从费马原理到物理信息神经网络：变分物理的统一计算方法

Aman Razdan,[1] Aditya Shankar Mazumdar,[1] AmitTanwar,[1] 和 Pragati Ashdhir[1]印
度德里*110007*，德里大学 *HinduCollege* 物理系
（*电子邮箱：pragatiashdhir@hinducollege.ac.in）

变分原理是物理学诸多领域共通的数学框架，但在本科阶段的教学仍以解析方法为主。本文提出一种注重教学且融入计算技术的变分建模方法，结合梯度下降、自动微分和物理信息神经网络（PINNs）等现代工具。本文将经典变分问题重新表述为优化任务，并使用 NumPy、Matplotlib、PyTorch 和 JAX 等开源 Python 库加以实现。我们通过一系列源自标准本科课程的问题展示这一方法，包括从费马原理推导斯涅尔定律、考虑和不考虑黏性阻力的抛体运动、简谐运动、带阻尼的非线性摆、由拉普拉斯方程和泊松方程描述且具有非线性温度相关内部热生成的稳态热传导、基于最小作用量原理的双摆，以及振动弦的变分分析。此外，本文还介绍了量子力学应用，通过氢原子、氦原子和硅原子核示意性核模型的变分解，展示该框架在经典物理、量子物理和核物理领域的广泛适用性。该方法旨在增强学生的概念理解，同时使他们接触现代计算研究方法。

arXiv:2601.01262v1 [physics.comp-ph] 3 Jan 2026

I. 引言

变分思想在物理学的众多分支中发挥着统一作用，但其教学潜力在本科教学中尚未得到充分发挥。从光学现象背后的几何原理，到涉及深度神经网络的当代应用，自然界与计算都揭示了一个共同的组织主题：物理行为受极值原理和泛函优化支配。这种共同结构为教学提供了绝佳机会——变分建模不仅可以作为传统上彼此分离的子领域之间的概念桥梁，也可以成为解决物理问题的实用且易于计算的框架。

从历史上看，变分推理始于费马在十七世纪提出的论断：光沿着最短时间路径传播[1,2]。当欧拉和拉格朗日为这一思想奠定严格的数学基础后，它发展为欧拉–拉格朗日方程，并使约束动力学得以现代化表述。随后，哈密顿驻作用量原理和诺特定理将变分结构提升到分析力学中的核心哲学地位，揭示物理定律源自对恰当泛函的极值化[3]。尽管本科课程往往只零散地讲授这些概念，但将它们结合起来，便能生动展现变分观点的统一力量。

计算物理的进步进一步拓展了变分方法的应用范围。离散极值化程序——以 Ritz 方法和伽辽金方法，以及后来的有限元法为代表——使得微分方程能够在保留其内在物理结构[4]的同时得到近似求解。由此，优化从纯粹的理论构想发展成一种实用的计算工具，如今已成为真实物理系统建模不可或缺的一部分。

随着自动微分的出现，变分计算获得了新的算法基础。通过将链式法则机械化，这些技术即使面对大型高维模型，也能提供机器精度导数[5,6]。反向模式微分——在物理教育中更广为人知的名称是反向传播——恢复了一种熟悉的结构：沿着泛函的梯度下降以求极值。从这个意义上说，深度学习优化并非偏离物理推理，而是变分法的现代延伸。

物理信息神经网络（PINNs）是这一融合在当代的一种体现。通过将控制微分方程直接嵌入损失泛函，PINNs 在单一变分框架[7]内整合了优化、微分和物理定律。对于本科生而言，这传达了一个重要的理念：机器学习模型不必完全依赖经验，也可以建立在他们于物理学入门课程中接触到的相同物理原理之上。

在本研究中，我们利用开源的 Python 库，将这些进展纳入适合本科生学习的框架：使用 NumPy 和 Matplotlib 进行数值计算与可视化，并使用 PyTorch 和 JAX 实现自动微分和基于神经网络的建模。我们通过一系列结构化递进的问题展示这一框架，这些问题常见于本科物理学课程。基础层级的示例包括从费马原理推导斯涅尔折射定律、考虑和不考虑黏性阻力的抛体运动，以及非线性摆的分析。中级层级的示例包括通过拉普拉斯方程和泊松方程建模稳态热传导，以及运用最小作用量原理研究双摆。最后，高级层级的问题运用瑞利–里兹法研究经典振动弦、量子力学处理

氢原子和氦原子，以及利用伍兹–萨克森势进行硅原子核的核结构建模。这些示例共同展示了变分原理与先进计算工具的无缝结合，使学生既能深入理解概念，也能接触计算物理当前研究前沿的方法。

II. 方法论框架

本节将介绍理解本文结果及其分析所需的背景知识。

A. 理论基础

我们从泛函的数学引入开始，构建本文的理论背景。标准函数$f : x \to y$ 将数值输入映射为数值输出，而泛函 $\mathscr{F}$ 则将整个函数映射为一个标量值：

$$\mathscr{F}[f] = \lambda, \quad \text{where } \lambda \in \mathbb{R} \tag{1}$$

泛函通常表示为如下形式的定积分：

$$\mathscr{F}[f] = \int_a^b f(x)\,dx. \tag{2}$$

在这里， $\mathscr{F}$ 以曲线$f(x)$的完整形状作为输入，并输出一个数。变分法的中心任务是找到使$\mathscr{F}$ 驻定[8]的特定函数 $f(x)$。类似于在普通微积分中寻找一个点x_0，使导数$f'(x_0) = 0$，我们要寻找一个函数$f(x)$（满足某些边界条件，例如$f(a) = A$ 且$f(b) = B$），使得变分 $\delta\mathscr{F} = 0$。在数学上，我们引入一条试探路径 $\tilde{f}(x) = f(x) + \varepsilon g(x)$，其中$g(x)$是一个在边界处为零的任意微扰。驻定条件要求泛函关于微扰幅度 ε 的导数为零，即

$$\frac{d}{d\varepsilon}\mathscr{F}[f + \varepsilon g]\Big|_{\varepsilon=0} = 0 \tag{3}$$

这意味着$f(x)$是函数无限维空间中的一个临界“点”。自然界似乎偏好这些驻定路径。许多经典现象都由特定基本泛函的最小化所支配[9]：

1. **距离（测地线）：** 曲面空间中的物体沿着使路径长度最短的路径运动。

2. **时间（费马原理）：** 光沿着传播时间最短的路径传播，从而解释了反射和折射现象。

3. **势能（狄利克雷原理）：** 肥皂膜或悬挂电缆等静态系统会形成使总势能最小的形状。

4.**作用量（哈密顿原理）：** 动力系统沿着使作用量保持驻定的轨迹演化。

本研究聚焦于对具有物理意义的泛函进行计算最小化，包括光学时间和力学作用量（表 I），以求得光线轨迹、动力学路径，以及本科物理学中遇到的典型逆变分问题。

表 I：本研究中最小化的基本泛函。此处，$n(x, y)$ 表示折射率，而 $L(t, y, \dot{y})$ 表示力学系统的拉格朗日量。

物理量	泛函形式
传播时间（光学）	$T[y] = \int x_1 x_0 n(x,y)\sqrt{1 + y'^2}\,dx$
作用量（力学）	$S[y] = \int_{t_0}^{t_1} L(t,y,\dot{y})dt$

B. 计算实现

将上一节讨论的变分思想转化为计算实践，需要学生掌握若干互补数值方法和算法的基本知识，具体如下。

1. 自动微分

自动微分（AD）是一类用于以机器精度计算函数导数的算法技术。它通过将复杂计算分解为一系列基本运算，并在每一步系统地应用链式法则来实现这一点。与符号微分不同，自动微分避免了表达式迅速膨胀的问题；与有限差分近似不同，它不会引入截断误差或舍入误差[10,11]。

为说明其基本思想，考虑函数

$$z = f(x,y) = x\,y + \sin(y). \tag{4}$$

引入中间变量

$$a = x\,y, \qquad b = \sin(y), \tag{5}$$

计算过程可表示为

$$z = a + b. \tag{6}$$

这种分解可以用有向无环图的形式直观呈现，即图 1 所示的计算图 。在这种图示中，节点表示变量和中间量，边表示基本运算。

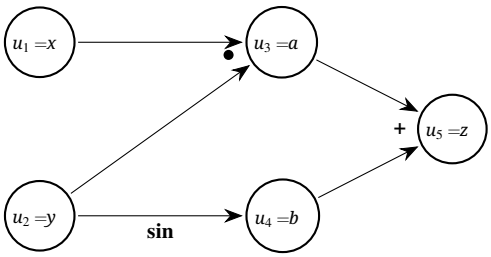


图 1：z的计算图

在这一框架中，计算导数就是利用链式法则，让导数信息 沿计算图传播。根据传播方向的不同，自动微分可分为两种不同模式。

在前向模式自动微分中，导数沿着计算本身的方向，从输入传播到输出。图中的每个节点u_i 都同时携带其数值及其相对于选定独立变量s的导数。如果u_j 表示u_i的子节点，则链式法则可写为

$$\frac{\partial u_j}{\partial s} = \sum_i \frac{\partial u_j}{\partial u_i}\frac{\partial u_i}{\partial s}. \tag{7}$$

