arXiv:2301.00923v1 [cs.LG] 3 Jan 2023

通过非凸优化探索复杂动力系统

Hunter Elliott

2022年12月

摘要

即使动力系统可以用简单的机理模型充分描述，对其复杂行为进行分类也可能颇具挑战。如果这类系统难以进行解析处理，蛮力模拟往往是唯一的选择。我们提出一种替代方案：利用机器学习工具，以优化为驱动的方法。我们将此方法应用于一种新颖、完全可优化的反应扩散模型，该模型纳入了复杂的化学反应网络（称为“稠密反应扩散网络”或“Dense RDN”）。借此，我们能够系统地识别新的状态与行为，包括图案形成、耗散最大化的非平衡态，以及类复制动力学结构。

1 引言

远离平衡态的驱动可使化学反应系统在其组成化学物种浓度的时间和空间变化上呈现出惊人的复杂性（图 1，附录 B.1, [1]）。这种复杂性可以用简单的反应扩散（RD）模型来刻画，并且人们已认识到，它对于理解从生物图案形成 [2][3] 到更广泛意义上产熵“耗散结构”的涌现 [4], 等各种非平衡现象都具有重要意义；甚至有人推测，它对生命起源的最早阶段也至关重要 [5][6][7]。其背后的微观物理化学原理已得到充分理解，而基于这些原理并仅采用最少假设建立的简单微分模型，与实验结果高度吻合 [8][9][10][11]。

尽管已经从机理层面理解了这些系统，但采用还原论方法研究这类系统往往收效甚微。写出描述



图 1：反应扩散系统中可能出现的复杂动力学示例。每个面板展示了 5 种化学物种的局部浓度，从 $t = 0s$（左）到 $t = 1000s$（右）。所有浓度可视化结果均经主成分分析投影为 3 种颜色，单位为任意单位（附录 A.2）。

系统的偏微分方程，并不能轻易通过解析方法全面揭示其可能呈现的行为，因此需要转而关注降阶模型或特定行为 [12][13][14][15]。一些解析方法或许能够确定对应于稳定或复杂动力学的参数范围，但一般情况下无法列举出哪些复杂行为和状态是可能实现的 [16]。此外，这些系统中有许多会表现出混沌行为，这意味着它们可能会随着时间推移遍历无穷多个状态 [17][18][19]。虽然可以高精度地进行正向数值模拟，但对于哪怕是复杂程度中等的系统，穷举探索其状态空间和参数空间也至少十分繁琐，最糟的情况下还会耗费难以承受的计算资源 [3]。

本文探讨的另一种“逆向设计”方法，是先选定一种感兴趣的特定行为，再询问能否在模型的构型空间中找到与该行为对应的点。如果感兴趣的行为可以表示为系统状态和/或动力学的易于评估的数学函数，那么就可以将对系统行为的探索表述为一个优化问题。尽管这一优化问题在大多数情况下是非凸的，现代机器学习（ML）的一些工具仍可能适用。因此，我们不再试图全面列举化学反应系统可能呈现的各种行为，而是针对某个感兴趣的特定状态或动力学检验一个假设，从而避开分析上的棘手难题和穷举搜索所带来的高昂复杂度。

以往曾将反应-扩散系统的逆向设计表述为一个优化问题，尽管采用的并非符合物理实际的化学模型，且应用往往面向图像处理或纹理合成 [20][21][22][23]。有些研究设计了符合实际的反应-扩散系统，但采用的是启发式方法 [24][25]。其他采用符合物理实际的反应-扩散模型的逆向设计方法则避开优化，而是旨在构建反应-扩散系统模块，以分解复杂性，并支持模块化或层次结构的程序化设计 [26] ，或实现类似元胞自动机的布尔动力学 [27]。有些方法可以对化学反应网络动力学进行逆向设计，但不包含扩散组分或空间组织 [28]。机器学习方法也曾应用于相关的实际模型，但用途不同：要么用于近似前向模型 [29][30][31], ，要么用于从真实世界数据中学习模型 [32][33][34]。本文既不近似物理模型，也不使用任何真实世界数据。相反，我们仅使用机器学习中的优化方法，探索一个未经近似、已知但具有灵活性的符合物理实际的前向模型，不仅考察稳态，也考察动力学和热力学。

我们提出的方法存在局限，不仅体现在可探索的系统和行为上，也体现在基于这些探索能够得出的结论上：优化未能收敛，并不意味着某种状态或动力学行为不可能存在，只是缺乏其可能存在的证明。尽管如此，我们仍希望在此表明，该方法可以补充现有方法，并有望用于研究驱动的非平衡化学反应系统，以及更广泛的复杂动力系统。

2 一种优化方法

从总体上看，这种方法要求我们指定一个模型，用来模拟感兴趣系统的动力学；同时还要指定一个损失函数，用以表达关于某种

系统可能具备的动力学行为的假设。损失函数只是模型的状态、动力学和参数的标量函数；当且仅当系统呈现出所关注的行为或状态时，该函数才取较小值。如果模型和损失函数都可微，我们就可以借助机器学习中的模型构建与优化方法来最小化损失函数，并尝试实现所关注的行为。本文重点介绍一种新颖的“稠密反应扩散网络”模型，即 Dense RDN。在该模型中，任意数量的化学物种通过反应网络相互转化，同时在两个空间维度上扩散。化学反应、反应的速率常数、扩散系数和初始条件均由优化确定。损失函数由研究者选择，可以表示稳定性或双稳态等简单行为，也可以表示更复杂的行为；第 3 节将展示几个示例。

2.1 预备知识

更正式地说，我们假设可以指定一个前向模型 $\Psi(\boldsymbol{X},\ \boldsymbol{\theta}_\Psi)$，使状态 $\boldsymbol{X}$ 随时间演化，并以 $\boldsymbol{\theta}_\Psi$ 为参数。该模型应给出微分量 $\frac{\partial \boldsymbol{X}}{\partial t} = \Psi(\boldsymbol{X},\ \boldsymbol{\theta}_\Psi)$，以便用于例如，前向欧拉积分：

$$\begin{aligned} \boldsymbol{X}_{t+1} &= \boldsymbol{X}_t + \Psi(\boldsymbol{X}_t, \boldsymbol{\theta}_\Psi)\Delta t \\ &= \boldsymbol{X}_t + \Delta \boldsymbol{X}_t \end{aligned} \tag{1}$$

给定时间步长 Δt，状态变化量为 $\Delta \boldsymbol{X}_t$（关于如何确保此步长足够小，详见附录 A.1）。初始条件 $\boldsymbol{X}_0$ 由随机向量 $\boldsymbol{z}$ 通过模型 $\boldsymbol{X}_0 = G(\boldsymbol{z},\ \boldsymbol{\theta}_G)$生成（在本研究中，该模型是一个神经网络），其参数为 $\boldsymbol{\theta}_G$。对这些初始条件重复应用式 (1)，即可得到时间演化 $\mathcal{X} = [\boldsymbol{X}_0, \boldsymbol{X}_1, ..., \boldsymbol{X}_T]$。为方便起见，我们定义 $\boldsymbol{\theta} = \boldsymbol{\theta}_\Psi \cup \boldsymbol{\theta}_G$，并要求 Ψ 和 G 关于 $\boldsymbol{X}$ 和 $\boldsymbol{\theta}$几乎处处可微。

最后，我们假设可以指定一个标量损失函数 $\mathcal{L}(\boldsymbol{\theta})$，使得当损失函数趋近于最小值时，$\mathcal{X}$ 中会呈现出所关注的行为。作为一个简单示例，如果我们希望识别稳态，可以定义 $\mathcal{L}(\boldsymbol{\theta}) = \|\Delta \boldsymbol{X}_t\|_2$。因此，总体而言，我们寻求参数 $\boldsymbol{\theta}^*$：

$$\boldsymbol{\theta}^* = \arg\min_{\boldsymbol{\theta}} \mathcal{L}(\boldsymbol{\theta}) \tag{2}$$

由于 $\mathcal{L}$ 完全可微，因此可以应用随机梯度下降（SGD）等基于梯度的优化技术。如果优化收敛到足够低的损失函数值，使所得动力学满足研究者的标准，那么我们就确定了能够产生相应时间演化 $\mathcal{X}$ 的初始条件和转移模型参数，从而证明该行为属于模型可能呈现的动力学。相反，如果优化未能得到这样的参数，我们不能据此断定所需行为是不可能的，只能说明这一过程未能对其进行参数化。

