引力及其他物理定律的奖励函数与最小成本原理

Rubén Moreno-Bote[1-3,*]

[1]庞培法布拉大学工程系，08002，西班牙巴塞罗那[2]庞培法布拉大学脑与认知中心，08002，西班牙巴塞罗那[3]庞培法布拉大学 Serra Húnter研究员计划，西班牙巴塞罗那，08002[*]通讯作者：ruben.moreno@upf.edu

arXiv:2603.25444v1 [q-bio.NC] 26 Mar 2026

摘要

如果宇宙遵循某种特定设计，那么一个核心问题便是观测到的力所优化的代价函数是什么。这就是逆最优控制或逆强化学习问题：从观测系统的动力学中推断奖励函数。我们首先确立最小代价原理，即运动定律可以通过最小化加速度代价减去状态依赖奖励函数的时间折扣积分而导出。在依据基本原理确定加速度代价的函数形式后，我们根据经典引力和库仑力所遵循的运动定律推断奖励函数。粒子对的相对速度越高，且其相对运动与距离向量越接近正交时，推断出的奖励函数值越高。总而言之，我们的研究表明，自然界中的中心力所优化的动力学特征和静态特征，分别是相对运动和准圆轨道。

如果宇宙具有智能设计，那么一个核心问题便是它最小化的代价函数是什么。代价函数——或者更自然地说，奖励函数——可能赋予宇宙某些“理想”特征，供自然力加以优化。一个自然的目标或许是产生丰富而结构化的运动。不过，这种运动应当在相互作用力尽可能少的干预下产生。即使对于牛顿引力或库仑力等经典相互作用，其代价函数和奖励函数的形式也尚未确定。

在这里，我们使用逆最优控制（IOC）[1–4]，从系统动力学中推导奖励函数。逆强化学习中的这些方法及相关方法，已被用于在线性[1,5,6]和非线性[2,7,8]控制问题（包括脉冲控制[9],）中推断代价函数，并广泛应用于根据观察到的动物行为推断奖励函数[10,11]。然而，IOC 受到所采用控制代价的复杂性，以及关于时间范围和时间折扣的假设的限制。

本文确立了最小成本原理。该原理表明，在具有中心力的经典系统中，例如牛顿引力和库仑力，其运动定律源于一个最优控制问题[12,13]的最优解；该问题旨在最小化加速度代价减去奖励函数的时间折扣积分。最小成本原理与最小作用量原理[14]，相似，但前者是一个最优控制问题，而后者是一个最优路径问题[12]。两种原理都会导出相同的运动定律，正如不可数多个不同的抛物线具有相同的最小值；它们各自的价值在于各自框架所提供的独特见解。

我们从基本原理出发——而非为了数学上的便利而将其作为假设[1,2,13]——证明，与最小成本原理相关的控制代价是二次的，其形式为对各粒子的加速度平方求和。运用最小成本原理和推导出的加速度代价进行 IOC，我们得到了牛顿引力和库仑力对应的未知奖励函数的形式。我们发现，推断出的奖励函数会促成两个显著特征：当粒子对之间的相对运动较强时，奖励为正；当相对运动方向不与其距离向量正交时，奖励为负。这些效应受粒子间距离的反向调节。

总而言之，我们揭示了自然力所优化的两个显著特征：相对运动和类圆轨道。优化这些特征或许能为宇宙的复杂性涌现提供基本要素。运动及其多样性的最大化，也是其他原理的核心思想，例如最大占据原理[15,16]和赋权[17]。本文表明，运动是在自然界中涌现出来的一种优化量，而非预先假定的原理。我们对奖励函数所作的定量刻画，也远远超越了将引力和其他中心力定性描述为在粒子对之间产生“吸引”或“排斥”的说法；我们证明，自然力会主动促成相对运动和类圆轨道的形成。

最小成本原理——我们将轨迹的时间折扣累积成本定义为$x_{ik}(t)$、 $t \geq t_0$、 $i = 1, \cdots, N$个具有笛卡尔坐标 k的粒子的成本

$$\mathcal{C} = \int_{t_0}^{\infty} dt\; e^{-\gamma(t-t_0)} \left[\sum_i \frac{1}{2} m_i \left\| \ddot{x}_i(t) \right\|^2 - R(x(t), \dot{x}(t)) \right] , \tag{1}$$

初始坐标 $x_0 = x(t_0)$和速度 $\dot{x}_0 = \dot{x}(t_0)$固定，其中我们使用记号$x = (x_1, \cdots, x_N)$，并令 $x_i = (x_{i1}, x_{i2}, x_{i3})$。这里， $\gamma > 0$是时间折扣指数，而项 $C_a(\ddot{x}) = \sum_i \frac{1}{2} m_i \left\| \ddot{x}_i(t) \right\|^2$是加速度（控制）成本。状态奖励函数 $R(\bar{x}, \dot{x})$取决于系统状态，其中包括粒子的坐标 x 及其速度 $\dot{x}$。重要的是，奖励并不显式依赖于加速度 $\ddot{x}$；对加速度的唯一依赖体现在加速度成本中。加速度成本对较大的加速度施加惩罚，从而抑制强力。我们将 $\mathcal{C}$ 简称为