随着计算的进行，导数信息与原始计算同步累积。在上面的例子中，对于$s = x$，有

$$\begin{aligned}\frac{\partial z}{\partial x} &= \frac{\partial u_5}{\partial x}\\ &= \frac{\partial u_5}{\partial u_3}\frac{\partial u_3}{\partial x} + \frac{\partial u_5}{\partial u_4}\frac{\partial u_4}{\partial x}\\ &= \frac{\partial z}{\partial a}\frac{\partial a}{\partial x} + \frac{\partial z}{\partial b}\frac{\partial b}{\partial x}\\ &= \frac{\partial a}{\partial x} + \frac{\partial b}{\partial x}\end{aligned} \tag{8}$$

因此，计算 $\partial z/\partial x$需要追踪x的变化如何影响中间变量a和b，以及这些变化随后如何传播到z。我们也可以对$s = y$进行类似计算。

在反向模式自动微分中，链式法则沿相反方向实现。计算首先向前进行，以评估并存储所有中间变量，从而记录计算轨迹。随后，导数从输出向输入反向传播。如果s 表示标量输出，且u_i 表示一个中间节点，则反向累积可表示为

$$\frac{\partial s}{\partial u_i} = \sum_j \frac{\partial u_j}{\partial u_i}\frac{\partial s}{\partial u_j}, \tag{9}$$

其中，求和遍历u_i的所有子节点u_j 。对于当前示例，z 相对于x和y 的导数可通过沿图进行一次反向遍历同时获得。下面以$s = z$ 和$i = 1 \implies u_1 = x$为例进行说明：

$$\begin{aligned}\frac{\partial z}{\partial u_1} &= \frac{\partial z}{\partial x}\\ &= \frac{\partial z}{\partial u_3}\frac{\partial u_3}{\partial x} + \frac{\partial z}{\partial u_4}\frac{\partial u_4}{\partial x}\\ &= \frac{\partial z}{\partial a}\frac{\partial a}{\partial x} + \frac{\partial z}{\partial b}\frac{\partial b}{\partial x}\\ &= \frac{\partial a}{\partial x} + \frac{\partial b}{\partial x}\end{aligned} \tag{10}$$

尽管式 7 和式 9 中前向模式与反向模式的数学表达形式上看起来相似，但它们的计算行为有着根本差异。前向模式在原始计算的同时计算导数，因此内存开销相对较小；但当输入变量数量较大时，它会变得低效，因为每个输入都需要单独进行一次前向扫描。相比之下，反向模式的内存开销较高，因为必须存储完整的前向轨迹；但它能够通过一次反向扫描，高效计算相对于所有输入变量的梯度。因此，前向模式最适合输入数量少而输出数量多的问题；反向模式则特别适用于高维输入空间，这类空间常见于变分形式、反问题以及现代基于物理学的机器学习中。

在本研究中，导数使用 JAX 库中的 jax.grad() 函数进行计算[12], ，该函数能够高效实现反向模式自动微分。

2. 梯度下降

梯度下降（GD）是一种用于确定函数最小值的迭代优化算法[13]。对于可微标量函数 $f(x)$，可通过以下更新规则逐步获得使函数达到最小值的自变量近似值

$$x_{i+1} = x_i - \eta \left.\frac{df}{dx}\right|_{x=x_i}, \tag{11}$$

其中，x_i 和 x_{i+1} 分别表示变量在第 i^{th}次和第（$i+1$）次 th迭代时的取值， $\eta > 0$为学习率（或步长），用于控制算法的收敛速度。

当 η足够小时，式（11）中的迭代会趋向于 $f(x)$ 的局部极小值，前提是 f 光滑且有下界。如果 η 选择得过大，迭代可能越过极小值而无法收敛；而 η 取值过小则会导致收敛缓慢。

在本研究中，目标函数不是单个变量的函数，而是一个泛函，其自变量是一个离散化的

场。形式如下的连续泛函

$$\mathscr{F}[y] = \int_a^b \mathscr{L}\big(y(x), y'(x), x\big)\, dx \tag{12}$$

当函数$y(x)$满足欧拉–拉格朗日方程时，泛函取得极值

$$\frac{\partial \mathscr{L}}{\partial y} - \frac{d}{dx}\left(\frac{\partial \mathscr{L}}{\partial y'}\right) = 0. \tag{13}$$

我们不对方程（13）进行解析求解，而是将定义域离散化，并将泛函 $\mathscr{F}[y]$ 近似为一个标量值多元函数

$$\mathscr{F}[y] \longrightarrow \mathscr{F}(y_1, y_2, \ldots, y_N), \tag{14}$$

其中，$\mathbf{y} = (y_1, y2, \ldots, y_N)$表示空间网格上的场值。在这种离散设定下，泛函导数可近似为 $\mathscr{F}$ 关于$\mathbf{y}$的梯度。

梯度下降更新随后自然推广为

$$\mathbf{y}_{i+1} = \mathbf{y}_i - \eta \left.\frac{\partial \mathscr{F}}{\partial \mathbf{y}}\right|_{\mathbf{y}=\mathbf{y}_i}, \tag{15}$$

其中$\mathbf{y}_i$ 和$\mathbf{y}_{i+1}$ 分别表示第i^{th}次和第$(i+1)^{\text{th}}$次迭代时的场值数组。

收敛时，梯度 $\partial\mathscr{F}/\partial\mathbf{y}$趋于零，而离散解则逼近连续极限下由欧拉–拉格朗日方程得到的驻函数。从这个意义上说，梯度下降提供了变分原理的一种计算实现，将求解微分方程的问题转化为迭代优化问题。

这一视角与现代计算框架自然契合。在这些框架中，自动微分能够高效、准确地计算梯度，从而可以将优化算法直接应用于物理学中的广泛问题。

3. 物理信息神经网络（*PINNs*）

物理信息神经网络（PINNs）融合了深度学习与基于物理的建模，其神经网络经过训练以满足物理系统的控制方程，而非仅仅依赖标注数据。在这一框架中，微分方程、边界条件和初始条件等形式的先验物理知识被直接嵌入网络的损失函数中[7]。因此，训练过程被表述为对残差（或损失）泛函进行最小化；该泛函将由神经网络参数化的函数映射为一个标量，用以衡量总误差。

设物理系统由常微分方程（ODE） $\mathscr{N}[u(t)] = 0$ 或偏微分方程（PDE） $\mathscr{N}[u(x,t)] = 0$描述，具体取决于独立变量的数量。在这两种情况下，方程都定义在某个定义域 Ω上，并结合适当的初始和/或边界条件。在 PINN 中，未知场u由神经网络 u_θ近似，其中 θ 表示可训练参数（权重和偏置）的集合。因此，神经网络充当一种灵活且可微的试探函数（见图 2），能够表示控制微分方程的解，同时通过优化来满足相关的物理约束。

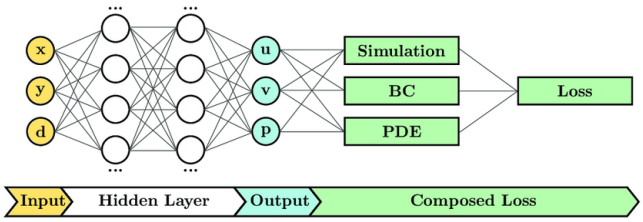


图2：用于求解正问题和逆问题的物理信息神经网络架构示意图。网络将空间坐标（x，y）和参数坐标（d）作为输入，使其通过隐藏层，并输出状态变量（u，v，p）。损失函数由数据模拟损失、边界条件（BC）损失和偏微分方程残差损失组成。转载自参考文献14

目标损失函数 $\mathscr{L}(\theta)$ 的构造旨在惩罚对控制方程和施加约束条件的违反。它通常表示为若干不同标量损失项的加权和，

$$\mathscr{L}(\theta) = w_{\text{data}}\mathscr{L}_{\text{data}} + w_{\text{phys}}\mathscr{L}_{\text{phys}}, \tag{16}$$

或者，等价地，分解为微分方程、边界条件和初始条件各自的贡献，

$$\mathscr{L}(\theta) = \mathscr{L}_{\text{PDE}} + \mathscr{L}_{\text{BC}} + \mathscr{L}_{\text{IC}}. \tag{17}$$

数据损失项 $\mathscr{L}_{\text{data}}$ 用于确保与观测测量值、边界条件或初始条件相符，通常通过网络预测值与可用数据点之间的均方误差（MSE）[7,15]进行评估。物理损失项 $\mathscr{L}_{\text{phys}}$ 起到基于物理学的正则化作用。如果控制方程——无论是常微分方程还是偏微分方程——可以表示为残差形式$f(\boldsymbol{\xi}) \approx 0$，其中 $\boldsymbol{\xi}$ 表示相关自变量，则网络会使用自动微分，在一组分散的配点 $\{\boldsymbol{\xi}_i\} \subset \Omega$处计算该残差，

$$\mathscr{L}_{\text{phys}} = \frac{1}{N_r}\sum_{i=1}^{N_r} \left|\mathscr{N}[u_\theta(\boldsymbol{\xi}_i)]\right|^2. \tag{18}$$

当网络预测的导数违反基础物理定律时，这一项便会对其进行惩罚。这里，配点是用户在定义域中选取的一组点，控制微分方程在这些点上得到近似强制执行，而无需引入空间或时间网格。