成功的希望? 乍看之下，这似乎是注定要失败的尝试。我们希望探索具有难以处理的复杂非线性动力学的模型 $\Psi(\boldsymbol{X},\ \boldsymbol{\theta})$。这使 $\boldsymbol{\theta}$ 与 $\mathcal{X}$ 之间的关系高度非线性，并且无法保证

损失函数 $\mathcal{L}$具有凸性。非凸优化很困难（NP难），但我们建议使用随机梯度下降等局部优化方法来寻找 $\boldsymbol{\theta}^*$。尽管如此，有两点观察仍让我们备受鼓舞。首先，我们不需要找到损失函数的全局最小值，只需找到一个“足够低”的局部最小值，使所得结构和/或动力学符合研究者对感兴趣行为的标准。其次，整个深度学习（DL，机器学习的一个子领域）领域取得了显著成功，而它依赖于这类优化即使看似难以实现也能成功。对于经验上观察到的这类非凸优化能够可靠收敛的原因，目前仍是一个活跃的研究领域，本文不予探讨。不过，我们推测有一种解释与此相关：当待优化的参数空间具有足够高的维数时，具有高损失值的真正局部最小值越来越不可能存在 [35]。我们没有系统地研究收敛所需的条件。不过，我们确实表明，通过将高维物理化学模型嵌入维数更高的神经网络，并采用深度学习中使用的优化方法，我们能够在各种损失函数下实现收敛。这至少从经验上表明，在数据驱动机器学习领域广受欢迎的优化方法，也可用于无数据、完全由正向模拟驱动的复杂动力学模型探索。

2.2 密集反应-扩散网络模型

我们选择反应扩散（RD）模型作为一种简单而准确的模型，它能够呈现时空复杂行为，并且在概念上至少与物理学、化学和生物学相关。实验中已经观察到，即使是简单的反应-扩散系统，也会产生复杂、甚至混沌行为，而这些行为能够由此类模型很好地描述 [8][9][10][11]。在反应物持续流入的驱动下，这些模型代表了热力学中所关注的非平衡系统，甚至可能与生命起源有关 [5][36]。以往的大量研究分析了人工设计的化学反应集合，而在本文中，我们从一个规模庞大且密集的化学反应网络入手，并允许在优化过程中确定适当的反应集合。

2.2.1 化学：化学反应网络

在本研究中，前向模型 Ψ 是一个反应扩散模型。该模型中的化学反应构成了我们所称的“密集”化学反应网络（CRN），因为它包含所有符合以下三个反应原型的可能反应：

$$\begin{aligned} A + B &\rightleftarrows 2C \\ A &\rightleftarrows B \\ A + 2B &\rightleftarrows 3B \end{aligned} \tag{3}$$

所谓“反应原型”，是指虽然化学反应网络可以包含任意数量的化学物种 N_s，但式 (3) 中的字母仅表示 A、B、C 必须是 3 种不同的物种，并具有指定的化学计量数（见附录 B）。每个反应的正向和逆向速率都是自由参数 $\boldsymbol{\kappa} = \{k_{f1}, k_{r1}, k_{f2}, k_{r2}, ...\} \subset \boldsymbol{\theta}_\Psi$，它们结合标准质量

作用动力学后，会产生由化学反应 $\Delta \boldsymbol{X}_{t_{Rxn}}$ 引起的状态净变化。式 (3) 中的反应类型会带来高度非线性的动力学，因此可以确定复杂行为至少是可能的（见附录 B.3），尽管参数空间中的许多区域会产生乏味的行为（例如单调弛豫至平衡态）。这种反应网络结构还涵盖了 3 种或更多化学物种之间可能发生的许多单分子、双分子和三分子反应。

虽然所有这些反应同时存在于现实世界的化学反应网络中或许并不太可能，但请注意，优化可以将反应速率设为 ~ 0，因此我们实际上优化的不仅是反应速率，还包括要纳入哪些反应。

2.2.2 非平衡：流动反应器驱动

我们将系统建模为处于"流动反应器"中，因此通过反应物的恒定流入使其保持在远离平衡态的状态。这种流入会产生相应的流出，从而使化学物种 $\boldsymbol{X}^i$ 的浓度因该驱动而产生净变化 [37][1]:

$$\Delta \boldsymbol{X}^i_{t_{Drv}} = \left(f x_i - f \boldsymbol{X}^i_t\right) \Delta t \tag{4}$$

同样，每种物种的进料浓度 x_i 和共享的流速 f 都是在优化过程中确定的。

2.2.3 空间：扩散与初始条件

为实现空间组织，化学反应网络存在于一个离散化的二维区域内，因此化学物种 i 在位置（$u,\ v$）处的浓度由 $\boldsymbol{X}^i\left(u,\ v\right)$ 给出。初始条件 $\boldsymbol{X_0}$ 由随机向量 $\boldsymbol{z}$ 通过 G 生成；G 被实例化为一个神经网络，类似于 DCGAN [38] 等卷积生成器模型（详情请参见附录 A.4）。

每种化学物种也会发生扩散:

$$\Delta \boldsymbol{X}^i_{t_{Dif}} = D_i \nabla^2 \boldsymbol{X}^i_t \Delta t \tag{5}$$

扩散系数 $D_i \subset \boldsymbol{\theta}_\Psi$ 可进行优化。状态的最终综合变化是反应、驱动和扩散所产生贡献的总和:

$$\Delta \boldsymbol{X}_t = \Delta \boldsymbol{X}_{t_{Rxn}} + \Delta \boldsymbol{X}_{t_{Drv}} + \Delta \boldsymbol{X}_{t_{Dif}} \tag{6}$$

这些项及其参数共同定义了一个非平衡物理化学模型空间，该空间既能呈现常见行为，也能呈现时空复杂行为，并且足够灵活，能够让优化过程确定模型实际采用的状态和动力学。

3 结果

要完整地指定一个优化过程，剩下的就是定义损失函数 $\mathcal{L}$，以表征所关注的行为。实际上，这个损失函数编码了关于模型可能状态和/或动力学的假设，我们将在此探究几种这样的假设。

3.1 图案形成

图案形成——化学物种浓度中涌现出的有组织空间变化——是反应扩散系统的一种标志性行为 [2][39]。自图灵在 1952 [40] 提出这一假设以来，该现象已在实验中被反复重现，人们也认识到它对生物图案形成具有重要意义 [2][3]，甚至有假设认为它与生命起源最早期的阶段有关 [5][6][7]。

我们旨在确定对应于系统能够支持任意结构稳定图案的动力学参数（$\boldsymbol{\theta}_\Psi$）范围。由于这些将是由非平衡过程驱动的图案，我们选择一种合适的任意结构作为我们的“目标”：Ilya Prigogine 的图像（图 2a，左）以此致敬他在耗散结构方面的开创性工作 [4][41]。因此，我们定义了一个损失项，以促使化学物种 i 的浓度分布与该目标 $\boldsymbol{X}$ 相匹配：

$$\mathcal{L}(\boldsymbol{\theta}) = \mathbb{E}\left[\|\Psi(\boldsymbol{X}_{t-1}, \boldsymbol{\theta}_\Psi)^i - \tilde{\boldsymbol{X}}\|_2\right] = \mathbb{E}\left[\|\boldsymbol{X}_t^i - \tilde{\boldsymbol{X}}\|_2\right] \tag{7}$$

其中，期望值同时取自时间和 $\boldsymbol{X}_0 \sim G(\boldsymbol{z}, \boldsymbol{\theta}_G)$，在实践中通过对时间和 $\boldsymbol{z}$ 进行随机采样来实现。此外，我们旨在识别稳定图案，因此引入了一个额外的损失项，以促进时间稳定性：

$$\mathcal{L}(\boldsymbol{\theta}) = \mathbb{E}\left[\|\boldsymbol{X}_t^i - \tilde{\boldsymbol{X}}\|_2 + \lambda_\Delta \|\Delta \boldsymbol{X}_t\|_1\right] \tag{8}$$

这确保我们优化的不只是与目标匹配的初始条件，还包括一个能随时间保持该目标的动力学模型。要成功收敛到模仿该目标的图案，需要运用许多在深度学习中已普遍采用的优化启发式方法，以及一些针对特定任务和模型所做的调整，详见附录D。尽管如此，这仍然是可行的，如图 2a 所示，其中一个在

