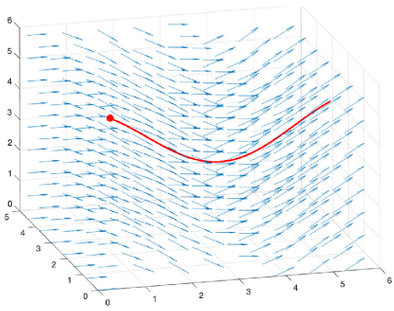
逻辑公式的「空间形状」

YKY [November 29, 2025]

看待 AGI 的其中一种观点、当然也是我最喜欢的观点，就是说 AGI 的目的，是 **学习一套逻辑公式去描述世界**。

这套逻辑公式本来是不存在的，它是从机器学习的过程中「无中生有」的。但既然逻辑法则可以任由我们创造，而目的是 maximize rewards，则似乎这个问题 "under-specified," 也就是说约束条件太弱。其实它的约束条件是因为 **记忆有限**，换句话说是一个 **资讯压缩** 的问题。所以这个问题是 well-defined 而且有 solution, 在数学上是一个很有意思的问题。本文试图准确地 描述 逻辑公式的空间结构。

首先，如果大家熟悉微分方程的，应该见过所谓 vector field 里的 "flow"（向量场的流动）：



(1)

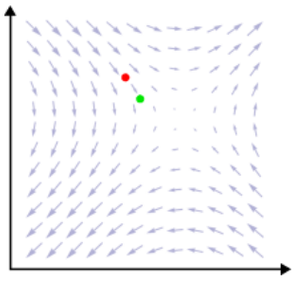
在状态空间里，我们会走出一条 **轨迹** (trajectory), 在轨迹上会收到 **奖励** (rewards). 强化学习的目的就是 maximize 在长远的 time horizon 上的奖励总和。

注意 以上的图像是 **连续的**，但 AGI 的符号逻辑的状态是 **离散的**。离散的 强化学习 服从 Bellman 方程，而 连续的 control theory 服从 Hamilton-Jacobi 方程，它描述一个粒子在某力场之下的运动方式，后者的向量流称为 Hamiltonian flow. 有时我会在这两个图像之间跳了跳去，以获得某些 insights，但这也不是必需的，只是我也稍为熟悉物理那边，所以比较方便。

在 AGI 里，状态 = Working Memory 的内涵，每个状态就是一个「故事」。比如说，x_0 =「现在是凌晨 3am ∧ 我很肚饿 ∧ 冰箱又没有食物 ∧ 钱包也没现金。」或者 x_7 =「我很爱她 ∧ 但她不爱我 ∧ 昨天还被她扇了一巴掌。」

而 vector field 则代表每个状态可以如何 transition 到另一状态。换句话说，vector field **等价于** 我们的逻辑知识库，但逻辑以特殊的方式定义每个状态点上的 tangent vector; 通常一个逻辑公式可以定义很多个状态上的 tangent vectors. 因此这个 向量场具有特殊的逻辑结构，形成数学上有趣的问题。

$\mathbb{X}$ = **state space (状态空间)**



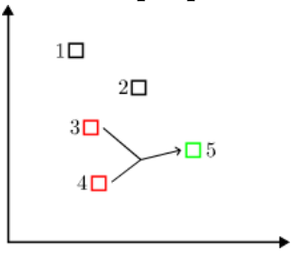
顺着向量场的状态转移

$\bullet \to \bullet$

$x_0 \mapsto x_1$

(2)

$\mathbb{X}$ = **proposition space (命题空间)**



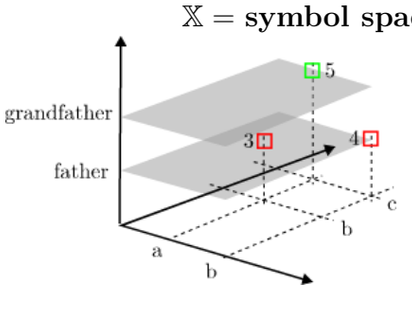
$\bullet = \{\boxed{1}, \boxed{2}, \boxed{3}, \boxed{4}\}$

$\bullet = \{\boxed{1}, \boxed{2}, \boxed{5}\}$

$\boxed{3} \wedge \boxed{4} \to \boxed{5}$

(3)

$\mathbb{X}$ = **symbol space (符号空间)**



$\boxed{3} = \text{father}(a, b)$

$\boxed{4} = \text{father}(b, c)$

$\boxed{5} = \text{grand-father}(a, c)$

(4)