剩余成本，并假设其有限。

剩余成本 $\mathcal{C}$ 是轨迹的泛函。运动方程由求得加速度轨迹 $\ddot{x}^*(t)$， $t \geq t_0$来确定；该轨迹在固定初始条件 x_0 和 $\dot{x}_0$下使剩余成本最小，

$$\mathcal{C}^*(x_0, \dot{x}_0) = \min_{\ddot{x}} \mathcal{C}, \quad \ddot{x}^* = \arg\min_{\ddot{x}} \mathcal{C} \,, \tag{2}$$

其中 $\mathcal{C}^*$ 是最小（最优）剩余成本；根据定义，它与时间无关，因此也独立于 t_0。

公式 1-2 定义了最小成本原理。因此，推导运动定律归结为一个最优控制问题[12,13,18]。实质上，这个问题要求在使用尽可能弱的力的同时，最大化奖励函数——其中包含所需的动力学和静态特征。最小成本原理与最小作用量原理[14]有关，但并不相同。作用量是拉格朗日量的泛函，与代价函数类似，但它并不显式依赖加速度，这与公式 1 中的剩余成本不同。

公式 1 中加速度代价的形式由对称性原理和不变性唯一确定（详见附录1）。首先，我们要求加速度代价在时间和旋转下保持不变，并且对粒子及其质量具有可加性。其次，我们要求内力引起的加速度代价变化在均匀加速参考系的选择下保持不变。这两组条件足以唯一确定加速度代价 $C_a(\ddot{x})$，其形式与公式 1 完全相同。在最优控制问题中，通常出于数学上的便利而假设采用二次成本项[12,13,18]。在这里，我们证明了这种形式源自封闭系统中的第一性原理。

我们注意到，当奖励函数为零时，剩余成本简化为$\mathcal{C} = \int_0^\infty dt e^{-\gamma t} \sum_i \frac{1}{2} m_i \left\| \ddot{x}_i(t) \right\|^2$。根据公式 2 最小化此剩余成本，便可得到运动方程 $\ddot{x}_{ik} = 0$。其解对应于每个 N 粒子以恒定速度做直线运动。因此，这个最优控制问题对应于 N 独立粒子。

一般力的逆最优控制——我们的目标是在已知运动定律的情况下，确定公式 1 中未知的奖励函数。我们表明，对于一般的力，在时间折扣剩余成本中推断奖励函数是可行的。

通过启发式地将时间划分为微分步长 dt，并将指数项和下一步的最优成本按 dt、 dx 和 $d\dot{x}$ 的一阶项展开，公式 1 中的最优剩余成本可以表示为贝尔曼方程的形式：

$$\begin{aligned} \mathcal{C}^* = \min_{\ddot{x}} \Bigg\{ & \left(\sum_{ik} \frac{1}{2} m_i \ddot{x}_{ik}^2 - R(x, \dot{x}) \right) dt \\ & + (1 - \gamma dt) \left(\mathcal{C}^* + \sum_{ik} \frac{\partial \mathcal{C}^*}{\partial x_{ik}} dx_{ik} + \sum_{ik} \frac{\partial \mathcal{C}^*}{\partial \dot{x}_{ik}} d\dot{x}_{ik} \right) \Bigg\} \,, \end{aligned} \tag{3}$$

其中 $\mathcal{C}^* = \mathcal{C}^*(x, \dot{x})$。使用恒等式 $dx = \dot{x} dt$ 和 $d\dot{x} = \ddot{x} dt$后，对加速度进行最小化即可得到最优运动方程

$$m_i \ddot{x}_{ik} = -\frac{\partial \mathcal{C}^*}{\partial \dot{x}_{ik}} \,, \tag{4}$$

以及哈密顿-雅可比-贝尔曼方程，该方程明确地将未知奖励函数表示为最优剩余成本的函数，形式如下

$$R(x, \dot{x}) = \sum_{ik} \left[\frac{1}{2m_i} \left(\frac{\partial \mathcal{C}^*}{\partial \dot{x}_{ik}} \right)^2 + \frac{\partial \mathcal{C}^*}{\partial x_{ik}} \dot{x}_{ik} + \frac{\partial \mathcal{C}^*}{\partial \dot{x}_{ik}} \ddot{x}_{ik} \right] - \gamma \mathcal{C}^*. \tag{5}$$

鉴于已知 wn 力 $F_{ik}(x)$，根据 $m_i\ddot{x}_{ik} = F_{ik}$ 和公式 4，最优剩余成本变为 s

$$\mathcal{C}^*(x,\dot{x}) = -\sum_{ik} F_{ik}(x)\dot{x}_{ik} + B(x)\ , \tag{6}$$

其中 $B(x)$是一个与速度 $\dot{x}$无关的任意函数。对于 $B(x) = 0$，最优剩余成本是功的时间导数的负值，$\frac{dW}{dt} = -\sum_{ik} F_{ik}(x)\dot{x}_{ik}$。若能找到原函数，推广到速度相关力的情形就直接明了。将公式 6 的解代入公式 5 后，未知的奖励函数可写为

$$\begin{aligned} R(x,\dot{x}) = &-\sum_{ik,jp}\dot{x}_{ik}\dot{x}_{jp}\frac{\partial F_{ik}}{\partial x_{jp}} - \frac{1}{2}\sum_{ik}\frac{F_{ik}^2}{m_i}\ , \\ &+\sum_{ik}\dot{x}_{ik}\left(\gamma F_{ik} + \frac{\partial B}{\partial x_{ik}}\right) - \gamma B(x) \end{aligned} \tag{7}$$