关键在于，所有时间导数、空间导数或两者兼有的导数——出现在 $\mathscr{N}[u_\theta]$ 中的——均通过自动微分计算，从而确保以机器精度进行求值，而无需诉诸数值离散。由此得到的优化问题，

$$\theta^* = \arg\min_\theta \mathscr{L}(\theta) \tag{19}$$

采用基于梯度的优化算法求解，例如梯度下降或其现代变体。在公式 19 中，极小值运算符 arg min 返回使损失函数达到最小值的参数值 θ^*，而非最小值本身。

从变分视角来看，物理信息神经网络可以视为在由神经网络参数化的函数空间上进行优化。网络参数并非通过显式求解与某个泛函相关联的欧拉–拉格朗日方程，而是通过迭代更新，使试探函数$u_\theta(\xi_i)$逐渐逼近控制方程的驻定解。从这个意义上说，物理信息神经网络为经典变分原理与现代基于机器学习的优化方法搭建了一座计算桥梁。

这一表述对物理学教育和研究尤其具有吸引力，因为它在单一计算框架中统一了微分方程、变分方法、自动微分和优化，从而能够以无网格方式求解复杂的物理问题[16]。

在我们的工作中，我们使用 PyTorch 库中的 torch.nn 模块（用于初始化神经网络）、torch.autograd 模块（用于计算自动微分梯度）以及 torch.optim 模块中的 torch.optim.Adam 类（作为优化器）实现了物理信息神经网络[17]。

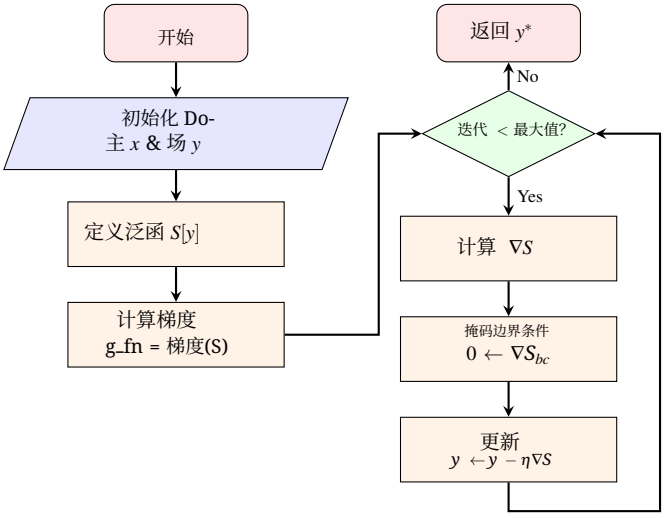


图 3：本文所用变分优化框架的通用算法流程。从定义在计算域 x上的初始试探场y 出发，构造特定问题的泛函$S[y]$。使用自动微分计算泛函梯度 ∇S，并以学习率 η通过梯度下降迭代使其最小化。通过在受约束点屏蔽梯度来施加边界条件。迭代持续进行，直至达到规定的最大迭代次数，从而得到优化后的场y^*。

4. 算法流程

由于所提出的方法适用于各种各样的物理系统，我们概述了一个通用的算法框架，只需提供适当的物理输入，即可将其应用于任何问题。图 3 所示的流程图总结了这一流程。对于给定的计算域（可以是一维或多维），我们首先使用随机值初始化试探场y （或路径），同时满足规定的边界条件或初始条件。利用这一试探构型，根据所考虑的物理问题构造相应的泛函$S[y]$。

随后，使用自动微分计算泛函相对于场变量的梯度。利用这一梯度信息，通过梯度下降算法最小化泛函，迭代更新场。该迭代过程将持续进行，直到达到收敛，或达到预先设定的最大迭代次数（记为 Max 或 Max_Iter）。参数 η表示学习率（或步长），用于控制每次更新的幅度；该参数此前已在梯度下降的相关内容中介绍过。

III. 示例

我们通过两类互补的问题展示所提出计算框架的适用范围和有效性：正向变分问题，其中控制泛函已知；以及逆向变分问题，其中泛函由给定的微分方程构造而成。在每种情形下，我们都着重介绍其背后的物理思想和主要结果，而具体的数值实现细节以及相关 Python 代码则收录于补充材料中的章节。每个示例的结尾都会明确注明相关章节编号。

A. 正向变分问题

在此，我们讨论控制泛函已明确给定并作为基本出发点的系统。对定义泛函的积分进行离散化，并通过直接优化数值求得其最小值。

1. 费马原理：光学中的最短时间

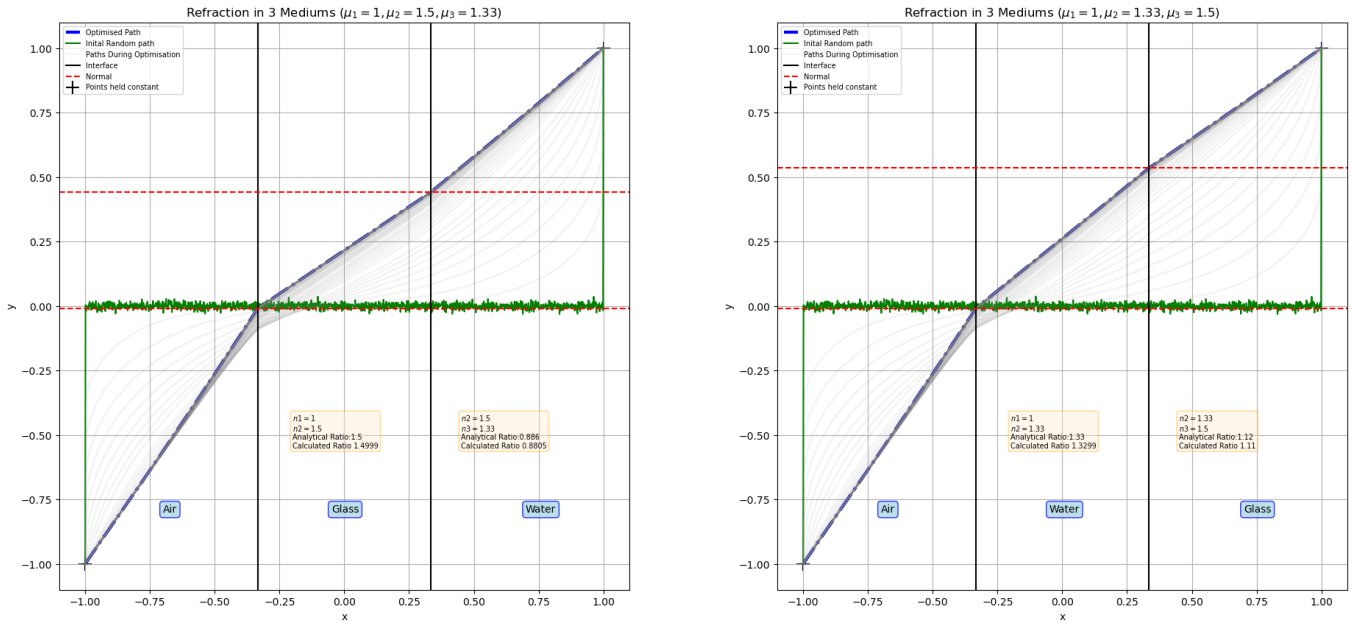
光学中的基本泛函是传播时间。费马原理 指出，光沿着使传播时间最短的路径传播，从而得到折射的斯涅尔定律：

$$n_1 \sin\theta_1 = n_2 \sin\theta_2, \tag{20}$$

其中n_i 表示i^{th} 介质的折射率，θ_i 表示光线与该介质中法线的夹角。通过离散化时间泛函（表 I）并应用梯度下降，我们可以直接确定光路。优化成功恢复了斯涅尔定律所预测的轨迹，验证了该计算框架适用于几何光学。

本研究考虑了两种不同的设置：

1. Transition from Air → Glass → Water (see Fig. 4a)



(a) 情形 1：低折射率-高折射率-低折射率

(b) 情形 2：低折射率-高折射率-高折射率

图 4：**穿过三种介质的折射**：为方便数值计算，取 c 的值为 1，将 x –定义域 从 -1 到 1 划分为 1，500 个点，因此每个界面有 500 个点。假设初始路径服从均值为 0、方差为 1 的正态分布，并将端点固定在 $(-1,\ -1)$ 和 $(1,\ 1)$，由此得到图中所示的图案，

2. 从空气 →水 → 玻璃的过渡（见图 图 4b)

为便于开展研究，我们将两个边界点固定在坐标（$-1,\ -1$）和（$1,\ 1$）处。我们构造了一个跨越该区间的x数组，并将其划分为三个相等的区段，为每个区段赋予相应的折射率，以表示三种不同的光学介质。固定端点之间的光路由离散数组y（x）表示。该数组采用随机方式初始化，同时满足边界值保持固定的约束，即y（-1）$=-1$ 和y（1）$=1$。这种随机初始化确保不会对解施加任何先验几何偏置，并使优化过程能够仅依据变分原理确定极值光路。从跨越三种介质的这条初始随机光路出发，使用梯度下降算法最小化光程泛函。经过 successive 迭代，光路逐渐演化为满足费马原理的驻定解，最终恢复出物理上正确的折射轨迹。

为了确定入射角和折射角，并据此验证斯涅尔定律，我们分析了材料界面附近优化后的光路。界面设置为平行于y轴，因此相应的表面法线与x轴对齐。在这种几何结构下，光线与法线之间的夹角等同于光线与x轴之间的夹角。通过计算优化路径的局部斜率，数值求得各介质中的光线角度。具体而言，利用界面相邻的离散点计算斜率：界面处的y坐标与相应介质中最近点的y坐标之差除以均匀网格间距dx。该斜率的反正切 即为相应的光线角度。由此提取的角度与解析计算的预测结果（表 II）高度吻合，从而证实了斯涅尔定律，并验证了变分计算方法。源代码见补充材料第 1 章。