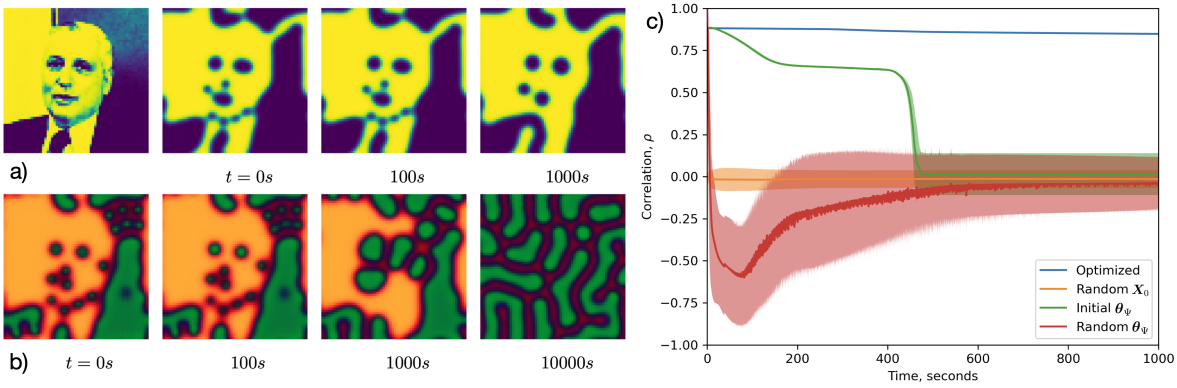

图 2：经过优化以匹配目标（a，最左侧面板）的4种化学物质稠密反应扩散网络支持半稳定的耗散结构（a，右侧面板）。第二个5种化学物质示例（b）的稳定性较低，显示出动态的图案形成。c）对 a) 中示例随时间的相关性进行测量后证实，对初始条件和反应网络同时进行优化对于保真度和稳定性是必要的。实线表示均值，带状区域表示在 $n = 32$ 次随机化中 +/- 一个标准差。

4物种化学反应网络中发现的图案与目标具有较高的保真度（在已优化的时间尺度 $32s$ 内，皮尔逊相关系数平均值为 .88 r ），并且在远长于优化过程中所采用的时间尺度上仍保持稳定（图 2c）。在其他实验中，目标图案的时间稳定性较低，优化显然收敛到了对应于动态图案形成的动力学区域（图 2b）。我们通过随机化初始条件 $\boldsymbol{X}_0$ 和反应网络参数 $\boldsymbol{\theta}_\Psi$，并比较由此得到的相关性随时间与目标的相关性，确认这两者都必须进行优化（图 2c，另见附录D.4）。

因此，这证实了我们可以通过纯粹的优化，从头识别出一组初始条件以及反应网络的动力学和拓扑结构（见附录B.4），使其能够形成任意图案。

3.2 耗散最大化

由于这些是受驱动的非平衡系统，自然会引出这样一个问题：随着熵产生率——即耗散率——增加，这些模型的行为会发生怎样的变化？在不趋向平衡态的系统中，耗散率就是输入驱动能量转化为熵的速率。这个 Dense RDN 系统的热力学定义明确（[37][42], 见附录 C）。熵产生率是时间演化 $\mathcal{X}$ 的可微函数，因此可用于构建损失函数，并按扩散和反应耗散率进行分解 $\sigma_{Tot} = \sigma_{Rxn} + \sigma_{Dif}$：

$$\mathcal{L}(\boldsymbol{\theta}) = \mathbb{E}\left[e^{-\sigma_{Rxn}(\boldsymbol{X}_t,\boldsymbol{\theta})-\sigma_{Dif}(\boldsymbol{X}_t,\boldsymbol{\theta})}\right] \tag{9}$$

请注意，我们对负耗散率取指数，使其成为一个下界为零的最小化问题；原始耗散率没有界限，且可能相差数个数量级，

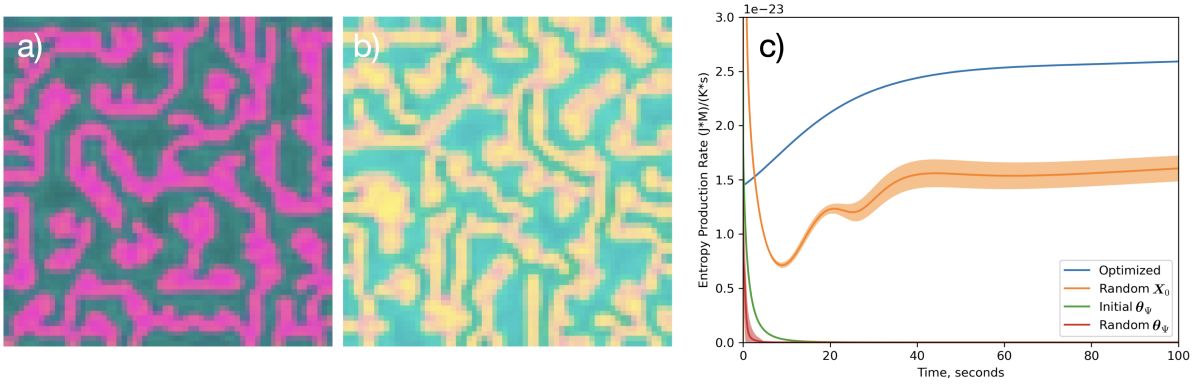

图3：最大化扩散产生熵的速率，会使由5种化学物质构成的 Dense RDN 中的局部浓度形成有趣的丝状结构（a-b）。我们通过比较 a) 中已优化示例（蓝色）的耗散率随时间的变化，以及初始条件经过随机化的模型（橙色）、优化初始时采用的参数（绿色）和动力学经过随机化的模型（红色），确认了各个已优化组件的贡献（c）。实线表示均值，色带表示在 $n = 32$ 次随机化中的 +/- 一个标准差。

这可能导致优化过程中出现数值不稳定。最后，我们寻求的是稳定地处于非平衡状态的状态，而非仅仅经历耗散性弛豫并趋近平衡态的状态，因此我们引入一项，用以鼓励耗散率保持不变的解：

$$\mathcal{L}(\boldsymbol{\theta}) = \mathbb{E}\left[e^{-\sigma_{Rxn}(\boldsymbol{X}_t,\boldsymbol{\theta}))-\sigma_{Dif}(\boldsymbol{X}_t,\boldsymbol{\theta}))}\right] + \lambda_\sigma \mathrm{Var}\left[e^{-\sigma_{Tot}}\right] \tag{10}$$

其中，方差仅针对时间计算，并且对驱动流速 f 进行约束，使最大耗散率保持有限。式（10）的最小化由反应耗散率主导；该耗散率在空间位置置换下保持不变，因此并不诱导任何空间结构，这一点并不令人意外。不过，我们仍可将所得反应网络视为高熵生成率化学反应网络的示例（有关该化学反应网络的可视化，请参见附录B.4）。请注意，在此示例中，驱动流速 f 被固定为 $.03s^{-1}$）。

如果我们改为仅纳入扩散耗散率项——该项要求浓度具有空间非均匀性才能不为零——便会发现有趣的结构：与未进行优化时相比，这些结构能够以稳定方式耗散远高于其水平的驱动能量，并且持续的时间尺度也远长于优化过程中使用的64秒（图3）。值得注意的是，这些结构的形式并非以任何方式预先指定，而是完全由耗散最大化损失函数与模型的物理化学性质之间的相互作用自然涌现出来的。

3.2.1 耗散性分布

在实践中，前述所有优化在生成模型 G 中都会遭遇“模式崩溃”，使其最终只生成单个初始条件（这意味着式 (8) 和 (10) 中的期望实际上仅针对时间）。我们或许希望生成的不仅是 $\boldsymbol{X}_0$ 的分布，也包括其时间演化的分布。为此，可以引入一个解码器模型 $\hat{}\boldsymbol{z}_t = E(\boldsymbol{X}_t)$，并通过相应的损失项对其进行优化，使其重构用于生成 $\boldsymbol{X}_0$的随机二进制向量 $\boldsymbol{z}$：

$$\mathcal{L}_z(\boldsymbol{\theta}, \hat{\boldsymbol{z}}) = \mathbb{E}\left[H(\boldsymbol{z}, \hat{\boldsymbol{z}}_t)\right] \tag{11}$$

其中，H 为交叉熵，E 由卷积神经网络实现。对于确定性前向模型，式 (11) 显然总能满足，因此我们还引入一个在时空上变化、服从均匀分布的噪声项 $\boldsymbol{\epsilon}_f(u,\ v)$，它对应于流速的随机变化：

$$\Delta \boldsymbol{X}^i_{t_{Drv}} = (f + \boldsymbol{\epsilon}_f) \odot \left(x_i - \boldsymbol{X}^i_t\right) \Delta t \tag{12}$$