这是我们的第二个核心结果。若力是保守力，且势能为 $U(x)$，那么第一项可改写成更对称的形式：$-\sum_{ik,\ jp}\dot{x}_{ik}\dot{x}_{jp}\frac{\partial^2 U}{\partial x_{ik}\partial x_{jp}}$。

正如我们将在下文看到的，公式 7 中的第一项包含了有利于高速度的贡献。第二项显然会对强力施加惩罚。

我们的推导表明，在公式 7 所示的奖励函数下，运动方程 $m_i\ddot{x}_{ik} = F_{ik}$ 会使公式 1 中的剩余成本最小化。这一点也可以明确证明（详见附录 2）：由动力学 $m_i\ddot{x}_{ik} = F_{ik}$ 生成、且 F_{ik} 不同于 F_{ik} 的轨迹˜ ˜，在奖励函数如公式 7 所示时，会使公式 1 中的剩余成本不低于最小值。

现在我们将式 7 专门应用于两个重要情形：

i) 经典引力的奖励函数 — 粒子 i受到由 N 个粒子组成的系统所施加的总引力为

$$F_{ik}(x) = -\sum_{j,\ j\neq i}\frac{Gm_im_j}{r_{ij}^3}(x_{ik} - x_{jk})\ , \tag{8}$$

其中 $r_{ij} = \|x_i - x_j\|$。经过一些直接但冗长的计算，式 7 中的奖励函数在 $B(x) = 0$ 时变为

$$\begin{aligned} R(x,\dot{x}) = &\frac{G}{2}\sum_{i\neq j}\frac{m_im_j}{r_{ij}^3}\|\dot{x}_i - \dot{x}_j\|^2 && \textit{(I)} \\ &-\frac{3G}{2}\sum_{i\neq j}\frac{m_im_j}{r_{ij}^5}[(x_i - x_j)\cdot(\dot{x}_i - \dot{x}_j)]^2 && \textit{(II)} \\ &-\frac{G^2}{2}\sum_{i\neq l, i\neq s}\frac{1}{m_i}\frac{m_i^2m_lm_s}{r_{il}^3r_{is}^3}(x_i - x_l)\cdot(x_i - x_s) && \textit{(III)} \end{aligned} \tag{9}$$

这里省略了与 γ 成正比的项，因为该项可以任意小。最后一个方程和下一个方程中的求和均遍历所有相关指标。为方便后续计算，质量没有约去。最优剩余成本由式 6 给出

$$\mathcal{C}^*(x,\dot{x}) = \frac{G}{2}\sum_{i\neq j}\frac{m_im_j}{r_{ij}^3}(x_i - x_j)\cdot(\dot{x}_i - \dot{x}_j)\ . \tag{10}$$

正如封闭系统所预期的那样，奖励函数和最优剩余成本都只取决于相对位置和速度。

根据方程 6–7，最优剩余成本和奖励函数的定义都只确定到一个任意函数 $B(x)$。因此，可以选择 $B(x)$，使得奖励函数中可以加入一个仅依赖于 x 的任意函数，以及一个关于 $\dot{x}$ 的任意线性项。这意味着只有项 I 和项 II 是本质性的，因为它们是速度的二次函数，因此无法并入 $B(x)$ 的贡献中。此外，项 III 源于剩余成本中存在推导得到的二次控制代价，而项 I 和项 II 则具有更内在的性质。

首先，项 I 的奖励为非负值，取决于成对粒子的速度向量之差的范数。因此，当粒子间的相对运动较强时，这一奖励项较大。粒子间的距离也会调节这一效应：短距离下会促进较强的相对运动，而距离较远时，奖励函数则趋于中性。

其次，项 II 的奖励为非正值，取决于成对粒子的距离向量与相对速度向量之间的点积。当一对粒子之间的相对运动与其距离向量垂直时，该奖励的成对贡献为零。因此，项 II 会促使粒子彼此围绕对方做类圆周运动。距离较远时，非圆周运动不会受到惩罚；而距离较近时，这种运动则代价较高。

数值模拟证实了我们的分析结果。正如预期的那样[14],，生成的轨迹大多是曲线，粒子彼此靠近时往往会形成类圆轨道（图1a； $N = 5$ 个粒子）。项 I 和项 II 的奖励贡献随时间波动（图 1b），其中项 I 始终为正，项 II 始终为负。大部分轨迹上的最小代价到达函数都相对较小，但其间夹杂着半周期性的正峰和负峰（图1c）。我们验证了，对剩余成本进行数值积分（红线，不可见）与公式 10 中解析推导的剩余成本（黑线）相符。对于包含 $N = 10$ 个粒子的更大系统，剩余成本的波动有所增加，但与包含 $N = 5$个粒子的系统相比，在更大比例的轨迹上，剩余成本仍保持较小。若粒子受到非牛顿力 $1/r^{2+\epsilon}$ 作用，剩余成本会增加（图1e）。当 $\epsilon = 0$时，最小代价到达函数达到最低值，这对应于纯粹的牛顿引力（图 1f）。

ii) 库仑力的奖励函数— 对于库仑静电力，奖励函数与公式 9 相同，只需将 G 替换为 $-C_{Coulomb}$，并将分子中的质量 m_i 替换为电荷 q_i，而分母中的质量保持不变。对于异号电荷，适用与引力中项 I 和项 II 相同的规则。然而，对于同号电荷，这些特征则相反：相对运动被最小化，粒子的运动倾向于与其距离向量平行，无论粒子彼此靠近还是远离，运动强度都相同。

