带乘积约束的极小极大优化的复杂度

Martino Bernasconi Matteo Castiglioni 博科尼大学 米兰理工大学martino.bernasconi@unibocconi.it matteo.castiglioni@polimi.it

arXiv:2602.04665v1 [cs.CC] 4 Feb 2026

摘要

我们研究具有非凸非凹效用函数 f 的博弈中，计算局部极小极大均衡的计算复杂性。根据 Daskalakis 、 Skoulakis 和 Zampetakis [DSZ21]，的研究，我们已知在一种受限情形下，该问题是困难的：参与者必须采用受到联合约束的策略，因此，在更自然的约束条件下，该问题的复杂度仍未明了。本文解决了这一问题，并证明即使在乘积约束下，尤其是在超立方体上，该问题仍是 PPAD 困难。

1 引言

本研究探讨了寻找连续光滑函数的局部极小极大点这一基本问题

$$f : [0,1]^d \times [0,1]^d \to [0,1].$$

鞍点的计算是数学优化中的一项基础工作，在许多问题中有着重要的实际应用，范围涵盖生成对抗网络（GAN）的训练 [Goo+14; ACB17] 和对抗训练 [Mad+17; SND18] ，以及对齐[Mun+24]。

过去十年间，这类问题的重要性推动了对能够收敛到此类解的算法的理论研究。例如，在非凸-凹情形下，许多算法以多项式速率收敛 [OLR21; Jin+21; LJJ20; The+19; LJJ25]。对于更一般的非凸-非凹情形，现有结果仅限于局部收敛 [Mer+18; WZB19; Bal+18; MSJ25] ，或者需要额外的结构条件 [DDJ21; ALW21; KYP24]。

尽管这一问题十分重要，其精确的算法刻画仍未明了。具体而言，目前没有已知算法能够在 poly(1/ϵ) 时间内以可证明的方式找到 ϵ-近似解。

这一理论上的差距，反映了实践中一种有据可查的挫折感，尤其是在训练生成对抗网络和对抗系统时 [XSC21; Kod+17; HMC21; VGFP19]。这种困难通常被归因于极小极大问题局部解的非凸性。然而，这种解释似乎并不令人满意，因为在非凸最小化中寻找驻点相当容易处理（另请参见 [DSZ21] 中的讨论）。本文证明了这一失败背后更根本的原因：寻找局部极小极大点在计算上是困难的。具体而言，我们证明，除非 PPAD =P，否则不存在 poly(1/ϵ) 时间算法。

更形式化地说，我们证明以下称为 GDA（梯度下降-上升，见问题 1）的问题是 PPAD 困难：给定一个可微函数 $f : [0,1]^d \times [0,1]^d \to [0,1]$，求 $(x,\ y)$，使得

$$\max_{x' \in [0,1]^d} \langle \nabla_x f(x,y), x' - x \rangle \le \epsilon \quad \text{and} \quad \max_{y' \in [0,1]^d} -\langle \nabla_y f(x,y), y' - y \rangle \le \epsilon.$$

该问题要求找到一个点（$x,\ y$），使最大化玩家 x 和最小化玩家 y 都处于局部最优点，或者等价地，找到梯度下降-上升动力系统的不动点。

Daskalakis 等人 [DSZ21], 首次从计算复杂性的角度分析了这个问题，并证明它属于 PPAD。他们还证明了其 PPAD 困难性，但仅限于这样的情形： x 和 y 玩家只能在超立方体 $[0,1]^{2d}$的一个耦合（即非乘积）子集中进行策略选择。

正如后续许多研究 [BR21； Fea+22； Ber+24； HMN24； Ana+25]所指出的，在更自然的乘积约束设定下刻画极小极大优化的复杂度，始终是一个根本性的未解决问题。最近，Bernasconi 等人 [Ber+24]简化了 Daskalakis 等人 [DSZ21], 的归约，改进并推广了这一结果，同时强调了耦合约束在该归约中的核心作用。

在本文中，我们最终解决了这一未决问题，并证明即使在乘积约束下，尤其是在超立方体上，寻找局部极小极大点也是 PPAD 困难的。

计算复杂性从形式上说，GDA 是 NP 中的一个全搜索问题；也就是说，它是一个总有解的 NP 问题。Megiddo 和 Papadimitriou [MP91]表明，这类问题（即 TFNP）不太可能存在任何完备问题。Papadimitriou [Pap94]