表 II：由列表材料常数（n_2/n_1）得到的相对折射率，与根据斯涅尔定律（$n_2/n_1 = \sin i/\sin r$）从数值优化光路中提取的相对折射率之比较。

界面	i (°)	r (°)	n_2/n_1	$\sin i/\sin r$
空气 →水	56.7479	38.95	1.3300	1.3303
水 →玻璃	38.9500	34.35	1.1278	1.1141

2. 哈密顿原理：力学中的驻定作用量

哈密顿原理指出，物理系统在两个固定时刻之间的实际运动沿着作用量取驻值的路径进行。Greydanus等人[18],则强调了一种互补且当代的阐释，他们将作用量称为“自然的代价函数”。从这一视角来看，物理演化可以理解为一个优化过程：在所有符合边界条件的可容许路径中，实际形成的轨迹使一个标量泛函达到极值——

这个泛函即作用量 $S[y]$，定义见表 I。为凸显这一阐释的重要性及其教学价值，我们将详细分析三个典型问题。

抛体运动

第一个是较简单的一维抛体运动。设质量为m的抛射体的高度随时间变化的函数记为$y(t)$。那么，该系统的拉格朗日量定义为动能（T）与势能（V）之差：

$$\mathscr{L}(t,y,\dot{y}) = T - V = \frac{1}{2}m\dot{y}^2 - mgy, \tag{21}$$

根据表 I，作用量定义为

$$S[y(t)] = \int_{t_0}^{t_1} \mathscr{L}(t,y,\dot{y})\,dt = \frac{m}{2}\int_{t_0}^{t_1} (\dot{y}^2 - 2gy)\,dt \tag{22}$$

通过对式 (22) 中的积分进行离散化，得到一个初始 $[y_0]$ 数组，并使用梯度下降进行最小化，我们能够确定最优轨迹，该轨迹与解析结果一致（见图 5a）。
源代码请参阅补充材料 第 2 章第 2.1 节。

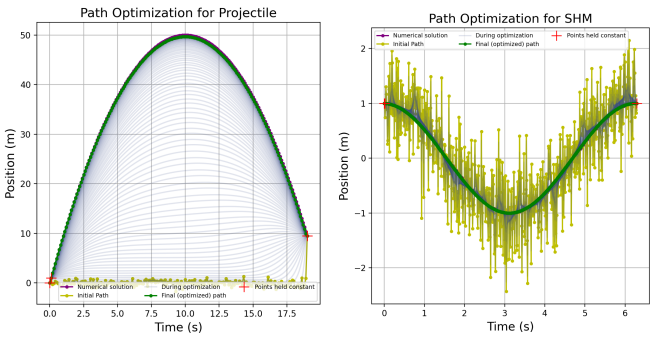


图 5：**通过作用量极小化进行路径优化。(a)** 抛体运动和**(b)** 简谐振子的轨迹演化。优化器逐步调整初始试探路径（黄色），经过中间状态（浅蓝色），最终得到驻定作用量路径（绿色），它与四阶龙格-库塔基准（紫色）完全重合。所有物理常数均设为 1。抛射体的时间步长为$dt = 0.1$ s（$t \in [0,19]$ s），振子的时间步长为$dt = 0.01$ s（$t \in [0,10]$ s）。

简谐运动

第二个问题聚焦于简谐振子（*SHM*）的周期运动。这是一个基本模型，构成了广泛物理系统的基础。设质量为m的质点连接在力常数为k的弹簧上，并在某一时刻t相对于平衡位置的位移为$x(t)$。那么，对于这个系统，我们有拉格朗日量

$$\mathscr{L}(t,x,\dot{x}) = \frac{1}{2}m\dot{x}^2 - \frac{1}{2}kx^2 \tag{23}$$

以及，作用量

$$S[y(t)] = \int_{t_0}^{t_1} \mathscr{L}(t,x,\dot{x})\,dt = \frac{1}{2}\int_{t_0}^{t_1} (m\dot{x}^2 - kx^2)\,dt \tag{24}$$

然而，由于一个细微的疏忽，直接将作用量最小化并不能得到正确结果，这被称为无约束能量效应[18]。基本上，在这种情况下，质点的周期运动本身可以视为一种运动约束，需要在计算中加以编码。我们通过施加能量守恒来解决这一问题：定义能量残差（总能量相对于其初始值的偏差），并将该残差与作用量的加权组合进行最小化。如果ws 和we 分别是作用量和能量残差对应的权重，那么修正后的作用量可表示为

$$S' = w_s S + w_e \Delta E. \tag{25}$$

经过调优权重后，该方法可恢复正确的数值解（见图 5b）。

源代码请参阅补充材料 第 2 章第 2.1 节。

双摆

基于前一个问题的结果，第三个示例扩展到一个耦合质点系统：平面双摆。两个质点m_1 和m_2 由长度分别为l_1 和l_2的刚性无质量杆连接。质点m_1 通过第一根杆连接到固定支点，与竖直方向的夹角为 θ_1；而质点m_2 通过第二根杆连接到m_1，与竖直方向的夹角为 θ_2。于是，该系统的拉格朗日量可写为

$$\begin{aligned}\mathscr{L}(t,\theta_1,\dot{\theta}_1,\theta_2,\dot{\theta}_2) = &\frac{1}{2}m_1 l_1^2\dot{\theta}_1^2 + \frac{1}{2}m_2\{l_1^2\dot{\theta}_1^2 + l_2^2\dot{\theta}_2^2 \\ &+ 2l_1 l_2\dot{\theta}_1\dot{\theta}_2\cos(\theta_1-\theta_2)\} \\ &+ m_1 g l_1\cos\theta_1 \\ &+ m_2 g\{l_1\cos\theta_1 + l_2\cos\theta_2\}\end{aligned} \tag{26}$$

相应的作用量按通常方式定义为拉格朗日量对时间的积分。由于该系统具有两个自由度，对应于两个角坐标，因此通过对两个角位移 θ_1 和 θ_2进行优化，使能量残差（如前一个问题所示）与作用量的加权组合最小化。这种变分优化得出的驻作用量轨迹与直接数值积分得到的解吻合良好。（见图 6）。源代码请参阅补充材料 第 2 章第 2.2 节。

B. 逆变分问题

在许多物理和数学情境中，系统通常由控制微分方程直接描述，而非

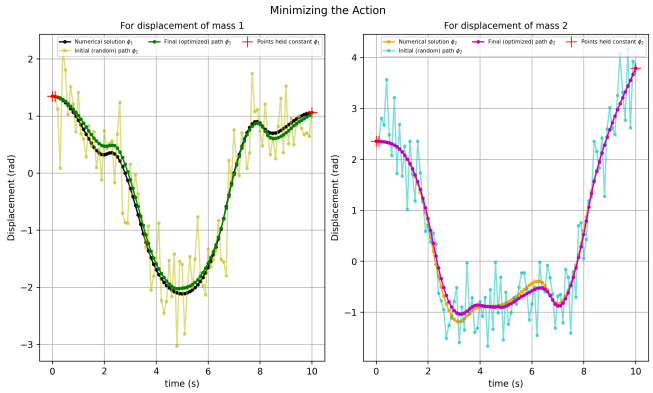


图 6：**双摆的轨迹优化**。图中展示了第一和第二个质点的位移曲线。优化器将随机生成的初始路径（黄色和青色）逐步优化为最终的驻作用量轨迹（绿色和洋红色），这些轨迹与数值解（虚线）相吻合。所有物理参数（m，l，g）均设为 1。优化采用的时间步长为 $dt = 0.1$ s，定义域为 $t \in [0, 10]$ s

由显式的变分原理描述。在这种情况下，可以运用变分法的逆问题，有意构造一个适当的泛函，使其欧拉–拉格朗日方程能够复现给定的微分方程。在此框架下，微分方程被视为一项约束，用以指导合成作用量泛函的设计。定义该泛函后，原有的边界值问题或初值问题便可改写为一项优化任务：通过在容许函数空间中最小化该泛函，求得所需的解。变分极小值对应的函数，其欧拉–拉格朗日条件能够确保目标微分方程成立。

这种视角使变分优化能够重新用作一种通用而灵活的微分方程求解器。具体而言，即使事先不知道底层的物理作用量原理，它也能运用基于梯度的优化、自动微分和受机器学习启发的工具来求解常微分方程和偏微分方程 *a priori*。从教学角度看，这种方法连接了传统的微分方程方法与现代计算优化，凸显了变分思维在物理学和应用数学中的统一作用。

1. 特征值问题：瑞利–里兹泛函

一大类物理边值问题都属于斯图姆–刘维尔理论[19]的范畴。一般的斯图姆–刘维尔微分方程为

$$\frac{d}{dx}\left[p(x)\frac{dy}{dx}\right] + \big(q(x) + \lambda\, w(x)\big)y = 0, \tag{27}$$

其中，$p(x)$、$q(x)$以及权函数$w(x)$均为定义在闭区间 $[a,b]$ 上的已知实值函数。目标是确定满足给定边界 λ 条件的特征函数$y(x)$及其对应的本征值。

方程（27）是与泛函[20]的驻值相对应的欧拉–拉格朗日方程

$$J[y] = \int_a^b \left[p(x)\left(\frac{dy}{dx}\right)^2 - q(x)y^2\right] dx, \tag{28}$$