与牛顿引力的情况一样，如果库仑力从 $1/r^2$ 扰动至 $1/r^{2+2\epsilon}$，则产生的剩余成本高于未受扰动的情况 $\epsilon = 0$ （图 1f）。

讨论 — 我们首先确立最小代价原理，并从第一性原理推导出加速度代价的形式，由此推断出经典力所优化的奖励函数。我们的分析揭示了牛顿引力、库仑相互作用以及更一般的力所优化的动力学和静态特征。具体而言，相对运动和类圆轨道受到青睐。可以说，这些特征是构建一个“有趣”宇宙的两个关键要素，因为结构化运动构成了复杂性涌现的基础。

选择非零的时间折扣指数 $\gamma < 1$ 具有多重影响。首先，这使数学处理更加可行，并避免在寻找时不变奖励函数[1,2]时进行繁琐的推导。其次，非零的时间折扣隐含地假设系统可能蒸发或解体，而这一假设是合理的。第三，假设$\gamma < 1$ 强调瞬态而非平稳分布；如果

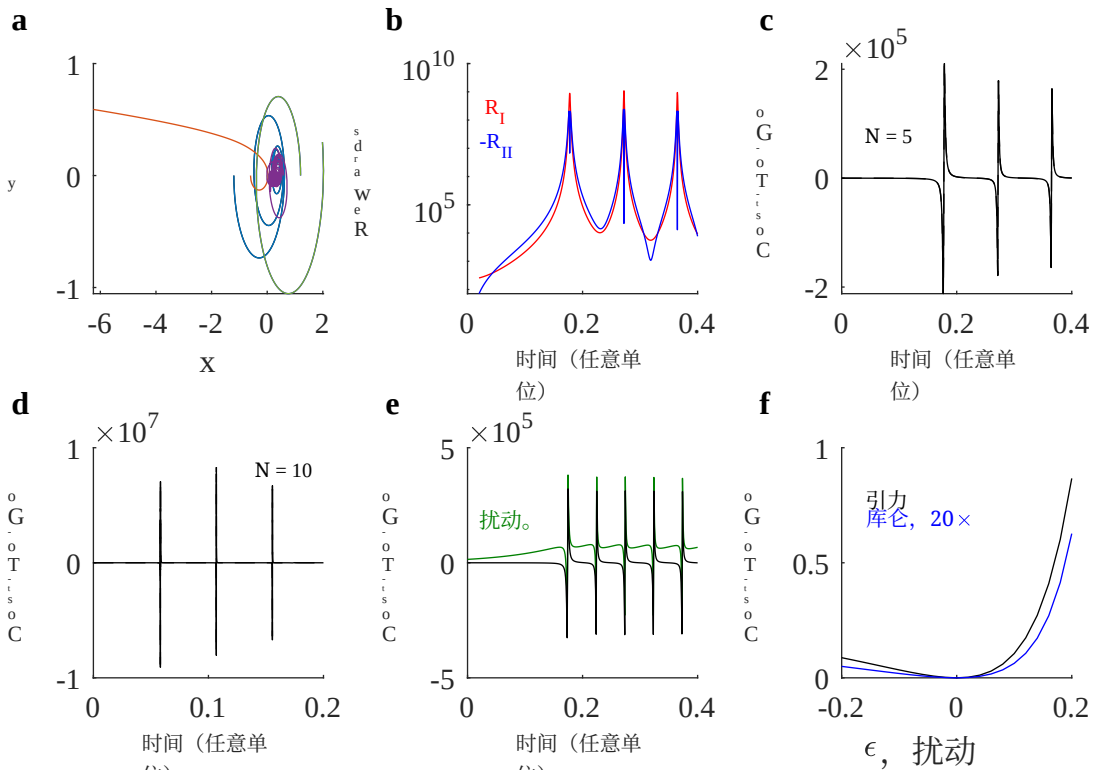

图1：牛顿引力使公式 9 中的奖励函数达到最大值。（**a**）五个粒子遵循牛顿引力运动的轨迹。（**b**）公式 9 中项 I 和项 II 的奖励贡献随时间的变化。由于奖励项 II 非正，因此图中显示其绝对值。（**c**）轨迹上各状态的公式 10 中最小剩余成本随时间的变化（黑线）。这与通过公式 1 数值计算得到的轨迹实际产生的成本一致（红线；不可见）。（**d**）随着粒子数量增加，$N = 10$，剩余成本的波动幅度更大，但处于相对较小值的时间更长。（**e**）将牛顿力从 $1/r^2$ 扰动至 $1/r^{2+\epsilon}$，并加入 $\epsilon = 0.05$，会产生高于最优值（黑线）的剩余成本（绿线）。剩余成本是在受扰动力产生的轨迹上逐个状态计算的。（**f**）受扰动的牛顿力 $1/r^{2+\epsilon}$ 产生的剩余成本（黑线）随 ϵ的变化，计算时使用与面板（b）相同的初始条件。剩余成本在 $\epsilon = 0$处达到最小值。受扰动的库仑力 $1/r^{2+2\epsilon}$ 产生的剩余成本（蓝线）随 ϵ的变化。参数详见附录 3。