其中，$\boldsymbol{\epsilon}_f(u,\ v) \sim$服从 Unif(0，0。8)（经高斯低通滤波以避免空间不连续），$\odot$ 表示 Hadamard 积。这样可以确保 $\boldsymbol{X}_t$的样本之间存在近似的 ϵ_f-球分离。

最小化 (10) 和 (11) 的和，得到 $\boldsymbol{\theta}^*$，其对应的状态分布能够在驱动随机波动的情况下，仍随时间保持其独特性（图 4）。在所示示例中，解码器能够以 100% 的准确率重构 16 位 $\boldsymbol{z}$ 向量，持续超过 1000 个时间点；这表明，尽管环境嘈杂，该动力系统仍能随时间传递信息，其信道容量与 $\boldsymbol{z}$ 的香农熵成正比。

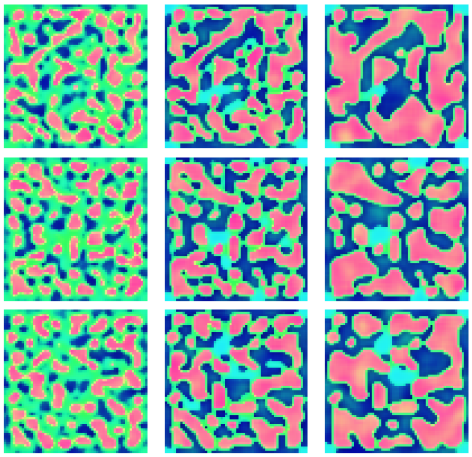

图 4：通过同时最大化扩散熵产生和一种损失函数，得到的 5种化学物质 Dense RDN 中耗散结构的分布。该损失函数要求用于生成初始条件的 $\boldsymbol{z}$ 向量在每个时间点都可重构。每一行是来自$\boldsymbol{z}$的一个样本，列从左到右分别对应 $t = 10$、100， $1000s$。

3.3 动态结构

到目前为止，我们已经求解了动力系统的特定状态 $\boldsymbol{X}_t$，或其状态转变的性质 $\Delta \boldsymbol{X}_t$。另一种有趣的做法是直接针对完整时间演化 $\mathcal{X}$ 的性质进行优化。此前已有研究表明，反应扩散模型中的耗散结构会发生类粒子运动 [43][44]，甚至会“复制”简单的斑点图案 [1][45][46]。在这里，我们不明确指定动力学模型或任何状态 $\boldsymbol{X}_t$，而是寻求结构重组与运动。我们通过优化一个损失函数来实现这一点，该损失函数要求 $\boldsymbol{X}_0$ 与 $\boldsymbol{X}_T$ 的两个平移版本相似（图 5a）：

$$\mathcal{L}(\boldsymbol{\theta}) = \|\boldsymbol{X}_0 - T_{-w}(\boldsymbol{X}_T)\|_2 + \|\boldsymbol{X}_0 - T_{+w}(\boldsymbol{X}_T)\|_2 \tag{13}$$

其中，$T_{-w}(\boldsymbol{X})$表示将 $\boldsymbol{X}$ 垂直平移距离 $-w$。然而，式（13）存在一个平凡解，即所有化学物种的浓度均匀且不随时间变化。因此，我们引入一个损失项，以促使时间序列末尾的浓度空间标准差接近或高于目标值 β^*:

$$\mathcal{L}_{STD}(\boldsymbol{\theta}, \beta^*) = \frac{1}{N_s} \sum_i \text{Max}\left(0, \left(\beta^* - \sqrt{\text{Var}\left[\boldsymbol{X}_T^i\right]}\right)\right)^2 \tag{14}$$

其中，方差是在空间位置上计算的（u， v）。随后，我们最小化（13）和（14）之和。通过要求时间序列开头的一个中心区域与时间序列末尾的两个不同区域 $\boldsymbol{X}_T$ 之间具有这种相似性（图 5a），优化过程必须收敛到不仅仅是平移、而是某种模糊类似于复制的状态，以最小化这一损失函数。这需要在更长的时间尺度上进行优化，因此必须采用经过修改的“增量式”优化过程（见附录 D.3）。尽管如此，优化最终还是收敛到一个化学反应网络（附录图 6b）和一组初始条件，它们大致符合预期的动力学（图5b）。复制保真度一般，

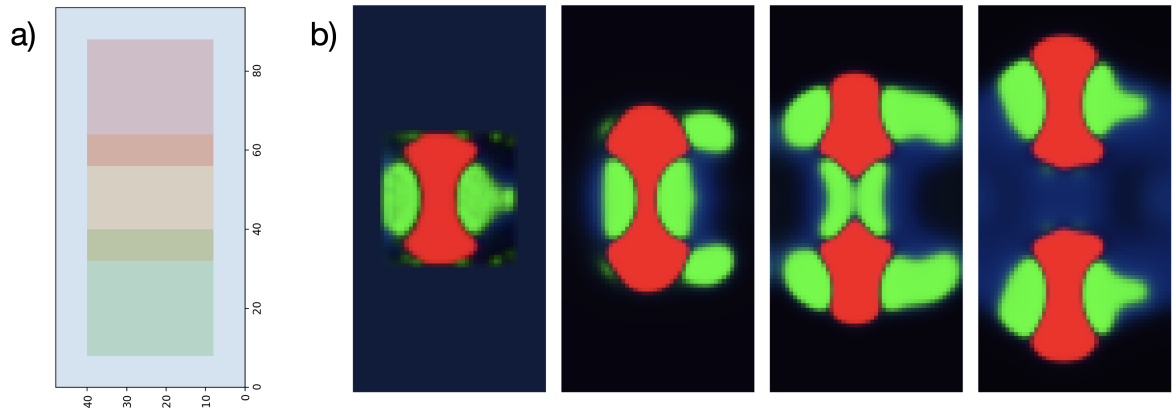

图 5：5种化学物质 Dense RDN 中优化导出的结构及其相关类复制动力学示例。(a) 展示了损失函数的结构：它要求 $t = 0$ 处橙色区域与 $t = T$处红色和绿色的两个区域相似。(b) 展示了从 $t = 0s$ （左）到 $t = 352s$ （右）的局部浓度随时间的变化。请将 b) 中最左侧面板的结构与最右侧面板中的两个结构进行比较。

在 $t = 0$ 处的“母体”结构与 $t = 352s$ 处的两个“子代”结构之间，平均皮尔逊相关系数为 0.893。

当然，这并不 是真正的复制。图 5b 右侧的两个“子代”模式本身并不能分裂；即使它们能够分裂，这种损失结构也无法应对其后代之间不可避免的拥挤和碰撞。尽管如此，这仍是一个示例，展示了相对复杂的动力学重组如何完全由损失函数的形式和物理化学模型的性质所产生。

4 讨论

我们通过若干示例证明，如果能将关于复杂系统可能状态和动力学的假设恰当地用数学形式表示为损失函数，那么就可以通过非凸优化来引导对该系统可能行为的组合爆炸的空间进行搜索。

这种方法存在一个根本性的局限：优化未能收敛，并不构成对该假设的证伪。这只是缺乏支持该假设的证据，而不是证明假设所描述的行为不存在。

尽管如此，这种方法仍通过优化驱动的方式，识别出了能够稳定支持预定空间图样的扩散耦合化学反应网络（3.1）。虽然此处展示的示例并非涌现而成，而是依赖于优化后的初始条件，但我们的方法仍可能有多种形式的应用，例如改进此前展示过的微纳加工技术 [47][48][24][49][26]。重要的是，我们的方法无需预先指定要纳入哪些具体反应，而是允许优化从初始的稠密反应网络中选择这些反应。

我们也可以不针对特定状态进行优化，而是搜索动力学；我们在此展示了，无需预先指定形状或运动（3.3），也能做到这一点。

虽然我们选择的损失函数形式产生了类似“复制”的动力学，但实际上，这种相似性相当表面：这种复制的保真度低且不稳定，而且可能需要更长的时间尺度、更复杂的化学反应网络、专门鼓励稳定性的损失函数，或将这些因素结合起来，才能让它与真正复制之间的类比不那么牵强。尽管如此，仅通过优化就能得到一种非各向同性的浓度分布，并使其以这种方式发生剧烈重组，这至少是朝着证明在没有区室化、而只有反应和扩散的系统中，复杂的真正复制体具有可行性迈出的一步。