并满足归一化约束

$$\int_a^b w(x)\, y^2\, dx = 1. \tag{29}$$

等价地，该问题也可表述为对瑞利商[21]进行无约束最小化

$$\mathscr{F}[y] = \frac{\displaystyle\int_a^b \left[p(x)\left(\frac{dy}{dx}\right)^2 - q(x)y^2\right] dx}{\displaystyle\int_a^b w(x)y^2\, dx}. \tag{30}$$

$\mathscr{F}[y]$ 的最小值对应于系统的最低本征值，而使其达到最小值的函数就是相应的基态特征函数。与斯图姆–刘维尔问题的解析解不同，解析解中的特征函数整体振幅是任意的积分常数，必须通过手动归一化来确定；而变分方法则在优化过程中通过瑞利商的分母自动实现归一化。

从计算角度来看，该过程很直接：首先初始化满足边界条件的试探函数$y_0(x)$，然后使用基于梯度的优化来最小化泛函 $\mathscr{F}[y]$，从而求得本征值–特征函数对。

这类特征值问题在量子力学中无处不在，也自然出现在经典振动系统中。为展示所提出的变分特征值求解器的通用性，我们将其应用于四种具有代表性的物理系统：经典振动弦、氢原子、氦原子，以及由质子束缚的 ^{28}Si原子核；其中核势采用含自旋–轨道耦合的伍兹–萨克森势进行建模。

振动弦

我们首先考虑经典振动弦，这是一个典型的斯图姆–刘维尔问题，具有明确的变分表述。通过在固定端边界条件下最小化弹性能泛函，可以得到平衡振型。对泛函进行离散化并应用基于梯度的优化算法后，我们求得基频振型及其对应的本征频率（见表 III 和图7(a)）。数值计算得到的本征值与解析结果高度吻合，从而验证了正向变分实现。

源代码请参见补充材料第 3 章第 3.1 节。

氢原子

作为具有精确解的量子力学示例，我们将该框架应用于氢原子。径向薛定谔方程被写成斯特姆–刘维尔形式，并结合库仑势，通过最小化瑞利商求解。所得基态能量和波函数与已知解析解高度吻合，为数值方法提供了基准测试（见表 III 和图 7(b)）。

源代码请参阅《补充材料》第 3 章第 3.2 节。

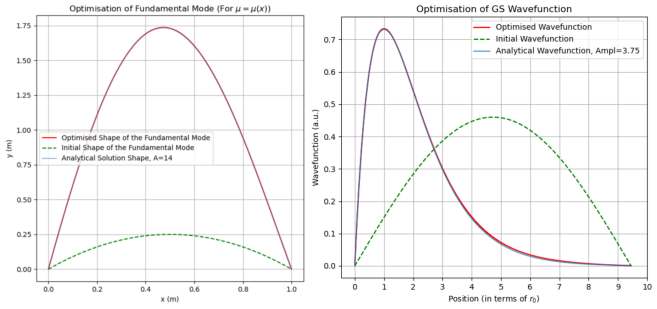


图 7：**（a）振动弦和（b）氢原子基态的瑞利–里兹优化**。试探函数（绿色虚线）弛豫至最终本征态（红色实线），与解析基准（蓝色）完全重合。空间坐标分别归一化为$L = 1$ 和玻尔半径a_0, 。

氦原子

氦原子 因电子间排斥项$e^2/4\pi\varepsilon_0 r_{12}$, 而增加了额外的复杂性，这使得薛定谔方程无法获得精确的解析解。为了使问题易于处理，我们基于两个具有物理依据的假设，采用变分近似：

1.**屏蔽：** 电子之间的库仑排斥被建模为屏蔽效应，因此每个电子感受到的有效核电荷Z_{eff} （其中 $1 < Z_{\text{eff}} < 2$）有所降低，而非完整的核电荷。

2.**可分离性：** 通过将显式的电子间相互作用替换为有效单体势函数，电子被视为彼此独立，从而使哈密顿量可分离。

在我们的计算框架中，这种近似将问题转化为一个斯特姆–刘维尔本征值方程，其无量纲参数为$p(r) = -\hbar^2/2m$、$w(r) = -1$，以及有效势$q(r) = -Z_{\text{eff}}e^2/4\pi\varepsilon_0 r$。

不同于标准的正向变分计算，我们将氦原子表述为一个逆变分问题。我们以实验测得的基态能量[23] of -79.0 eV为目标值，优化参数Z_{eff}

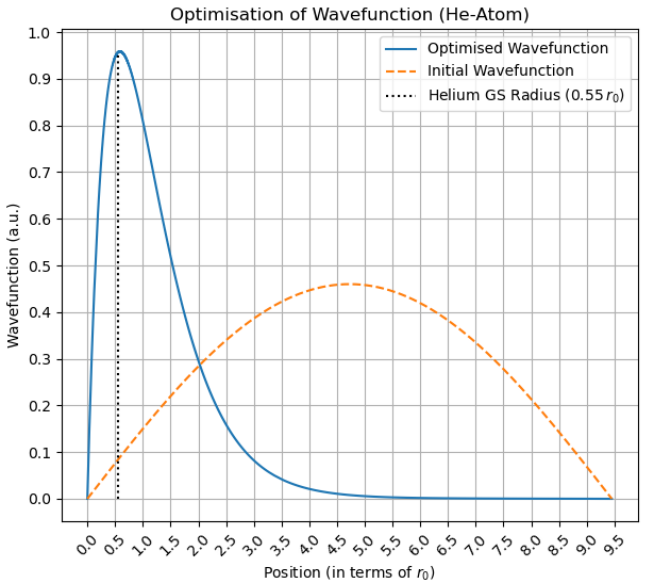


图 8：**氦原子径向波函数的逆向优化**。随着Z_{eff} 得到优化，单电子函数从试探态（橙色虚线）弛豫至物理基态（蓝色实线）。径向峰值与理论平均半径 $\approx 0.55a_0$ （垂直点线）对齐。

使计算得到的最低本征值与该能量相匹配。优化收敛至$Z_{\text{eff}} = 1.7037$，与理论基准高度吻合，较哈特里–福克估计[24] of $Z_{\text{eff}} = 1.6875$的偏差不到 1%。相应计算得到的基态能量为 -78.99eV，优化后的径向波函数如图 8 所示。

有关源代码，请参阅补充材料 第 3 章第 3.3 节。

硅原子核

作为对本征求解器的最终验证，我们计算了硅原子核 ^{28}Si（$Z = 14$）的质子空穴态。由于其闭合的 $1d_{5/2}$ 亚壳层，这为对自旋–轨道相互作用（V_{ls}）的处理提供了严格检验。

单粒子运动采用组合有效势进行描述

$$V(r) = V_{\text{WS}}(r) + V_{\text{SO}}(r) + V_{\text{CF}}(r) + V_{\text{C}}(r), \tag{31}$$

其中

$$V_{\text{WS}}(r) = \frac{-V_0}{1 + \exp\left(\frac{r-R}{a}\right)} \tag{32}$$

即为伍兹–萨克森势，这是一种示意性平均场模型，通过唯象参数V_0 （势阱深度）、a （表面弥散度）和$R = R_0 A^{1/3}$ （核半径）来描述核饱和、表面弥散度和有限束缚，其中A 为质量数，$R_0 \approx 1.2$ fm 为经验核半径常数。

$$V_{\text{SO}}(r) = V_{ls}\left(\frac{1}{r}\frac{dV_{\text{WS}}}{dr}\right)\langle \vec{L}\cdot\vec{S}\rangle \tag{33}$$

表 III：总结了经典与量子力学系统的结果。计算得到的特征值分别代表振动系统的基频和氢原子的基态能量，与理论值高度吻合。

系统	泛函系数[a]	计算得到的特征值	理论特征值	估计中的误差
经典振动弦	$p(x)=\mu(x)$[b]; $w(x)=\mu(x)$; $q(x)=0$; $\lambda=\omega_0^2$(在 弧度²秒$^{-2}$)	$\lambda_{comp}\simeq 9.7989$	$\lambda_{theo}\approx 9.7987$[c]	0。0021%
氢原子（一维径向薛定谔方程）	$p(r)=-\hbar^2/2m$; $w(r)=-1$; $q(r)=V(r)$[d]; $\lambda=E_0$（单位：eV）	$\lambda_{comp}\simeq -13.609$	$\lambda_{theo}\approx -13.60$[22]	0.0662%

a 泛函系数的值通过将系统微分方程的形式与式（27）进行比较得出 b $\mu(x)=\mu_0+(x-L/2)\Delta$ 是长度为$L=1$的弦的线性质量密度，其中 $\mu_0=0.954gm^{-1}$ 和 $\Delta=0.5gm^{-2}$ c 有关该值的推导，请参阅补充材料第 3 章 d $V(r)=-\frac{e^2}{4\pi\varepsilon_0 r}$ 是中心库仑势

是强度为的自旋–轨道耦合势V_{ls}，并且

$$V_{\mathrm{CF}}(r)=\frac{\hbar^2}{2m}\frac{l(l+1)}{r^2} \tag{34}$$

是第l个分波的离心势。式（31）中的前三项作用于质子和中子。最后一项是仅质子所受的排斥性库仑势，表示为

$$V_{\mathrm{C}}(r)=\begin{cases}\dfrac{Ze^2}{2R}\left(3-\dfrac{r^2}{R^2}\right), & r\le R,\\[2mm] \dfrac{Ze^2}{r}, & r>R.\end{cases} \tag{35}$$