宇宙的预期寿命是有限的。相比之下，采用 $\gamma = 1$ 意味着后续成本由渐近稳态分布主导；也就是说，时间范围会长到足以使瞬态效应变得可以忽略。尽管强化学习中有时会采用这种方法[19,20]，，但它可能会阻碍跨多个时间尺度的动力学复杂性涌现。

最后需要指出，奖励项 I 和 II 源自式 7 中奖励函数一般表达式的第一项，该项关于力呈线性。一个重要推论是，如果已知一组力中每个力所优化的奖励函数，那么这些力之和所优化的奖励函数就对应于各个奖励函数之和。进而，这意味着我们推导出的运动和循环性特征，即使在系统中加入新的力后，仍然适用。

致谢— 感谢 Dmytro Grytskyy 发现方程中的拼写错误。本项目获得了西班牙科学、创新与大学部资助的拨款（MICIU/AEI/10.13039/501100011033），以及加泰罗尼亚研究与高等研究院资助的 ICREA ACADÈMIA（2022）的支持。

参考文献

1. Kalman, R.E. (1964). 线性控制系统何时是最优的？Basic Eng, 86(1): 51-60. 2. Moylan, P., and Anderson, B. (1973). 非线性调节器理论与一个逆最优控制问题。IEEE 自动控制汇刊*18*, 460–465. 3. Casti, J. (1980). 关于最优控制理论的一般逆问题。优化理论与应用杂志*32*, 491–497. 4. Ab Azar, N., Shahmansoorian, A., and Davoudi, M. (2020). 从逆最优控制到逆强化学习：历史回顾。控制年度评论*50*, 119–138. 5. Wu, H., Hu, Q., Zheng, J., Shao, X., Liu, Y., and Li, D. (2026). 利用折扣逆强化学习对线性连续时间系统进行输出反馈控制。IEEE 控制论汇刊。6. Xue, W., Kolaric, P., Fan, J., Lian, B., Chai, T., and Lewis, F.L. (2021). 基于逆最优控制的跟踪控制逆强化学习。IEEE 控制论汇刊*52*, 10570–10581. 7. Ziebart, B.D., Maas, A.L., Bagnell, J.A., Dey, A.K. et al. (2008). 最大熵逆强化学习。载于 Aaai 第8卷。美国伊利诺伊州芝加哥，第1433–1438页。8. Xu, Y., Xie, J., Zhao, T., Baker, C., Zhao, Y., and Wu, Y.N. (2022). 基于能量的连续逆最优控制。IEEE 神经网络与学习系统汇刊*34*, 10563–10577. 9. Wang, W., Li, C., and Huan, M. (2025). 基于拟变分不等式的动力系统最优脉冲控制与稳定化。Mathematics*13*, 3864. 10. Drugowitsch, J., Moreno-Bote, R., Churchland, A.K., Shadlen, M.N., and Pouget, A. (2012). 知觉决策中证据累积的代价。Journal of Neuroscience*32*, 3612–3628.

11. Rothkopf, C.A., and Ballard, D.H. (2013). 视觉运动行为的模块化逆强化学习。《生物控制论》*107*, 477–490. 12. Kirk, D.E. (2004). 最优控制理论：导论。Courier Corporation. 13. Fleming, W.H., and Soner, H.M. (2006). 受控马尔可夫过程与黏性解。施普林格。14. Landau, L.D., and Lifshits, E. (1960). 《力学》第1卷。CUP Archive. 15. Moreno-Bote, R., and Ramirez-Ruiz, J. (2023). 赋权、自由能原理与最大占据原理的比较。载于 NeurIPS 2023 研讨会：认知系统中的信息论原理。16. Ramirez-Ruiz, J., Grytskyy, D., Mastrogiuseppe, C., Habib, Y., and Moreno-Bote, R. (2024). 从内在动机到占据未来动作—状态路径空间的复杂行为。《自然·通讯》*15*, 6368. 17. Klyubin, A.S., Polani, D., and Nehaniv, C.L. (2005). 赋权：以智能体为中心的通用控制度量。载于2005年IEEE进化计算大会第1卷。IEEE，第128–135页。18. Lewis, F.L., and Vrabie, D. (2009). 用于反馈控制的强化学习与自适应动态规划。《IEEE电路与系统杂志》*9*, 32–50. 19. Hazan, E., Kakade, S., Singh, K., and Van Soest, A. (2019). 有可证明效率的最大熵探索。载于国际机器学习会议。PMLR，第2681–2691页。20. Grytskyy, D., Ramírez-Ruiz, J., and Moreno-Bote, R. (2023). 用于动作—状态熵正则化奖励最大化的一般马尔可夫决策过程形式化框架。arXiv预印本 arXiv:2302.01098.