或许最有趣的是，这种方法使我们能够考察在寻找具有极端热力学性质（例如熵产生率最大（3.2））的模型时涌现出的耗散结构。这里我们只展示定性观察，并指出，对于该模型中涌现的空间结构，只应最大化耗散率中的扩散组分。这些结构是在简单且时空不变的驱动下发现的。然而，有人假设，促成持续非平衡现象（如生命）涌现的相关适应，是在耗散来自更复杂、难以利用的驱动的能量时诱发的 [50][51]。如果结合适当的时空可变且化学性质复杂的驱动，这种优化方法或许有助于阐明这些假设。

最后，虽然本文聚焦于一种特定的反应-扩散类型模型，但所提出的方法适用于任何可微模型和损失函数。物理学、化学与生物学的交汇领域中，充斥着这样的系统：还原论已为复杂系统建立了简单且准确的微分模型，但迄今为止，这些模型仍未能使人们全面理解其所描述的现象 [52]。这一知识鸿沟也体现在图灵本人的一项观察中；图灵是最早研究反应-扩散系统数学模型的人之一：

> 这是一种假设：一旦某个事实呈现于 *Mind*，关于该事实的所有推论便会同时涌入 *Mind*。这种假设在许多情况下都很有用，但人们很容易忘记它并不成立 [53]。

类比而言，当我们已知某个复杂系统的机理模型这一“事实”，但现有方法仍难以描绘由此产生的整个动力学图景时，我们希望本文所述方法能够提供一种互补途径，弥合这一理解上的鸿沟。

所有源代码均可在 `https://github.com/hunterelliott/dense-rdn` 获取

参考文献

[1] John E Pearson。《简单系统中的复杂图案》。载于：《科学》261.5118（1993），第 189–192 页。[2] Shigeru Kondo 和 Takashi Miura。《作为理解生物图案形成框架的反应扩散模型》。载于：《科学》329.5999（2010），第 1616–1620 页。

[3] Amit N Landge 等。《自组织反应-扩散系统的图案形成机制》。载于：《发育生物学》460.1（2020），第 2–11 页。[4] Ilya Prigogine 和 Ren´e Lefever。《耗散系统中的对称性破缺不稳定性。II》。载于：《化学物理学杂志》48.4（1968），第 1695–1700 页。[5] Ilya Prigogine 和 Gregoire Nicolis。《生物有序性、结构与不稳定性1》。载于：《生物物理学季评》4.2-3（1971），第 107–148 页。[6] P´eter Szab´o 等。《计算机模拟揭示，有限扩散的复制体会进化出更高的效率和保真度》。载于：《自然》420.6913（2002），第 340–343 页。[7] Paul Adamski 等。《从自我复制到复制子系统：迈向从头起源的生命》。载于：《自然综述：化学》4.8（2020），第 386–403 页。[8] Kyoung J Lee 等。《相互作用的化学前沿形成的图案》。载于：《科学》261.5118（1993），第 192–194 页。[9] Gavin R Armstrong 等。《对波跨越一系列间隙的传播进行建模》。载于：《物理化学化学物理》6.19（2004），第 4677–4681 页。[10] Gerhard Ertl。《固体表面反应中的振荡动力学与时空自组织》。载于：《科学》254.5039（1991），第 1750–1755 页。[11] Chad T Hamik 和 Oliver Steinbock。《具有非单调色散关系的反应-扩散介质中的激发波》。载于：《新物理学杂志》5.1（2003），第 58 页。[12] M Or-Guil 等。《三组分反应-扩散系统中的斑点分岔：传播的开始》。载于：《物理评论E》57.6（1998），第 6432 页。[13] AH Khater 等。《双曲正切法：一种简单变换及非线性反应-扩散方程的精确解析解》。载于：《混沌、孤子与分形》14.3（2002），第 513–522 页。[14] Stephen Smith 和 Neil Dalchau。《超越激活子-抑制子网络：广义图灵机制》。载于：arXiv 预印本 arXiv:1803.07886（2018）。[15] Shigeru Kondo。《用于研究生物图案形成机制的更新版基于核函数的图灵模型》。载于：《理论生物学杂志》414（2017），第 120–127 页。[16] Martin Feinberg。《化学反应网络理论基础》。第 202 卷。施普林格，2019。[17] Vitaly Volpert 和 Sergei Petrovskii。《生物学中的反应-扩散波》。载于：《生命物理学评论》6.4（2009），第 267–310 页。[18] Reuben H Simoyi、Alan Wolf 和 Harry L Swinney。《多组分化学反应中的一维动力学》。载于：《物理评论快报》49.4（1982），第 245 页。[19] Otto E R¨ossler。《化学湍流：简单反应-扩散系统中的混沌》。载于：Zeitschrift f¨ur Naturforschung A 31.10（1976），第 1168–1172 页。[20] Ly Phan 和 Cindy Grimm。《绘制反应-扩散纹理》。载于：第三届欧洲计算机图形学协会基于草图的界面与建模会议论文集。2006，第 107–114 页。

[21] Ezio Bartocci 等人。“反应扩散网络中模式识别与合成的形式化方法”。载于：*IEEE* 网络系统控制汇刊5.1 (2016)，第 308–320 页。[22] Alexander Mordvintsev、Ettore Randazzo 和 Eyvind Niklasson。“反应扩散图案的可微编程”。载于：*arXiv* 预印本 *arXiv:2107.06862* (2021)。[23] Alexander Mordvintsev、Eyvind Niklasson 和 Ettore Randazzo。“利用神经元胞自动机生成纹理”。载于：*arXiv* 预印本 *arXiv:2105.07299* (2021)。[24] Grigory Tikhomirov、Philip Petersen 和 Lulu Qian。“具有任意图案的微米级 DNA折纸阵列的分形组装”。载于：《自然》552.7683 (2017)，第 67–71 页。[25] Natalie S Scholes 和 Mark Isalan。“编程图案形成的三步框架”。载于：《化学生物学当前观点》40 (2017)，第 1–7 页。[26] Dominic Scalise 和 Rebecca Schulman。“设计用于复杂图案形成的模块化反应扩散程序”。载于：《技术》2.01 (2014)，第 55–66 页。[27] Dominic Scalise 和 Rebecca Schulman。“在化学反应扩散网络中模拟元胞自动机”。载于：《自然计算》15.2 (2016)，第 197–214 页。[28] Niall Murphy 等人。“合成和调节具有指定行为的随机化学反应网络”。载于：《英国皇家学会界面期刊》15.145 (2018)，第 20180283 页。[29] Leon O Chua 等人。“自主细胞神经网络：图案形成与主动波传播的统一范式”。载于：*IEEE* 电路与系统汇刊第一辑：基础理论与应用42.10 (1995)，第 559–577 页。[30] Angran Li 等人。“基于卷积神经网络的反应扩散系统预测”。载于：《科学报告》10.1 (2020)，第 1–9 页。[31] Jaideep Pathak 等人。“从数据中无模型预测大型时空混沌系统：一种储备池计算方法”。载于：《物理评论快报》120.2 (2018)，第 024102 页。[32] Jaideep Pathak 等人。“利用机器学习复现混沌吸引子并根据数据计算李雅普诺夫指数”。载于：《混沌：非线性科学跨学科期刊》27.12 (2017)，第 121102 页。[33] Behzad Zakeri 等人。“求解偏微分方程的弱监督学习技术：一维反应扩散方程案例研究”。载于：高性能计算与大数据分析国际会议。施普林格。2019，第 367–377 页。[34] Kyongmin Yeo 和 Igor Melnyk。“用于含噪动力系统数据驱动仿真的深度学习算法”。载于：《计算物理学杂志》376 (2019)，第 1212–1231 页。[35] Yann N Dauphin 等人。“识别并攻克高维非凸优化中的鞍点问题”。载于：《神经信息处理系统进展》27 (2014)。

[36] Istv´an Scheuring 等人。“前生物进化的空间模型：汤先于披萨？”载于：生命起源与生物圈演化33.4 (2003)，第 319–355 页。[37] Dilip Kondepudi 和 Ilya Prigogine。《现代热力学：从热机到耗散结构》。John Wiley & Sons，2014。[38] Alec Radford、Luke Metz 和 Soumith Chintala。“利用深度卷积生成对抗网络进行无监督表征学习”。载于：*arXiv* 预印本 *arXiv:15*