随后定义了 TFNP 的若干子类，每个子类都对应一种非构造性（且低效）的证明原理，用以保证问题总有解。与我们的工作最相关的是 PPAD，它依赖于有向图上的奇偶性论证。人们发现，这一类适用于各种不动点问题，最著名的例子是纳什均衡 [DGP09; CDT09]。

最近，Fearnley 等人 [Fea+21; Fea+22] 证明，对于一个光滑连续函数，寻找其梯度下降的不动点是 CLS = PPAD ∩ PLS 类的完全问题。这表明，GD 同时是一个不动点问题和一个局部搜索问题。具体而言，很容易看出，寻找梯度下降不动点的问题可以归约为以下（非正式）问题（该问题曾用于定义 CLS [DP11]）：给定一个连续的“势函数”$p : [0,1]^d \to [0,1]$ 和一个连续的“改进函数” $g : [0,1]^d \to [0,1]^d$，寻找一个 x，使得 $p(g(x)) \geq p(x) - \epsilon$。显然，$g$ 的不动点就是这个问题的解。本文证明，与梯度下降不同，梯度下降-上升（GDA）是一个不含势函数的纯不动点问题。这与极小极大问题中广为人知的现象相符：在这类问题中，算法往往会自然地陷入循环。

技术挑战

近期关于纳什均衡计算的困难性结果，源于模仿[BR21; Bab16; Rub17; Rub15]这一思路。从较高层面来看，给定一个 F 向量值函数，其不动点很难找到，x 玩家（不动点玩家）会最小化 x 与 $F(y)$之间的距离，而 y 玩家（模仿玩家）会最小化 x 与 y之间的距离。然而，这种方法要求 x 玩家和 y 玩家采用不同的目标函数。极小极大问题的根本挑战在于，我们不得不让两个参与者采用相反的目标函数。因此，迫使一名参与者模仿另一名参与者，就会迫使另一名参与者逃离。这是需要克服的主要概念性障碍之一。

直观而言，我们不再试图强制实现 $x \simeq y$。相反，我们的构造结合了两个 PPAD 困难问题：LinVI [Ber+24]，当 $x \simeq y$时很难求解；以及 Pure-Circuit [Del+22]。具体来说，我们允许 Pure-Circuit 构件失效，但在这种情况下，构造会强制 $x \simeq y$，从而得到 LinVI 问题的解；该问题同样是 PPAD 困难问题。在下一节中，我们将对证明作高层次的技术概述。

2 证明概述

耦合约束设定中的所有已知归约 [DSZ21；Ber+24；Ana+25]都使用形如 $f(x,y) = F(x)\ x - y$ 的效用函数。特别是，Bernasconi 等人 [Ber+24] 明确揭示了带耦合约束的极小极大优化与变分不等式（VIs）之间的简单联系。变分不等式是寻找一个点 z，使得对所有$z' \in [0,1]^d$均有 $F(z)\ z' - z \geq 0$ 的问题。众所周知，对函数 F 的许多选择（例如线性函数）都会使相应的变分不等式成为 PPAD 困难问题。将 F 作为变分不等式算子，并结合耦合约束$x \simeq y$，足以证明耦合约束下的 PPAD 困难性。事实上，我们可以观察到，$\nabla_x f(x,y) = F(x) + J_F(x)(x-y)$，且 $\nabla_y f(x,y) = -F(x)$。“噪声项” $\Delta(x,y) = J_F(x)(x-y)$ 在耦合约束下可以忽略，从而使该问题实质上等价于原始变分不等式实例。[1] 控制 $\Delta(x,y)$ 在 $\nabla_x f(x,y)$ 中的大小，正是$x \simeq y$ 约束的用途。

[1]在本节中，我们相当非正式地使用符号 Δ 来表示我们希望保持较小的量。

摆脱耦合约束的主要挑战在于寻找处理噪声成分 $\Delta(x,\ y)$ 的替代方法。下面的讨论旨在直观地概述我们所采用的技术。

Pure-Circuit

我们首先考虑一个不同的算子 F。具体而言，我们从 Pure-Circuit [Del+22] 归约而来，该问题涉及为图中的节点寻找一个赋值，使其满足一组门（约束），这些门对顶点的取值施加限制。Pure-Circuit 问题特别好处理，原因在于：(1) 它作用于连续值 $[0,1]$，但只关心纯值 {0,1}，并将中间值 (0,1) 视为"垃圾值"（记为 $\perp$）；(2) 其依赖图结构非常规整。这使我们能够具体地推理变量之间的影响，并处理 $\Delta(x,y)$ 分量。