附录

1 N 独立粒子的加速度代价

一个根本问题是，为什么公式 1 中的加速度代价具有如下形式

$$C_a(\ddot{x}) = \sum_i \frac{1}{2} m_i \left\| \ddot{x}_i \right\|^2 \ . \tag{11}$$

例如，原则上也可以使用 $\|\ddot{x}_i\|^n$ 次幂，其中 $n \neq 2$ 。事实上，这些形式都会导出相同的直线运动方程 $\ddot{x}_{ik} = 0$。接下来，我们将说明推导公式 11 中加速度代价形式的充分条件。

首先，由于时间齐次性，加速度代价不能显式依赖于时间。其次，对于 N 个独立粒子，由于可加性，加速度代价应当是各单位质量粒子的加速度代价之和 C_a（$\ddot{x}_i$），

$$C_a(\ddot{x}) = \sum_i m_i C_a(\ddot{x}_i) \ , \tag{12}$$

其中 $\ddot{x}_i$ 是粒子 i的加速度向量。第三，由于空间各向同性，单位质量单粒子的加速度代价应当是加速度模的平方的函数，也就是说，C_a（$\ddot{x}_i$）$= C_a$（$\|\ddot{x}_i\|^2$）。需要指出的是，当我们对不同函数使用同一个符号时，这种记法并不严谨。我们采用如下记号约定：不同的自变量表示函数可能不同。

最后，我们要求由内力引起的加速度代价变化在不同参考系之间保持不变。考虑一个由 N 个粒子组成的系统，以及两个不同的参考系。粒子最初互不相互作用，其加速度为 $\ddot{x}_i$。例如，这些加速度可能由引力等外场引起。现在，假设在某一时刻 t，粒子之间通过相互作用力发生相互作用。这种相互作用导致加速度从 $\ddot{x}_i$ 突然变为 $\ddot{x}'_i = \ddot{x}_i + b_i$，其中 b_i 是粒子 i的加速度变化。我们要求，力突然开始作用前后的加速度代价变化不受所选参考系影响。对于任何惯性参考系，这一点显然成立，因为无论选择哪个参考系，所有粒子的加速度都相同。因此，我们提出更严格的条件：无论选择何种均匀加速参考系，加速度代价的变化都应保持不变。例如，我们的粒子系统的加速度代价变化在引力场内外都应相同。正如我们接下来将要展示的，这一条件足以完全确定加速度代价的函数形式，即公式 11。

在第一个参考系中，系统粒子的加速度在内力突然开始作用后由 $\ddot{x}_i$ 变为$\ddot{x}'_i = \ddot{x}_i + b_i$。因此，加速度代价的变化为

$$\Delta C_a = \sum_i m_i (C_a(\|\ddot{x}'_i\|^2) - C_a(\|\ddot{x}_i\|^2) \ . \tag{13}$$

在第二个参考系中，存在一个额外的均匀加速度 $a(t)$，其第 k个分量为 $a_k(t)$；所有粒子的加速度矢量都相同，而且可以随时间变化。在这个参考系中，内力突然出现后，粒子 i 的加速度由 $\ddot{x}_i + a(t)$变为 $\ddot{x}_i + b_i + a(t)$。因此，加速度代价的变化为

$$\Delta \tilde{C}_a = \sum_i m_i (C_a(\|\ddot{x}'_i + a\|^2) - C_a(\|\ddot{x}_i + a\|^2) \ . \tag{14}$$

˜因此，我们的不变性条件要求恒等式 $\Delta C_a = \Delta C_a$成立，因为加速度代价的变化在任意均匀加速参考系中都应相同。

将最后一个方程按加速度 a展开至一阶，

$$\Delta\tilde{C}_a = \Delta C_a + \sum_i m_i \left(\frac{d}{dy} C_a(\|\ddot{x}_i'\|^2)\, \ddot{x}_i' - \frac{d}{dy} C_a(\|\ddot{x}_i\|^2)\, \ddot{x}_i \right) \cdot a + O(\|a\|^2)\,, \tag{15}$$

其中，$\frac{d}{dy} C_a(\|\ddot{x}_i\|^2)$ 是实值函数 $C_a(y)$ 在 $\|\ddot{x}_i\|^2$ 处的导数。不变性意味着一阶项 a 为零。若导数 $\frac{d}{dy} C_a(\|\ddot{x}_i\|^2)$ 为常数，则该条件成立，因为根据牛顿第二定律，在封闭系统中 $\sum_i m_i b_i = 0$。因此，我们得到 $C_a(\|\ddot{x}_i\|^2) = \frac{1}{2}\|\ddot{x}_i\|^2$；取标准因子 $\frac{1}{2}$，便可得回公式 11 中 N 个粒子的加速度代价。

很容易看出，公式 11 中加速度代价的变化在任何均匀加速参考系中不仅在 $a(t)$ 的一阶近似下保持不变，而且在所有阶上都严格不变。

此外，不存在其他可微函数 $C_a(y)$ 能满足不变性。通过考察两个粒子 $N = 2$ 的情形，可以更简单地看出这一点。我们定义函数

$$\frac{d}{dy} C_a\, \|\ddot{x}_i\|^2\, \ddot{x}_{ik} \qquad 15 \qquad m_1\, g_k\, \ddot{x}_1' - g_k\, \ddot{x}_1 + (\ddot{x}_i)$$