11.06434

(2015).[39] Vladimir K Vanag 和 Irving R Epstein。“反应扩散系统中的图案形成机制”。载于：《国际发育生物学杂志》53.5-6 (2009)，第 673–681 页。[40] AM Turing。“《形态发生的化学基础》”。载于：《伦敦皇家学会哲学汇刊》。*B*辑：生物科学237.641 (1952)，第 37–72 页。[41] Ilya Prigogine。“时间、结构与涨落”。载于：《科学》201.4358(1978)，第 777–785 页。[42] Hitoshi Mahara、Tomohiko Yamaguchi 和 Masatsugu Shimomura。“二维可逆 Gray-Scott 系统中的熵产生”。载于：《混沌：非线性科学跨学科期刊》15.4 (2005)，第 047508 页。[43] Henry C Tuckwell。“反应扩散系统中的孤子”。载于：《科学》205.4405 (1979)，第 493–495 页。[44] Mathias Bode 和 H-G Purwins。“反应扩散系统中的图案形成——物理系统中的耗散孤子”。载于：《物理学*D*辑：非线性现象》86.1-2 (1995)，第 53–63 页。[45] Kyoung-Jin Lee 等。“反应扩散系统中自我复制斑点的实验观察”。载于：《自然》369.6477 (1994)，第 215–218 页。[46] Nathaniel D Virgo 等。“生命系统的热力学与结构”。博士论文。萨塞克斯大学，布莱顿，2011。[47] Irving R Epstein 和 Bing Xu。“纳米和微观尺度上的反应扩散过程”。载于：《自然·纳米技术》11.4 (2016)，第 312–319 页。[48] Marcin Fialkowski 等。耗散（动态）自组装的原理与实现。2006。[49] Gabriele M Coli 等。“基于深度学习的进化策略对软物质进行逆向设计”。载于：《科学进展》8.3 (2022)，eabj6731。[50] Nikolay Perunov、Robert A Marsland 和 Jeremy L England。“适应的统计物理学”。载于：《物理评论*X*》6.2 (2016)，第 021036 页。[51] Jeremy L England。“驱动自组装中的耗散适应”。载于：《自然·纳米技术》10.11 (2015)，第 919–923 页。[52] Philip W Anderson。“多即不同：对称性破缺与科学的层级结构本质”。载于：《科学》177.4047 (1972)，第 393–396 页。

[53] A. M. Turing。“I.—计算机器与智能”。载于：*Mind*LIX.236 (1950)，第 433–460 页。[54] Lothar Collatz。“微分方程的数值处理”。载于：柏林：施普林格(1966)。[55] 谢尔盖·约费和克里斯蒂安·塞格迪。“批归一化：通过减少内部协变量偏移加速深度网络训练”。载于：国际机器学习大会。PMLR。2015，第 448–456 页。[56] P·格雷和 S·K·斯科特。“等温连续搅拌釜反应器中的自催化反应：孤立环及其他多稳态形式”。载于：化学工程科学 38.1 (1983)，第 29–43 页。[57] 戴维·安杰利。“化学反应网络动力学教程”。载于：*2009 年欧洲控制会议（ECC）*。IEEE。2009，第 649–657 页。[58] Hitoshi Mahara 和 Tomohiko Yamaguchi。“非平衡反应扩散系统中熵平衡方程的计算”。载于：《熵》 12.12 (2010)，第 2436–2449 页。[59] 迪德里克·P·金马和吉米·巴。“Adam：一种随机优化方法”。载于：国际学习表征会议（海报）。2015。[60] Fran¸cois Chollet 等。*Keras*。 `https://keras.io`。2015。

附录

A 杂项方法详情

我们在此进一步以文字和数学形式描述这些方法的关键方面。更多详情请参阅源代码，地址为 `https://github.com/hunterelliott/dense-rdn`。

A.1 最小化数值积分误差

对于高度非线性的微分模型和简单的欧拉积分格式，我们必须谨慎确保优化结果仍能反映物理化学模型，而非仅仅是数值积分误差造成的假象。我们通过多种方式将这些误差的影响降至最低。

首先，我们通过在损失函数中加入惩罚项来控制步长 $\Delta \boldsymbol{X}_t$。在 3.1 中，这是损失函数（公式 8）的自然组成部分，旨在促进稳定图案形成；在这些实验中，我们设为 $\lambda_\Delta = 100$。对于其余实验，我们仅对超过阈值 δ_{Max} 的步长施加惩罚：

$$\mathcal{L}_\delta(\boldsymbol{\theta}, \delta_{Max}) = \lambda_\delta \frac{1}{UVT} \sum_{(u,v),t} \mathrm{Max}\left[|\Delta \boldsymbol{X}_t(u,v)| - \delta_{Max}, 0\right] \tag{15}$$

除 3.3 中的实验使用 $\lambda_\delta = 1.0$ 和 $\delta_{Max} = .3$ 外，所有实验均使用 $\lambda_\delta = 1000$ 和 $\delta_{Max} = .05$。因此，完整的损失函数是 $\mathcal{L}_\delta$ 与各实验中所述损失函数之和。

此外，对于所呈现的所有结果，我们都验证了：如果将时间步长 Δt 缩小两个数量级（从 1.0 降至 .01）来运行正向模拟，结果仍然不变。缩小时间步长应能显著降低数值误差，因此结果保持不变这一事实表明，它们不太可能是数值伪影。

A.2 局部浓度可视化

当待可视化的化学物种数量 N_s 为 >3时，我们首先对浓度维度进行主成分分析（PCA）投影。我们仅保留最大的 3 个组件，并分别将它们映射到 RGB图像中红、绿、蓝三个通道的强度，以便进行可视化。此处所呈现的所有情况中，这三个组件解释了 $> 90\%$ 的浓度方差。请注意，这种方法可以近似可视化任意复杂的化学物种混合物，但无法加入浓度颜色标尺。不过，这并不重要，因为浓度和反应速率常数都是在优化过程中同时确定的；二者都可以任意重新缩放，因此我们实际上使用的是任意浓度单位。

A.3 扩散建模

我们使用拉普拉斯算子的各向同性离散近似 [54] 来近似方程（5）。在所有情况下，扩散系数均被限制在.05 至 $.02 \times 10^{-5} \mathrm{m}^2/\mathrm{s}$ 之间。下限用于防止化学物种不断累积以及浓度出现空间不连续。鉴于 1s 的时间步长和离散拉普拉斯算子近似，上限则用于避免数值伪影（棋盘格图案、振荡）。采用这些扩散系数单位时，$\boldsymbol{X}$ 中单个空间元素（即“像素大小”）的物理尺寸为.01m；不过，由于扩散系数和浓度都是同时优化的，且可以任意重新缩放，因此所有单位实际上都是任意的。除 3.3 节外，所有空间域均为 64×64 像素；3.3 节中的空间域尺寸为 48×96。

A.4 神经网络

生成器模型 $G(\boldsymbol{z},\ \boldsymbol{\theta}_G)$ 是一种简单的架构，由重复的步幅为 2 的转置卷积、批归一化 [55] 和 a $tanh$ 激活组成。因此，最终生成的空间域宽度和高度均为 $2^{N_{blocks}}$，相应地，所展示结果中的模块数量也有所不同。输出还会经过一个固定的二维高斯核（$\sigma = 1.0$ 像素）滤波，以避免引入空间不连续性，从而避免扩散模拟和扩散熵率计算中的数值误差。最后，我们在最后一层省略批归一化（与 DCGAN [38] 中的做法相同），改为仅添加一个可优化的缩放因子，并对其加以约束，以防止产生负浓度，并将生成的最大浓度限制为 10.0（任意单位）。

3.2.1 中使用的编码器模型采用了类似的结构：步幅为 2 的卷积后接 a $tanh$ 激活和批归一化。

实验并不充分，且受到硬件限制；但总体而言，我们没有发现使用 ReLU 激活函数或更高容量的生成器或编码器架构，在收敛特性或最终已优化值方面有明显改善。

B 化学反应网络

B.1 入门示例反应网络

图 1 所示的动力学源自一个“耦合 Gray-Scott”反应系统，该系统由两个标准的 Gray-Scott反应系统 [56] 组成，并通过一个额外的反应相连，从而使两种自催化剂能够相互转化。

B.2 稠密化学反应网络

此处使用的反应网络之所以称为“稠密”，是因为它们包含了与特定原型匹配的所有可能反应。也就是说，如果一个反应网络包含物种$\mathcal{S} = \{A, B, C, D\}$，那么对于式 (3) 中的第一个反应，我们将纳入：

$$\begin{aligned} A + B &\rightleftarrows 2C \\ A + C &\rightleftarrows 2B \\ B + C &\rightleftarrows 2A \\ A + B &\rightleftarrows 2D \\ &\dots \end{aligned} \tag{16}$$