为获得壳层能量，我们首先抑制自旋–轨道项（$V_{ls}=0$），以分离出中心伍兹–萨克森势函数，从而得到 1s、1p 和 1d 壳层的几何质心（图 9 左）。恢复自旋–轨道相互作用后，求解器正确再现了各轨道角动量流形内简并被解除的典型现象（图 9 右）。

计算所得的能级分裂呈现出预期的角动量依赖性：1d 壳层的分裂（$\Delta E\approx$6.7 兆电子伏）大于 1p 壳层的分裂（$\Delta E\approx$3.1 兆电子伏），符合 $(2l+1)$ 标度关系。表 IV 将计算能量与准自由（$e,e'p$）数据进行了比较。模型准确再现了接近 -46兆电子伏的深束缚 1s 态，以及能量为 -13.14 兆电子伏的价 1$d_{5/2}$ 态；后者相较实验值（-11.58 兆电子伏）略微过度束缚，这是全局参数集的已知局限。总体而言，结果证实该求解器能够捕捉中质量原子核中的壳层结构和自旋–轨道效应。

源代码请参阅补充材料 第 3 章第 3.4 节。

2. 边值问题：狄利克雷原理

稳态热传导中展现了狄利克雷原理的一个直观物理示例。当温度场$T(\mathbf{r})$达到平衡后，便与时间无关，并由热

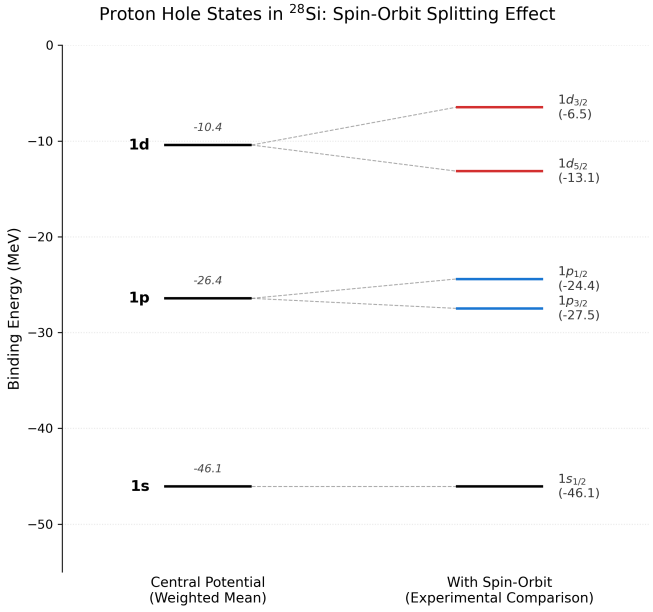


图 9：计算得到的 ^{28}Si 质子能级。（左）利用中心 Woods-Saxon 势函数（$V_{ls}=0$）计算的几何质心。（右）加入自旋轨道耦合后，简并被解除，1d 和 1p 壳层发生分裂。图中突显了较高轨道角动量下更大的分裂（$l=2$ 与 $l=1$）。

表 IV：计算得到的 ^{28}Si 质子结合能（兆电子伏）。“中心”列给出几何质心（$V_{ls}=0$），而“自旋轨道”列则包含$L\cdot S$耦合。实验数据引自参考文献 25。

态 (荷兰 谭.)	中心 ($V_{ls}=0$)	With 自旋轨道	Exp. (e,e'p)
1秒1/2	−46.07	−46.07	≈ −50
1段落3/2	−26.43	−27.48	−35 ± 4
1p1/2		−24.41	−25 ± 3
1d5/2	−10.42	−13.14	−11.58
1d3/2		−6.48	–

扩散和内部热源之间的平衡所支配。从傅里叶定律和能量守恒出发，控制方程可写为

$$\nabla \cdot (k \nabla T) + q(T) = 0, \tag{36}$$

其中，k 是热导率，$q(T)$表示单位体积内部热生成率。

根据狄利克雷原理，该边值问题的解是使能量泛函最小化的温度场

$$\mathscr{E}[T] = \int_{\Omega} \left[\frac{1}{2} k |\nabla T|^2 - Q(T) \right] dV, \tag{37}$$

其中，势函数$Q(T)$满足$dQ/dT = q(T)$。要求T的一阶变分为零，$\delta \mathscr{E}[T] = 0$在满足给定边界条件的约束下，可得欧拉–拉格朗日方程，该方程还原了上述稳态热传导方程。

这种变分形式使稳态热传导问题能够作为优化问题处理：物理上实际形成的温度分布对应于满足给定边界条件的最低能量构型。我们求解以下代表性边值问题，以说明这一思想。

均匀正方形板：精度基准测试

对于热导率均匀为 κ 且没有内部热生成的均匀正方形板，稳态温度场满足拉普拉斯方程。在二维情况下，控制方程的形式为

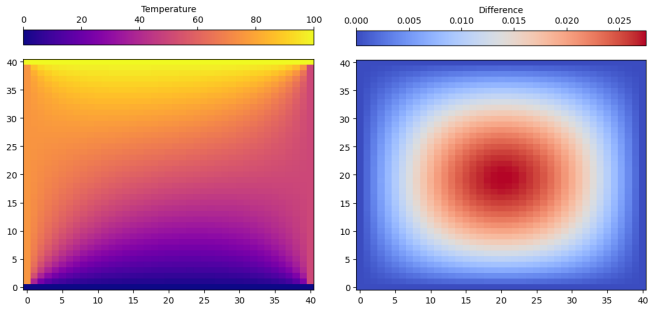
$$\nabla \cdot (\kappa \nabla T) = 0, \tag{38}$$

该方程可化简为

$$\nabla^2 T = 0 \tag{39}$$

其中 κ 为常数。该方程描述了仅由施加的边界条件驱动的纯扩散传热问题。此问题为验证求解椭圆型偏微分方程的数值方法提供了一个标准基准测试。我们将变分优化框架应用于此问题，以评估其准确性和稳健性。

从一个随机初始化且符合指定边界条件的内部温度场出发，优化器通过迭代最小化离散化的狄利克雷能量泛函，并收敛至其基态。将所得温度分布与使用传统雅可比迭代求解器[26]得到的解进行比较，结果验证了该方法的有效性，如图 10a 所示。图 10b 显示，两种解之间的差异很小，最大偏差出现在板中心附近，该处温度梯度最弱。这证实，变分方法与迭代方法得到的稳态解几乎完全一致。源代码见补充材料第 4 章第 4.1 节。



(a) (b)

图 10：**在 41 x 41 方形网格板上的温度分布模拟**，边界条件为 $u(x,\ 0) = 0$、$u(x,\ 40) = 100$、$u(0,\ y) = 75$，以及$u(40,\ y) = 50$。（**a**）使用变分求解器得到的温度分布。（**b**）变分求解器与迭代求解器得到的网格值之间的绝对差值。

非均匀杆：向变系数和源项的扩展

为展示变分方法的可扩展性，接下来考虑一根热导率随空间变化且内部存在热量生成的一维杆。当热导率 $\kappa(x)$ 沿杆变化，且存在源项$q(x)$ 时，式 36 变为非齐次泊松形式

$$\frac{d}{dx}\left(\kappa(x) \frac{dT}{dx} \right) + q(x) = 0. \tag{40}$$

在此情形下，直接采用有限差分求解会变得繁琐，因为微分算子作用于$\kappa(x)$与温度梯度的乘积。相比之下，变分形式通过最小化泛函，自然地将这一复杂性纳入其中

$$E[T] = \int_0^L \left[\frac{1}{2} \kappa(x) \left(\frac{dT}{dx} \right)^2 - q(x)\, T \right] dx. \tag{41}$$

离散化后，泛函变为

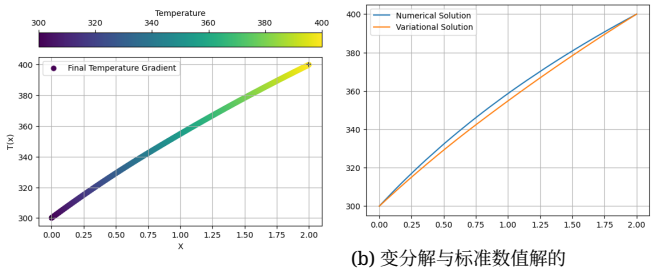
$$E_h[T] = \sum_i \left[\frac{1}{2} \kappa(x_i) \left(\frac{T_{i+1} - T_i}{\Delta x} \right)^2 - q(x_i)\, T_i \right] \Delta x. \tag{42}$$

通过基于梯度的优化最小化$E_h[T]$，可得到沿杆的稳态温度分布，与施加的边界条件一致，如图11所示。

源代码请参阅《补充材料》第4章第4.2节。

非线性微分方程

迄今为止，我们只讨论了线性微分方程。然而，此处建立的变分框架并不限于线性系统，同样适用于



(a) 杆上的温度分布由变分求解器获得。

(b) 变分解与标准数值解的解和标准数值方法解。

图11：长度为$L = 2$的杆中的热传导，其边界温度固定为$T(0) = 300$和$T(L) = 400$。热导率随位置线性变化为$\kappa(x) = \kappa_0(1 + \beta x/L)$，内部热源为$q(x) = q_0 \sin(\pi x/L)$。变分解和数值解之间的高度一致性证明了基于优化的方法具有稳健性。