( )。公式的一阶消失条件意味着 (( ) ( )) $m_2(g_k(\ddot{x}_2') - g_k(\ddot{x}_2)) = 0$ 对所有 k 都成立，这意味着 $m_2/m_1 = -(g_k(\ddot{x}_1') - g_k(\ddot{x}_1))/(g_k(\ddot{x}_2') - g_k(\ddot{x}_2))$。（解 $g_k(\ddot{x}) = d$，其中 d 为常数，不允许采用，因为这会导致零点处的代价不可微；另一种情况则是平凡解，即对于 $d = 0$，加速度代价为常数。）牛顿第二定律意味着 $m_2/m_1 = -b_{1k}/b_{2k} = -(\ddot{x}_{1k}' - \ddot{x}_{1k})/(\ddot{x}_{2k}' - \ddot{x}_{2k})$。由于函数 g_k 应当与质量无关，令两边相等可得 $(g_k(\ddot{x}_1') - g_k(\ddot{x}_1))/(g_k(\ddot{x}_2') - g_k(\ddot{x}_2)) = (\ddot{x}_{1k}' - \ddot{x}_{1k})/(\ddot{x}_{2k}' - \ddot{x}_{2k})$。对于任何与正质量条件相容的坐标取值，这一条件都应成立。例如，可以令所有 k $\ddot{x}_{1k} = c$ 任意取值，$\ddot{x}_{1k}' > c, \ddot{x}_{2k} = 1$ 和 $\ddot{x}_{2k}' = 0$，这意味着 $g_k(\ddot{x}_1') = g_k(c) + (g_k(1) - g_k(0))(\ddot{x}_{1k}' - c)$（我们用 c、1 和 0 分别表示各分量均相同的向量 c、1 和 0）。根据定义，$g_k(0) = 0$，因此 $g_k(\ddot{x}_1') = (g_k(1) - g_k(0))\ddot{x}_{1k}'$，这意味着 g_k 与加速度的第 k 个分量成正比。因此，导数 $\frac{d}{dy} C_a$ 再次被证明为常数，于是我们得到公式 11 中的解。

2 固定奖励函数的最优力

这里我们证明，对于由某些力函数 $F_{ik}(x)$（如公式 7 所示）确定的固定奖励函数 $R(x, \dot{x})$，在满足某些正则性条件的情况下，使公式 1 中的剩余成本达到绝对最小值的轨迹由运动方程 $\ddot{x}_{ik} = F_{ik}$ 给出。换句话说，˜任何其他力 $F_{ik} \neq F_{ik}$ 都会导致剩余成本不低于该值。

根据公式 1，剩余成本为

$$\mathcal{C} = \int_{t_0}^{\infty} dt\; e^{-\gamma(t-t_0)} \left[\sum_i \frac{1}{2} m_i \|\ddot{x}_i(t)\|^2 - R(x(t), \dot{x}(t)) \right]\,, \tag{16}$$

其中，x 及其时间导数表示粒子的坐标、速度和加速度，并且我们取初始时间为零，$t_0 = 0$。根据公式 7，与运动方程 $m_i \ddot{x}_{ik} = F_{ik}(x)$ 相关的奖励函数为：

$$R(x, \dot{x}) = -\sum_{ik,jp} \dot{x}_{ik} \dot{x}_{jp} \frac{\partial F_{ik}}{\partial x_{jp}} - \frac{1}{2} \sum_{ik} \frac{F_{ik}^2}{m_i} + \gamma \sum_{ik} \dot{x}_{ik} F_{ik} \tag{17}$$

其中，我们取 $B(x) = 0$。因此，将公式 17 代入公式 16，初始条件 x_0 和 $\dot{x}_0$ 对应的剩余成本为：

$$\mathcal{C}(x_0, \dot{x}_0) = \int_0^\infty dt\ e^{-\gamma t} \left[\sum_{ik} \frac{1}{2} m_i \ddot{x}_{ik}^2 + \sum_{ik,jp} \dot{x}_{ik} \dot{x}_{jp} \frac{\partial F_{ik}}{\partial x_{jp}} + \frac{1}{2} \sum_{ik} \frac{F_{ik}^2}{m_i} - \gamma \sum_{ik} \dot{x}_{ik} F_{ik} \right] . \tag{18}$$

右端第二项的积分可通过分部积分改写为

$$\begin{aligned} \int_0^\infty dt\ e^{-\gamma t} \sum_{ik,jp} \dot{x}_{ik} \dot{x}_{jp} \frac{\partial F_{ik}}{\partial x_{jp}} &= \int_0^\infty dt\ e^{-\gamma t} \sum_{ik} \dot{x}_{ik} \frac{d}{dt} F_{ik}(x) \\ &= e^{-\gamma t} \sum_{ik} \dot{x}_{ik} F_{ik}(x) \Big|_0^\infty - \int_0^\infty dt\ e^{-\gamma t} \sum_{ik} (-\gamma \dot{x}_{ik} + \ddot{x}_{ik}) F_{ik}(x) \\ &= -\sum_{ik} \dot{x}_{ik,0} F_{ik}(x_0) - \int_0^\infty dt\ e^{-\gamma t} \sum_{ik} (-\gamma \dot{x}_{ik} + \ddot{x}_{ik}) F_{ik}(x) , \end{aligned} \tag{19}$$