对于 Pure-Circuit 实例中的每个节点 v，我们引入一个由两个变量 x^v, y^v 进行的局部博弈，并将对 v 的赋值解释为：若 $(x^v - y^v)^2$ 较大则为 1，若 $(x^v - y^v)^2$ 较小则为 0。对于一个门的每个输出节点 q，不难实现一个光滑函数 $s_q(x,y)$，用以计算应赋给该门输出的数值。特别地，$s_q(x,y)$ 仅依赖于作为 q 之门输入的节点 v 的取值 $(x^v - y^v)^2$。

挑战在于确保一致性：我们必须强制要求，在该实例的解中，输出节点 q 上的值 $(x^q - y^q)^2$ 与 $s_q(x,y)$ 计算出的值一致。在这种语言中，一致性意味着，如果 $s_q(x,y) = 1$，那么 $(x^q - y^q)^2$ 就必须很大；如果 $s_q(x,y) = 0$，那么 $(x^q - y^q)^2$ 就必须很小。

蓝图

这里，我们概述一种初步的归约尝试。我们将效用函数定义为

$$f(x,y) = \sum_{q \in V} s_q(x,y) H_q(x,y), \tag{1}$$

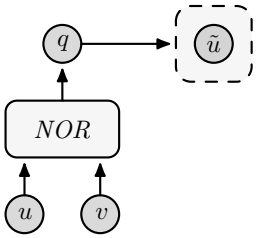
其中，关键在于，$s_q(x,y)$ 是用于计算 q 所属门的值的函数（因此只取决于节点 q 输入端的变量），而 $H_q(x,y)$ 是一个"连接函数"，它只取决于与节点 q 本身相关的变量（稍后将解释 H_q 这一术语）。

对于固定节点 q，我们可以将 f 写成更能突出其对变量 (x^q, y^q) 的依赖关系的形式。确实：

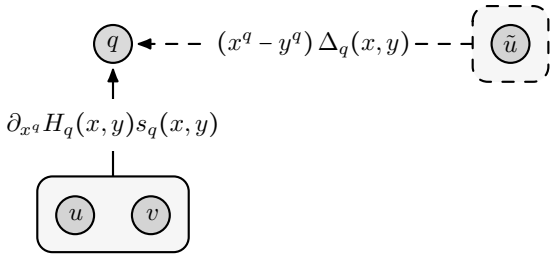
$$f(x,y) = \underbrace{s_q(x,y) H_q(x,y)}_{(A)} + \underbrace{\sum_{\tilde{u}:\, q \text{ is input to } \tilde{u}} s_{\tilde{u}}(x,y) H_{\tilde{u}}(x,y)}_{(B)} + \underbrace{\sum_{\text{all other nodes } u} s_u(x,y) H_u(x,y)}_{(C)}. \tag{2}$$

在（C）中，变量（$x^q,\ y^q$）并未出现；而在（B）中，它们只出现在 $s_{\tilde{u}}$，在（A）中则只出现在 H_q（$x,\ y$）中。这就是我们将 H_q 称为"连接函数"的原因，因为若没有它，f 关于（$x^q,\ y^q$）的偏导数就不会依赖于 s_q。特别是，为了保持一致性，我们希望节点 q 处的参与者激励只取决于 q 作为输出所对应节点中的变量。

因此，计算效用函数的偏导数时，我们只需考虑 (A) 和 (B) 的贡献。项 (A) 表示对节点 q 的期望影响。另一方面，那些



(a) 一个简单的 Pure-Circuit 实例示例。虚线框内是电路的所有其他节点 $\tilde{u}$ 和门。



(b) x^q 对 f的影响。虚线箭头表示噪声项，而实线箭头表示期望的那一项（分别来自公式（2）中的（B）项和（A）项）。

图1：一个简单 Pure-Circuit 实例的依赖图，以及变量 x_q 对公式（1）中定义的效用函数 f 的影响图。

来自（B）的是噪声分量，它们对节点 q处的变量引入了额外的依赖。我们将该噪声项封装为一个通用项 $\Delta_q(x,y)$，希望它仅依赖于 s_q，以反映 Pure-Circuit 实例的依赖结构。因此，偏导数为

$$\partial_{x^q} f(x,y) = \partial_{x^q} H_q(x,y) \cdot s_q(x,y) + 2(x^q - y^q)\Delta_q(x,y),$$

and

$$\partial_{y^q} f(x,y) = \partial_{y^q} H_q(x,y) \cdot s_q(x,y) - 2(x^q - y^q)\Delta_q(x,y),$$

