真类似，例如 Schroeder（2015）的仿真；这类仿真常用于帮助学生在化学和物理学入门课程中理解气体的性质。此仿真并未针对特定材料或问题进行优化，因此我们能从中获取的定量信息有限。为简化起见，我们使用 Lennard-Jones 势（图5）来确定相互作用力，但它只能描述两两相互作用。也可以采用更复杂的力场。即使使用这种简单的仿真，我们也能生成类气态和类固态（图6），类似于其他采用 Lennard-Jones 势的简单分子动力学仿真所能实现的效果（Schroeder，2015；Sweet 等，2018）。

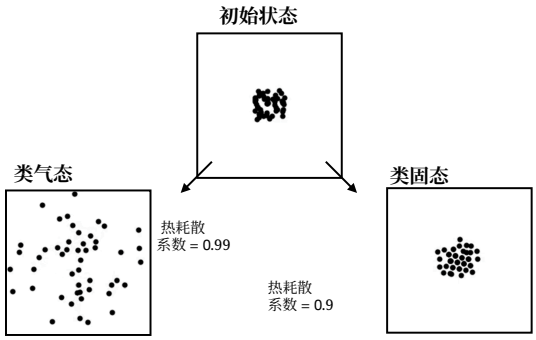

图6：使用两种不同的热耗散系数生成的固态和气态可视化结果。

星历实验。我们下载了太阳系内行星在2022日历年内的原始星历数据（分辨率为1天）。为此，我们使用了 NASA 的 Horizons 项目提供的在线界面。我们采用太阳系质心（SSB）作为坐标中心。在构建仿真时，我们使用了简单的引力势阱势能 $\frac{Gm_i m_j}{r_{ij}}$，并将万有引力常数 G、行星质量 m 和时间长度均采用国际单位制。为了探究作用量最小化能否重建内行星轨道，我们只扰动内行星的路径，不扰动太阳的路径。我们将本实验的时间长度设为两个月，因为金星和水星的轨道周期比地球短（随着轨道开始跨越多个周期并与自身尾部重叠，初始和最终路径的定性可视化变得困难）。

A.3 补充量子内容

最小作用量原理是量子路径积分的一种极限情况。最小作用量原理不过是路径积分的一种极限情况：在这种情况下，所考虑路径的作用量可以趋近于经典尺度。在此范围内，作用量存在变化的路径及其邻近路径所贡献的相位彼此不同步。对这些区域的贡献求和，类似于对快速振荡的正弦函数进行积分：每个具有正（相长）贡献的部分，都必然会被相邻的负（相消）贡献抵消。随着空间尺度增大，沿任意给定路径的振荡变得愈加频繁，任意一组路径的净贡献都趋近于零。唯一的例外是附近路径组成的群组中，作用量的变化趋于零的情况。即使在较大的空间尺度下，这些路径贡献的相位也相似，并会发生相长干涉。从这个角度来看，经典最小作用量原理实际上描述的是量子路径积分的共振模态。

图4相关的补充讨论。图4以图示方式呈现了量子—经典共振极限，其中在三个不同的空间尺度下运行了相同的量子模拟。请注意，随着尺度增大、趋近经典性，即使沿着非对角线发生微小移动，也会引发一系列快速振荡。这些是高作用量路径：粒子在消失的 dt 时间内，沿直线在空间网格中的相距甚远的点之间穿行。这些路径产生的贡献几乎在每次试验中都会彼此相消。我们可以从每次模拟按时间截取的快照中直接看到这一结果；在较大的尺度下，同一个初始波包退相干得更慢，与频率较低、因而更加量子的对应情况相比，只呈现出精细的高频干涉结构。随着我们趋近经典极限，整个波包将越来越精确地沿着粒子撞击墙壁时的经典最小作用量路径运动。