至于第二个反应 $A \rightleftarrows B$，我们会纳入所有可能的可逆单分子相互转化，例如 $A \rightleftarrows B$、$A \rightleftarrows C$、 $A \rightleftarrows D$、 $B \rightleftarrows C$，等等。更正式地说，我们会为（3）中给出的每种反应原型生成相应的反应集合：

$$\begin{aligned} \mathcal{R}_{3.1} &= \{X + Y \rightleftarrows 2Z | \forall X, Y, Z \in \mathcal{S}, X \neq Y, Y \neq Z, X \neq Z\} \\ \mathcal{R}_{3.2} &= \{X \rightleftarrows Y | \forall X, Y \in \mathcal{S}, X \neq Y\} \\ \mathcal{R}_{3.3} &= \{X + 2Y \rightleftarrows 3Y | \forall X, Y \in \mathcal{S}, X \neq Y\} \end{aligned} \tag{17}$$

这些反应集合中的每个反应都由独立的正向和逆向反应速率控制，这些速率是优化过程中确定的自由参数。例如，一个含 5 种物种的稠密化学反应网络共有 40 个反应（包括正向和逆向反应），而一个含 4 种物种的稠密化学反应网络共有 24 个反应。

为对这些化学反应的动力学建模，我们假设其遵循标准的质量作用动力学。采用化学反应网络理论中常用的矩阵表示，反应速率的计算 $\Delta \boldsymbol{X}_{t_{Rxn}}$ 是矩阵乘法 [16][57], ，可在用于前向模拟和优化的图形处理器上高效完成。

B.3 化学反应网络的动机

我们选择了上述反应网络，既因为它包含多种反应，也因为它具有产生多样行为的潜力。这为优化过程提供了广阔的空间，可供探索各种可能的反应集合和动力学，从而无需预先指定，就有机会产生有趣的行为。

这些反应涵盖了许多，甚至是大多数典型的单分子、双分子和三分子反应。在时间尺度足够长的情况下，更复杂的反应机制可以用这些更基元的反应来近似。具体而言，它们包含了经深入研究且表现出复杂行为的反应扩散模型中的反应，例如 Gray-Scott反应系统 [56] 或布鲁塞尔振子 [4]。因此，化学反应网络理论根据这些反应的拓扑结构和化学计量关系指出，复杂动力学至少是可能的，这并不令人意外 [16]。

B.4 已优化的反应网络

优化过程实际上是从上述模型能够表示的众多可能网络中选出一个特定的反应网络。我们可以将这个反应网络可视化为一张图：箭头从反应物指向产物，箭头样式则表示该反应的“强度”：颜色较深、尺寸较大的箭头表示速率常数较大；图的布局则使反应较强的物种彼此更靠近（箭头更短）。

这使我们能够比较不同损失函数得到的已优化反应网络，例如，3.2 节中用于最大化耗散的损失函数（图 6a）与 3.3 节中用于诱导类似复制动力学的损失函数（图 6b）所得到的网络。值得注意的是，这些只是反应速率常数，而通过这些反应途径的净通量取决于状态 $\boldsymbol{X}$ 及其随时间的变化。

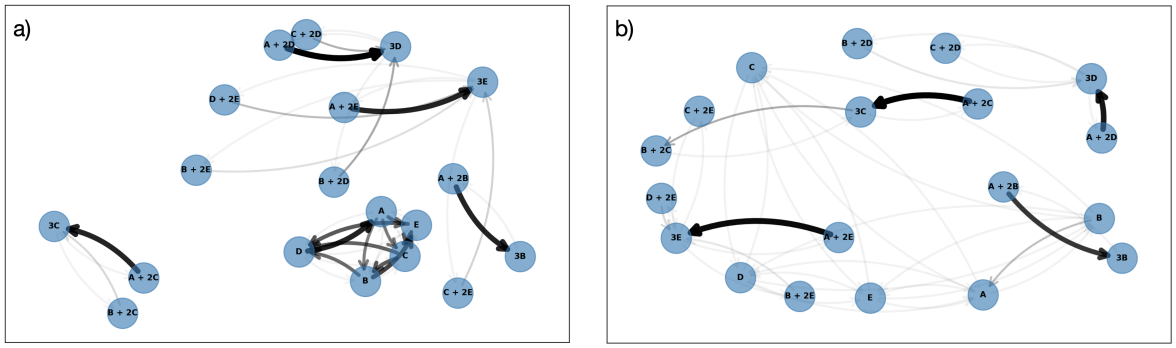

图6：耗散最大化得到的已优化反应网络图的可视化结果（a），对应图3a中的结果；以及复制样动力学得到的反应网络（b），对应图 5中的结果。颜色更深、线条更粗的箭头表示更大的反应速率常数。

C 热力学

单个体积元 $\boldsymbol{X}(u,\ v)$ 因化学反应而产生的瞬时熵产生率为 [37][58]:

$$\left(\frac{\partial S}{\partial t}\right)_{Rxn} = \sigma_{Rxn} = k_B \sum_j \left(\nu_j^+ - \nu_j^-\right) \ln\left(\frac{\nu_j^+}{\nu_j^-}\right) \tag{18}$$

其中，ν_j^+ 和 ν_j^- 是该体积元中反应 j 的局部正向和逆向反应速率，k_B 是玻尔兹曼常数，因此单位为 $\frac{\partial S}{\partial t}$ J K·L·s。在实际计算中，为保证数值稳定性，~~我们将~~ν_j ~~替换为~~ Max $(\nu_j, 1\times10^{-5})$。

体积元$\boldsymbol{X}(u,v)$中由扩散导致的瞬时局部熵产生率为 [58]:

$$\left(\frac{\partial S}{\partial t}\right)_{Dif} = \sigma_{Dif} = k_B \sum_i \frac{D_i}{\boldsymbol{X}^i(u,v)} \left(\nabla \boldsymbol{X}^i(u,v)\right)^2 \tag{19}$$

其中，D_i 是物种 i 的扩散系数，而 $\boldsymbol{X}^i(u,v)$ 是位置 (u,v) 处物种 i 的浓度。我们在公式（19）的分母中加入一个很小的数值稳定项（0.1），以避免除以零。

The 总熵产生率就是这两个组件之和 :

$$\sigma_{Tot} = \sigma_{Rxn} + \sigma_{Dif} \tag{20}$$

我们不对该速率进行积分来得到熵的净变化 $\Delta S_{Tot} = \sigma_{Tot}\Delta t$，而是报告并最大化瞬时速率的平均值，以避免对优化的时间和长度尺度产生显式依赖。

D 优化

D.1 初始化

正如神经网络优化中普遍观察到的那样，我们发现初始化对收敛至关重要。具体而言，在我们的案例中，使用均匀随机值初始化、但反应速率较低的模型（$k_{fi},\ k_{ri} < 1\times10^{-3}\ \forall\ i$），以及驱动流速和浓度同样较低的模型，都未能收敛。初始化于耗散性更强状态的模型——其反应速率和驱动流速更高——则更容易收敛。全面探索初始化依赖性所需的计算量大到难以承受，因此我们转而模拟了用于 [1], 的动力学参数，并将其推广到我们的稠密化学反应网络：每种化学物种初始化一个自催化反应，其速率常数为 1。所有其他反应速率均设为 $1\times10^{-3}+\epsilon_r$，其中 $\epsilon_r \sim$ 均匀分布（$-1\times10^{-4},\ 1\times10^{-4}$），因为我们发现这一数值足够低，能够避免浓度过高及其相关的数值误差和/或反应速率失控。微小的均匀随机噪声 ϵ_r 打破了反应速率的对称性。

流动反应器驱动的进料浓度 x_i 初始化为 $\frac{1}{4^{N_{a_i}}}$，其中 N_{a_i} 是产生物种 i~~的自~~催化反应数量。这种缩放方式经验证可产生一个耗散性很强、但在动力学和数值上具有一定稳定性的初始状态。

扩散系数初始化为均匀分布 $(.05, .2) \times 10^{-5} \mathrm{m}^2/\mathrm{s}$，在数值稳定区域内均匀取值。

D.2 启发式方法和超参数

图	T	lr	学习率按 1 2 在学习率早停耐心值时降低		早停耐心值	Adam β_1
2a	32	1×10^{-3}	(5，10，15) $\times 10^2$	5×10^2	1×10^3	.95
2b	64	1×10^{-3}	(5，10，15) $\times 10^2$	5×10^2	1×10^3	.95
3a	64	1×10^{-3}	-	2×10^3	6×10^3	.995
3b	64	1×10^{-3}	-	2×10^3	6×10^3	.995
4	128	2.5×10^{-4}	(1，2) $\times 10^4$	2×10^3	6×10^3	.995