其中我们利用了 $\partial_{x^q} s_v(x,y) = -\partial_{y^q} s_v(x,y)$，以及 s_v依赖于$(x^q - y^q)^2$。在推导这些偏导数时，我们依赖于 Pure-Circuit 的两个关键特性：（1） Δ_q 仅包含 $s_{\tilde{u}}$ 在其他少数节点 $\tilde{u}$ 处的导数（即那些以 q 为输入的节点），以及（2） s_q 不依赖于与节点 q相关的变量。我们还参阅图1作为说明。

现在，我们确实希望项 $|\Delta_q(x, y)|$ 足够小。例如，假设 $|\Delta_q(x, y)| \ll 1$，并考虑简单的连接函数 $H_q(x, y) = x^q + y^q$，一阶最优性条件很快就能得出所需的一致性：

- If $s_q(x,y) = 1$, then

  $$\partial_{x^q} f(x,y)(x' - x^q) = (1 + 2(x^q - y^q)\Delta_q(x,y))(x' - x^q) \le 0 \quad \forall x',$$

  这意味着 $x^q = 1$。另一方面，

  $$-\partial_{y^q} f(x,y)(y' - y^q) = -(1 - 2(x^q - y^q)\Delta_q(x,y))(y' - y^q) \le 0 \quad \forall y',$$

  这意味着 $y^q = 0$，并且 $(x^q - y^q)^2$ 很大。

- 如果 $s_q(x, y) = 0$，那么

  $$\partial_{x^q} f(x,y)(x' - x^q) = 2(x^q - y^q)\Delta_q(x,y)(x' - x^q) \le 0 \quad \forall x',$$

  and

  $$-\partial_{y^q} f(x,y)(y' - y^q) = 2(x^q - y^q)\Delta_q(x,y)(y' - y^q) \le 0 \quad \forall y'.$$

  这两个方程意味着 $x^q = y^q$（根据 $(x^q - y^q)\Delta_q(x, y)$ 的符号，它们都等于 0 或 1），因此 $(x^q - y^q)^2$ 很小。[2]

[2]仅当 $\Delta_q(x, y) \neq 0$时，这才成立，但我们暂时可以忽略这个问题。在最终构造中，我们将通过复制变量来处理这一问题。

这基本上就完成了归约。然而，几乎显而易见的是，我们无法保证 $|\Delta_q(x,\ y)|$ 始终很小。以往的方法 [DSZ21；Ber+24；Ana+25] 基本上选择了$H_q(x,\ y) = x^q - y^q$，这只能确保 Δ_q 在超立方体中一个外生施加的、满足 $x \simeq y$的子集内很小。

这里我们采用一种不同的方法，使用不同的函数 H_q，并以另一种方式处理Δ_q。关于 H_q 的构造和一致性的讨论暂且搁置，我们先解释所用的处理 Δ_q的技巧。

猜测与副本

为克服额外项 Δ_q的影响，我们采用一种基于重复（副本）和正则化的技巧。对于每个节点 q，我们为变量 n 创建 $x^q = \{x_i^q\}^{i\in[n]}$ 个副本（ y^q也同样如此），并用 $\|x^q - y^q\|^2$定义门函数 s_q。随后，我们添加一个二次正则项$\varphi(x,\ y) = \sum_{q\in V}\sum_{i\in[n]} M_i(x_i^q - y_i^q)^2,$，其中 $\{M_i\}^{i\in[n]}$ 构成一个正则网格，包含从 $-M$到 M的 n 个点（ M 的范围稍后讨论）。 M_i充当项 Δ_q的“猜测值”。事实上， $f(x,\ y) = \sum_{q\in V} s_q(x,\ y)H_q(x,\ y) + \varphi(x,\ y)$的偏导数现在变为：

$$\partial_{x_i^q} f(x,y) = \partial_{x_i^q} H_q(x,y)\cdot s_q(x,y) + 2(x_i^q - y_i^q)(\Delta_q(x,y) + M_i), \tag{3a}$$

and

$$\partial_{y_i^q} f(x,y) = \partial_{y_i^q} H_q(x,y)\cdot s_q(x,y) - 2(x_i^q - y_i^q)(\Delta_q(x,y) + M_i). \tag{3b}$$