非线性边值问题。作为一个代表性示例，我们考虑如下非线性方程：

$$\frac{d^2y}{dx^2} = y^3 + y, \tag{43}$$

满足适当的边界条件。

我们不直接求解该方程，而是将其解释为与泛函最小化相对应的欧拉–拉格朗日方程

$$E[y] = \int_a^b \left[\frac{1}{2}\left(\frac{dy}{dx}\right)^2 - \left(\frac{1}{4}y^4 + \frac{1}{2}y^2\right) \right] dx. \tag{44}$$

稳态解对应于函数$y(x)$，该函数在满足规定的边界条件的同时，使$E[y]$最小。

在间距为Δx的均匀网格上进行离散化后，泛函化为

$$E_h[y] = \sum_i \left[\frac{1}{2}\left(\frac{y_{i+1} - y_i}{\Delta x}\right)^2 - \left(\frac{1}{4}y_i^4 + \frac{1}{2}y_i^2\right) \right] \Delta x, \tag{45}$$

从而将非线性边值问题转化为有限维优化问题。图12比较了使用基于优化的方法获得的解与解析解。

从物理角度看，该模型可解释为具有非线性内部热源的稳态热传导，其中平衡温度分布通过最小化底层能量泛函，在扩散与非线性源项效应之间实现最佳平衡。

有关源代码，请参阅补充材料 第 4 章第 4.3 节。

3. 非保守系统：残差损失泛函

由于耗散系统不存在可供极值化的守恒量（如能量或作用量），基于最小化原理的变分

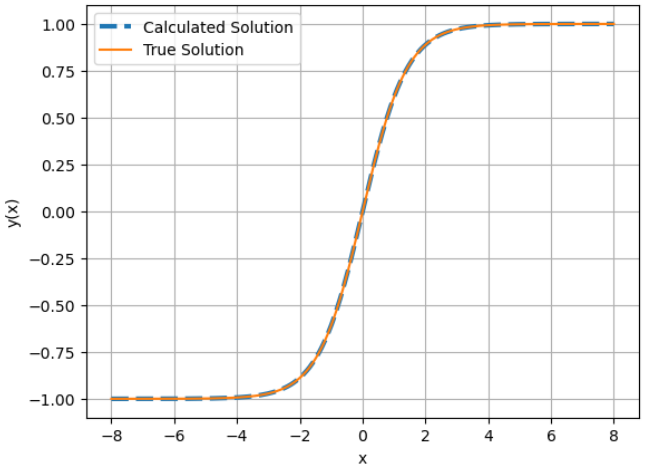


图 12：为求解此问题，我们假设边界条件为 y(-L)=-1 和 y(L)=1，可以看到，我们的解与解析解$y(x) = tanh(x\sqrt{2})$完美吻合

表述通常不可用。为解决这一局限，我们采用物理信息神经网络（*PINNs*），通过最小化由控制微分方程残差平方构造的泛函来获得解。在这一框架中，微分方程本身充当软约束，引导学习过程。

我们通过将物理信息神经网络求解器应用于两个具有代表性的**非保守动力系统**，展示了该方法的灵活性；在这些系统中，机械能会因摩擦而持续耗散。在这两种情况下，该框架不仅用于正向轨迹预测，也用于逆向参数估计。

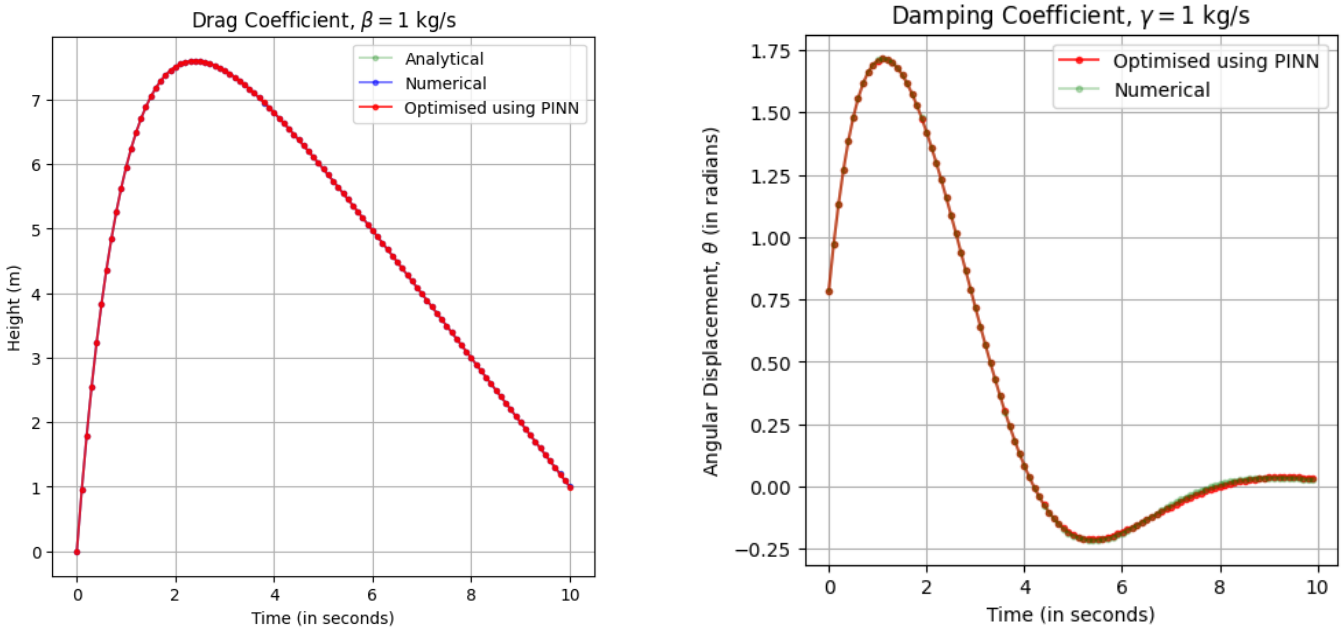
线性阻力下的抛体运动

受重力和线性空气阻力作用的粒子的垂直运动由[27]支配

$$m\ddot{y} + \beta\dot{y} + mg = 0. \tag{46}$$

这里，PINN 使用稀疏轨迹数据进行训练，同时将阻力系数 β 和初始条件视为可训练参数。如图 13a 所示，该网络成功重建了耗散运动所具有的非对称轨迹。在相应的逆问题中（表 V），学习得到的阻力系数与真实值非常接近，以高精度恢复出 $\beta_{\text{actual}} = 1.0$ kg/s $\beta_{\text{opt}} \approx 1.0015$ kg/s。

源代码请参阅补充材料 第 5 章第 5.1 节。



(a) (b)

图 13：**使用物理信息神经网络（PINNs）对非保守系统进行正向建模**。网络重建了**(a)** 受线性空气阻力作用的抛射体，以及**(b)** 阻尼单摆的时间演化。在这两种情况下，优化后的物理信息神经网络预测结果（红色/橙色）均与数值基准测试结果（绿色/虚线）高度吻合，展现了求解器无需直接监督即可捕捉耗散动力学的能力。

阻尼摆

接下来，我们考虑一个非线性耗散系统——一个长度为l、受黏性阻尼作用的摆，其描述方程为[27]

$$ml\ddot{\theta} + \gamma l\dot{\theta} + mg\sin\theta = 0. \tag{47}$$

该示例检验了求解器处理非线性回复力与依赖速度的耗散作用相结合时的稳健性。如图 13b 所示，物理信息神经网络准确捕捉了振荡振幅逐渐衰减的过程。在逆向建模任务中（表 V），网络从随机初始猜测出发，收敛到正确的阻尼系数（$\gamma \approx 1.007$ kg/s）和初始角位移（$\theta_0 \approx 0.785$ rad）。

源代码请参阅补充材料 第 5 章第 5.2 节

表 V：非保守系统的逆向建模。PINN 准确地从稀疏轨迹数据中恢复了隐藏的物理参数（摩擦系数和初始条件）。

系统	参数	实际值	优化值	Abs. 误差
抛射体带阻力的	阻力 β (kg/s)	1.0000	1.0015	0.0015
	$y(0)$（米）	0.0000	0.0001	0.0001
	$\dot{y}(0)$ (米/秒)	10.0000	10.0062	0.0062
阻尼摆	阻尼 γ (千克/秒)	1.0000	1.0071	0.0071
	$\theta(0)$ (弧度)	0.7854	0.7857	0.0002
	$\omega(0)$ (rad/s)	2.0000	2.0049	0.0049

IV. 讨论与研究范围

本文所述框架的主要优势在于其概念上的统一性。我们已经表明，同一种算法策略——对标量泛函进行最小化——能够一致地应用于各种物理系统。无论是从费马原理推导斯涅尔定律，利用最小作用量原理计算双摆的混沌轨迹，还是运用里兹–瑞利法确定 ^{28}Si 原子核的本征谱，其底层计算引擎基本保持不变。这种方法在... ...