其中我们假设 $\dot{x}_{ik} F_{ik}(x)$ 有界，因此在长时间极限下，衰减的指数项占主导。将此表达式代回公式 18，

$$\begin{aligned} \mathcal{C}(x_0, \dot{x}_0) &= -\sum_{ik} \dot{x}_{ik,0} F_{ik}(x_0) + \int_0^\infty dt\ e^{-\gamma t} \sum_{ik} \left(\frac{1}{2} m_i \ddot{x}_{ik}^2 - \ddot{x}_{ik} F_{ik} + \frac{1}{2m_i} F_{ik}^2 \right) \\ &= -\sum_{ik} \dot{x}_{ik,0} F_{ik}(x_0) + \int_0^\infty dt\ e^{-\gamma t} \sum_{ik} \frac{1}{2m_i} (m_i \ddot{x}_{ik} - F_{ik})^2 . \end{aligned} \tag{20}$$

右端第一项恰好是公式 6 中的最小代价到达函数。这意味着，我们需要证明：当动力学遵循 $m_i \ddot{x}_{ik} = F_{ik}(x)$ 时，该积分为零；而当 ˜ ˜动力学遵循 $m_i \ddot{x}_{ik} = F_{ik}(x)$，且 $F_{ik}(x) \neq F_{ik}(x)$ 时，剩余积分非负。通过直接代入可知，当 $m_i \ddot{x}_{ik} = F_{ik}(x)$ 时，积分为零，这一点显而易见。此外，由于质量非负，任何不同于 $m_i \ddot{x}_{ik} = F_{ik}(x)$ 的动力学选择都会使积分非负。由此证明了我们最初的论断。

3 参数

在图1的 a–c 面板中，我们对牛顿引力作用下的 5 个粒子进行模拟。它们标记为 $k = \{0, 1, 2, 3, 4\}$，其中粒子 $k = 2$ 质量较大，质量为 $m_1 = 10$；其他粒子的质量较小，分别为 $m_2 = 1 + 0.01k$、$k = \{0, 1, 3, 4\}$。我们使用 $G = 1$、$\gamma = 10$，模拟时间为 $T = 2$（图中仅绘出其中一段），积分时间步长为 $\delta t = 10^{-7}$。x 坐标上的初始条件为 $x_0 = 2(k-2)/5$，而在 y 和 z 坐标上，初始条件均为零。速度仅在 y 方向上分量非零，数值为 $v_{y,\ 0} = 4(k-2)/5$。

在图1d中，我们对牛顿引力作用下的 10 个粒子进行模拟。它们标记为 $k = \{0, \cdots, 9\}$，其中粒子 $k = 5$ 质量较大，质量为 $m_1 = 10$；其他粒子的质量较小，分别为 $m_2 = 0.01(k+1)$、$k = \{0, \cdots, 4, 6, \cdots, 9\}$。我们使用 $G = 1$、$\gamma = 10$，模拟时间为 $T = 1$（图中仅绘出其中一段），积分时间步长为 $\delta t = 10^{-10}$。x 坐标上的初始条件为 $x_0 = 3(k-4.5)/10$，而在 y 和 z 坐标上，初始条件均为零。速度仅在 y 坐标上分量非零，数值为 $v_{y,\ 0} = 3(k-4.5)/5$。

在图1e中，我们模拟了受扰动牛顿引力作用的5个粒子。除牛顿力从 $1/r^2$ to$1/r^{2+\epsilon}$ with $\epsilon = 0.05$ 受到扰动外，所有参数均与图1a–c相同。在扰动力作用下生成轨迹，并以两种不同方式计算轨迹上每个状态的剩余成本。首先，使用公式 10 中的最小代价到达函数，

仿佛从该状态开始，之后的力都恢复为牛顿力。其次，利用公式 1 对生成的轨迹进行线积分，并严格按照公式 9 中的奖励函数（不加入任何扰动），数值计算扰动力产生的剩余成本。结果表明，扰动力产生的剩余成本始终高于最小代价到达函数。

在图 1f 中，我们分别对受扰动牛顿引力作用和受库仑力作用的 3 个粒子进行模拟。对于引力，我们选择 $G = 1$ 和 $\gamma = 5$，模拟时间为 $T = 5$，积分时间步长为 $\delta t = 10^{-7}$。这 3 个粒子标记为 $k = \{0, 1, 2\}$，其中粒子 $k = 1$较重，质量为 $m_1 = 10$，其余粒子较轻，质量分别为 $m_2 = 0.01(k + 1)$、$k = \{0, 2\}$。x 坐标上的初始条件为 $x_0 = k - 1$，而 y 和 z 坐标上的初始条件均为零。速度仅在 y 坐标上非零，值为 $v_{y,\ 0} = 2(k - 1)$。

对于库仑力，较重粒子的质量为 $m_1 = 10$，其余粒子较轻，且质量相同，均为 $m_2 = 0.01$。较重粒子带有正电荷 $q = 1$，较轻粒子带有负电荷 $q = -1$。我们选择 $C_{Coulomb} = 1$ 和 $\gamma = 5$，模拟时间为 $T = 5$，积分时间步长为 $\delta t = 10^{-7}$。x 坐标上的初始条件为 $x_0 = k - 1$，而 y 和 z 坐标上的初始条件均为零。速度仅在 y 坐标上非零，值为$v_{y,\ 0} = 2(2 + 0.01k)(k - 1)$。

在测试不同参数（包括初始条件）时，我们没有观察到所呈现结果出现任何定性变化。