表 1：所呈现结果的超参数。lr：学习率；lr decrease：按固定计划降低学习率时对应的迭代次数；lr patience：自动降低学习率的耐心参数；es patience：自动早停的耐心参数。lr decrease 和 patience 列的单位均为迭代次数（在此语境中，“epoch”没有意义）。

我们使用 ADAM 优化器 [59], ，学习率和移动平均参数 β_1 均如表 1 所示（所有实验中 β_2 均设为 .999）。对于这些高度非线性的动力学模型，我们发现，尤其是在时间尺度较长的优化过程中，梯度裁剪对于保证收敛稳定性至关重要。我们使用 Keras [60]中实现的梯度范数裁剪，将阈值设为 0.5。我们通过 Keras 的回调函数，按照表 1 所示的耐心参数自动降低学习率并自动终止优化。在一些实验中，我们使用均方根偏差代替式 7 中给出的 L2范数，以便在改变空间域大小时无需调整损失函数的量级；正如预期，这样得到的结果与使用 L2范数相近。

使用这些模型计算梯度需要通过整个时间序列 $\mathcal{X}$进行反向传播，这会带来显著的内存开销。因此，考虑到大多数情况下我们只针对单个 $\boldsymbol{X}_0$进行优化，除第 3.2.1 节使用 2 外，我们将批量大小设为 1。对于包含耗散损失的实验，除第 3.2.1 节设为 0.04 外， λ_σ 均设为 1.0。优化过程中，反应速率常数和进料浓度被限制为非负值，流速被限制在 0.01 到 1.0 之间。扩散系数被限制在其初始化范围（A.3）内。

D.3 增量优化

对于 3.3 中展示的实验，我们必须针对足够长的时间尺度进行优化，以使完整的类复制动力学得以发生。在超过 ~ 250 个时间点（相当于 250 秒）的情况下，从随机初始化开始优化极不稳定，这在很大程度上是因为在高度非线性的化学反应网络中，难以平衡梯度消失与梯度爆炸。最终，我们通过一种“渐进式”优化方法实现了收敛：先在较短的时间尺度上进行优化，再使用较短时间尺度下的参数 $\boldsymbol{\theta}^*$ 来初始化较长时间尺度下的优化。此外，在最长的时间尺度下，方差目标

优化	T	β^*	lr	学习率在 1 2 处降低
1	256	1.0	1×10^{-3}	(1、2、3、50、100、150) $\times 10^2$
2	320	1.0	5×10^{-6}	(5、10) $\times 10^2$
3	352	0.5	1×10^{-6}	(20、40) $\times 10^2$

表 2：产生图 5 中结果的增量优化阶段。每次优化的结果都用于初始化下一次优化。

β^* 似乎与复制保真度损失（式 13）相悖；降低这一数值后，收敛时的亲代与子代相关性更好。值得注意的是，所需的迭代次数非常多，导致优化耗时很长；此处给出的最终结果总共需要 > 600k 次权重更新，实际运行时间超过 26 天。生成图 5 的各项优化所用的超参数列于表 2。所有优化中，ADAM 优化器 β_1 均设置为 .95。

D.4 对照分析

我们进行了额外分析，以检验模型各组件优化的必要性和充分性。这些计算性“消融实验”对应以下几种条件：

- “已优化”——所有组件均已优化。
- “随机 $\boldsymbol{X_0}$”——将优化后的初始条件 $\boldsymbol{X_0}$ 替换为正态分布随机噪声，其均值和方差与优化后的初始条件相同（截断噪声以防止出现负浓度）。
- “初始 $\boldsymbol{\theta}_\Psi$”——将已优化的反应网络参数和扩散系数替换为优化开始时的取值（如 D.1 所述）。
- “随机 $\boldsymbol{\theta}_\Psi$”——将已优化的反应网络参数和扩散系数替换为正态分布随机噪声，其均值和方差与已优化参数相同（截断噪声，以防止反应速率为负，或扩散系数超出数值稳定范围）。

这些实验的结果见图 2c 和图 3c，另有一些结果展示于此并在下文描述。这些结果证实，所有已优化的组件对于与目标保持持续相关性以及维持较高的耗散率都是必不可少的。

图 7a 展示了与正文相同的对照实验，但针对的是第二个示例（见图 2b）。随后，我们对两个示例重复相同的分析，但将时间步长缩短了两个数量级，以进一步证实稳定性和保真度并非优化利用数值误差所致（图 7b-c）。

在图 8a 中，我们展示了与正文相同的结果，但针对第二个示例（见图 3b）。我们还以更长的时间尺度运行了这两个示例，以证明

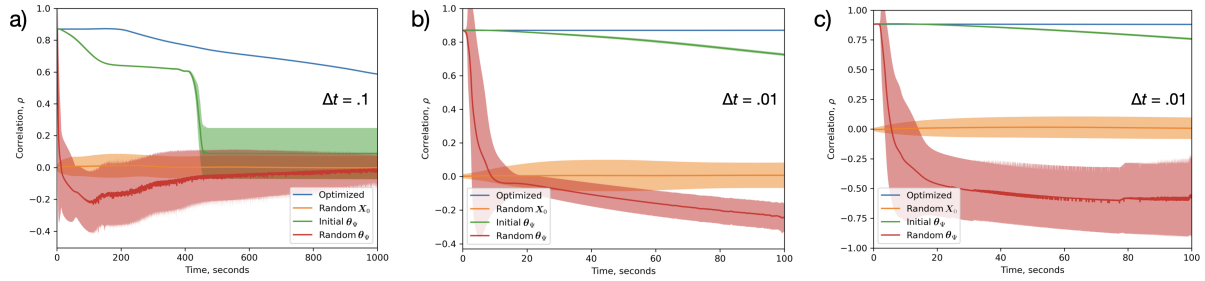

图7： (a) 图 2b 中示例与目标的相关性随时间的变化。本图和图 2c 均采用优化过程中所用时间步长的十分之一 $1s$ 。(b-c) 对图 2b 中的模型（见 b）和图 2a 中的模型（见 c）进一步将时间步长减小至 $.01s$ ，提供了更多证据，表明这些差异并非由数值误差导致。实线表示均值，色带表示在 $n = 32$ 次随机化中，均值上下一个标准差的范围 +/-。

优化后的状态具有稳定的耗散性，其熵产生速率始终高于含有未优化组件的状态（图 8b-c）。请注意，在这种情况下，均值和标准差匹配的随机动力学参数模型（“Rando m θ_Ψ”，以红色显示）极不稳定，并表现出显著的数值积分误差（其行为会随较小时间步长而改变，这印证了这一点）。这进一步证实了 A.1 节所述数值误差规避方法的必要性和有效性。

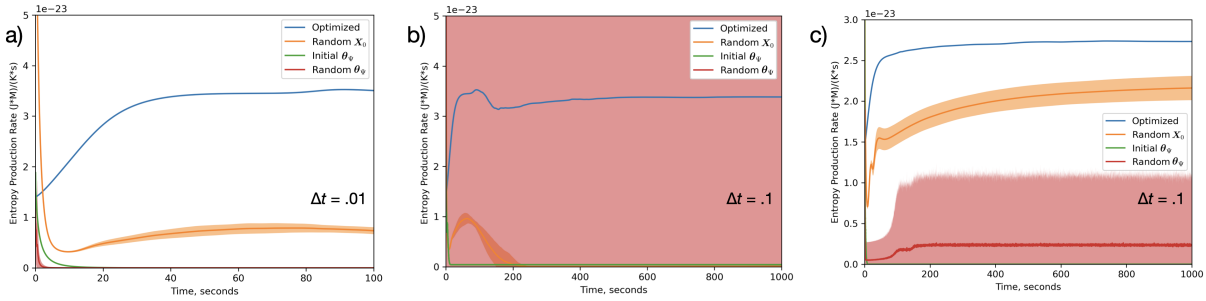

图8： (a) 图 3b 中示例的耗散率随时间的变化。本图和图 3c 均采用优化过程中所用的时间步长 $1/100th$ $1s$ ，以确认这些差异并非由数值误差导致。(b-c) 将时间尺度增大十倍以上后，证实这些是稳定的非平衡态，且优化所带来的耗散率增加在图 3b（见 b）和图 3a（见 c）中的模型上均持续存在。实线表示均值，色带表示在 $n = 32$ 次随机化中，均值上下一个标准差的范围 +/-。请注意，在 b) 和 c) 中采用 $.1s$ 时间步长时，具有随机动力学的模型表现出较高的变异性，这通常是由数值误差导致的（红色色带较宽）。