线性和非线性系统中的适用性，使学生能够将这些问题视为同一优化原理在物理学中的多种体现，而非彼此孤立的数学练习。

与此同时，明确传统变分方法的数学边界至关重要。一个微分方程组要能由驻值原理导出——也就是说，要存在拉格朗日泛函——就必须满足亥姆霍兹条件；这些条件定义了变分法的逆问题[28]。从物理上看，这些条件将标准拉格朗日力学限制在由自伴算子支配的保守系统中。空气阻力导致的抛体运动，或阻尼摆中的摩擦等耗散现象，会引入奇数阶时间导数（例如 $\gamma\dot{x}$），从而违反这些准则。因此，

这类系统无法通过极小化物理作用量来处理，因为在传统形式体系中不存在相应的标量势。

这一局限促使我们引入物理信息神经网络（PINNs），将其作为变分范式的自然推广。通过最小化控制微分方程的残差平方，PINNs 以定义明确的损失函数取代作用量泛函，而无需辛结构或自伴性。唯一的基本要求是微分方程本身连续。这样一来，基于优化的工作流程便可顺畅地扩展至非保守系统和耗散系统。从这个意义上说，传统变分原理虽然构成自然界基本相互作用的理论基础——包括《经典力学》、电磁学和量子理论——PINNs 则完善了教学工具箱，使变分思想也能应用于以摩擦、阻尼和衰变为特征的现象学系统。

为便于在教学中采用，本文所展示的所有数值示例均使用开源 Python 工具实现，并在补充材料中提供了附有完整文档的 Python 代码。这些资源有望帮助教师复现结果、根据自身课程调整工作流程，并以极少的额外负担实施所提出的教学方法，从而支持本科物理学中可复现、以计算为驱动的教学。

V. 教学方面的说明

本研究开发的数值求解器旨在弥合解析理论与计算实践之间的鸿沟。我们不把计算视为后处理工具，而是将数值优化直接融入物理问题的建模中，使学生能够以明确且具建设性的方式学习变分原理。我们设想了三种主要方式，可将这些内容纳入本科物理学课程。

1. **统一的计算框架：** 本科物理学课程常将《经典力学》、光学和量子理论等主题划分为彼此独立的领域。通过采用统一的求解器框架，计算振动弦的本征频率、确定力学中的最小作用量轨迹，以及评估 ^{28}Si中的核壳结构，学生能够认识到物理学背后的数学普适性。这种方法鼓励学生将变分原理视为统一的算法思想，而非特定主题下的推导。

2.**科学机器学习导论：** 本文介绍的物理信息神经网络（PINN）示例为学习机器学习提供了易于理解且以物理学为导向的入门途径。与传统的数据拟合练习不同，对阻尼摆等系统进行逆向建模，展示了物理定律如何作为损失函数中的微分方程残差，发挥正则化约束的作用。这种方法使学生不仅能利用机器学习进行预测，还能发现非保守系统中的参数，而传统变分方法在这类系统中往往会失效。

3. **推荐的学生项目：**

(a) 变分法可视化： 学生可以通过最小化经典作用量，将一条初始随机试探路径迭代松弛为物理轨迹，并为这一过程制作动画。这一过程能培养对欧拉–拉格朗日方程的几何直觉，而静态的解析推导往往会掩盖这种直觉。作为一个经典基准测试，还可以用同一框架通过最小化下降时间，重新发现最速降线（摆线）。

(b) 通用特征值求解器： 学生可以利用 Ritz–Rayleigh 变分框架，通过最小化适当的能量相关量来求解特征值问题。在经典系统中，这包括最小化弹性能或应变能，以确定振动膜和梁的固有频率；还包括最小化总势能，以确定柱状结构的临界屈曲载荷。在量子力学中，该方法通过最小化哈密顿量的期望值来获得束缚态能量[29]。通过最小化伍兹–萨克森能，可以将这一方法拓展至核结构研究，探索 ^{28}Si中的壳层演化和幻数的出现。

(c)物理信息引导的发现（物理信息神经网络）：对于不满足物理作用量原理的系统，学生可以改为最小化物理信息损失函数，该函数由控制微分方程的残差平方定义。利用阻尼振子产生的含噪合成数据，或更一般的实验测量值，可以采用逆向物理信息神经网络求解器来推断摩擦系数等未知参数或其他相关系统参数。高级扩展包括耦合系统，例如洛特卡–沃尔泰拉捕食者–猎物模型[30]，，展示了物理约束如何在数据受限情形下引导学习。

VI. 结论

本文介绍了一套面向本科物理学的综合变分工具包，展示了如何调整单一的基于最小化的策略，以描述广泛的物理现象。该框架的灵活性体现在可根据具体定义域选择适当的泛函

，涵盖折射中的光程、动力系统中的作用量极小化、量子力学中的能量泛函以及稳态热传导。物理信息神经网络（PINNs）的引入进一步扩展了这套工具包，克服了传统变分力学的内在局限。经典泛函受到亥姆霍兹判据的约束，在存在耗散时会失效；而 PINNs 将微分方程残差本身视为伪泛函，从而能够对非保守系统和反问题进行变分推理。从这一角度看，现代基于梯度的优化和机器学习并非偏离经典物理学，而是费马、欧拉和拉格朗日所提出的变分思想的自然计算延伸。

致谢

作者谨向德里印度理工学院光学与光子学中心教授 Kedar Khare 教授致以诚挚谢意，感谢他提出促成本研究开展的最初构想，并在研究早期给予宝贵建议。同时，作者衷心感谢印度政府生物技术部通过 DBT明星学院计划（拨款编号：HRD-11011/20/2022-HRD-DBT）提供的经费支持。

作者声明

作者声明不存在任何利益冲突。

[1]P. Tannery 和 C. Henry，《费马著作集》*OeuvresdeFermat*（Gauthier-Villars，1891年）。[2]A. E. Shapiro，“运动学光学：十七世纪光的波动理论研究”，《精密科学史档案》**11**，134–266（1973年）。[3] C. Bailey，“动力学与变分法”，《计算机在应用力学与工程中的方法》**60**，275–287（1987年）。[4]G. M. Hulbert，“结构分析的有限元方法”，《航空航天工程百科全书》（2010年）。[5]A. G. Baydin、B. A. Pearlmutter、A. A. Radul 和 J. M. Siskind，“机器学习中的自动微分：综述”，《机器学习研究杂志》**18**，1–43（2018年）。[6]A. Griewank 和 A. Walther，《导数求值：算法微分的原理与技术》*Evaluating derivatives: principles and tech- niquesof algorithmic differentiation*（SIAM，2008年）。

[7]M. Raissi、P. Perdikaris 和 G. E. Karniadakis，“物理信息神经网络：用于求解涉及非线性偏微分方程的正向和逆向问题的深度学习框架”，《计算与应用物理学杂志》378，686–707（2019年）。[8]T. O. University，“变分法导论”（2016年）。[9]I. M. Gelfand、R. A. Silverman 等，《变分法》（Courier Corporation，2000年）。[10] R. Scratchpad，“反向模式自动微分：教程”（2016年），最后访问于2025年10月6日。[11]M. Bartholomew-Biggs、S. Brown、B. Christianson 和 L. Dixon，“算法的自动微分”，《计算与应用数学杂志》124，171–190（2000年）。[12]T. J. authors，“JAX 101：自动微分”（2024年），最后访问于2025年10月11日。[13]A. Tapkir，“梯度下降及其优化算法综述”，《国际科学、工程与技术研究杂志》10，37–45（2023年）。[14]F. Pioch、J. H. Harmening、A. M. Müller、F.-J. Peitzmann、D. Schramm 和 O. e. Moctar，“物理信息神经网络的湍流建模：后向台阶流中不同雷诺平均纳维－斯托克斯模型的比较”，《流体》8，43（2023年）。[15]A. Farea、O. Yli-Harja 和 F. Emmert-Streib，“理解物理信息神经网络：技术、应用、趋势与挑战”，《人工智能》5，1534–1557（2024年）。[16]J. Seo，“使用物理信息神经计算求解现实世界优化任务”，《科学报告》14，202（2024年）。[17]P. Foundation，“PyTorch 文档”（2023年），最后访问于2025年10月25日。[18]S. J. Greydanus、T. D. Strang 和 I. Caruso，“自然的代价函数：通过最小化作用量模拟物理学”，载于 ICLR 2023年物理与机器学习研讨会（2023年）。[19]J. Lützen，“Sturm 和 Liouville 关于常微分线性方程的工作：斯图姆－刘维尔理论的出现”，《精密科学史档案》29，309–376（1984年）。[20]G. B. Arfken、H. J. Weber 和 F. E. Harris，《物理学家数学方法：综合指南》（Academic Press，2011年），第1072–1074页。[21]A. Basilevsky，《统计科学中的应用矩阵代数》（Courier Corporation，2013年），第451–453页。[22] R. Eisberg 和 R. Resnick，《原子、分子、固体、原子核与粒子的量子物理》（1985年）。[23]C. L. Pekeris，“双电子原子的基态”，《物理评论》112，1649（1958年）。[24]D. J. Griffiths 和 D. F. Schroeter，《量子力学导论》（剑桥大学出版社，2018年）。[25]J. Mougey、M. Bernheim、A. Bussiere、A. Gillebert、P. X. Ho、M. Priou、D. Royer、I. Sick 和 G. Wagner，“关于 12c、28Si、40ca 和 58ni 的准自由（e, e’p）散射”，《核物理A》262，461–492（1976年）。[26]S. C. Chapra、R. P. Canale 等，《工程师数值方法》，第1221卷（McGraw-Hill，纽约，2011年）。[27]J. Taylor，《经典力学》，G—— 参考、信息与跨学科主题丛书（University Science Books，2005年）。[28]K. Nigam 和 K. Banerjee，“亥姆霍兹条件简述”（2016年），arXiv:1602.01563 [math-ph]。[29]D. V. Schroeder，“用于求解量子束缚态的变分松弛算法”，《美国物理学杂志》85，698–704（2017年）。[30]M.-C. Anisiu 和 R. Beattie，“洛特卡、沃尔泰拉及其模型”，载于《DIDACTICA MATHEMATICA》第32卷（2014年），第9–17页。