这为什么有帮助？如果 $|\Delta_q|$ 在 n 中增长得足够慢（例如次线性增长），那么至少会有一个$i \in [n]$（更准确地说，是少数几个），使得 $|\Delta_q + M_i|$ 很小，从而有效消除这些副本的 Δ_q 项。然而，我们添加的副本 i 越多，项 H_v、 $v \in V$就越大（因为它们依赖于更多变量），进而 Δ_q也会变大。

界定 $|\Delta_q|$

对于 $|M_i + \Delta_q|$ 较大的坐标 i，梯度中的主导项（来自正则化项）对参与者具有不同的符号，因而目标函数是对齐的。这迫使均衡满足 $|x_i^q - y_i^q|$ 很小。具体而言，我们将证明在均衡处， $|x_i^q - y_i^q| = O(|M_i + \Delta_q|^{-1})$，这表明 $\|x^q - y^q\|_1$ 被调和级数上界约束，因而其阶为 $O(\log n)$。现在我们回到对 $|\Delta_q|$ 的界定。到目前为止，构件 H_q 一直尽可能保持一般性，因为我们仅使用了它依赖于变量 x^q 和 y^q 这一事实。现在，我们为 H_q选择一种特定结构，并要求连接函数具有如下形式： $H_q(x,y) = \sum_{i\in[n]} G_i^q(x,y)\cdot(y_i^q - x_i^q)$。在此选择下，我们有，

$$|H_q(x,y)| = O(\|y^q - x^q\|_1) = O(\log n).$$

特别地，由于这对每个节点 $q \in V$都成立，我们可以从公式（2）看出，对每个 $q \in V$都有 $|\Delta_q| = O(\log n)$，其中我们忽略对其他参数的依赖。因此，只需选取网格 M 的任意一个在 n 上以超过对数级增长的范围，即可保证对许多索引i而言， M_i 接近 $-\Delta_q$，即 $|\Delta_q + M_i| = O(1)$。我们注意到，对于该量，我们只需证明一个次线性上界。

在 LinVI 构件中隐藏一致性违例

现在，我们必须证明电路中节点的一致性，即：1）若 $s_q(x,y) = 1$ 则$\|x^q - y^q\|^2$较大；2）若 $s_q(x,y) = 0$ 则 $\|x^q - y^q\|^2$ 较小。第二个条件很容易成立。如公式（3）所示，当 $s_q = 0$ 时，第 i 个变量的偏导数中存留的部分将与$(x_i^q - y_i^q)$成比例，并且对参与者而言符号相反（因此两名参与者具有一致的目标函数）。这迫使在均衡处对所有 i 都有 $x_i^q \simeq y_i^q$。

主要挑战在于第一个条件：如果 $s_q = 1$，我们需要 $\|x_q - y_q\|^2$ 足够大。对于"猜测正确"的索引 i（此时 $|\Delta_q + M_i|$ 很小），梯度的主导项来自 H_q。因此，我们必须设计一个特定的构件 H_q 来强制满足该条件。为了界定 $|\Delta_q|$，我们已经约束了 $H_q(x,\ y) = \sum_{i\in[n]} G_i^q(x,\ y)\cdot(y_i^q - x_i^q)$，所以现在必须设计 G_i^q 函数，使得当 $x_i^q \simeq y_i^q$ 时，H_q 没有均衡点。遗憾的是，根据不动点论证，不存在一个函数$H_q(x,\ y)$，它在 $x^q = y^q$ 处为零，却在 $x^q = y^q$ 处没有均衡点。[3]

为绕过这一障碍，我们采用构造的核心思路：将对称均衡隐藏在构件 H_q 中编码的一个 PPAD 困难问题里。如果无法找到一个不存在对称均衡的函数，我们就可以使用一个具有"难以找到的"（近似）对称均衡的函数。具体而言，我们允许违反一致性条件，但任何此类违反都会导向隐藏在 H_q 中的 PPAD 困难问题的一个解。

值得注意的是，所需的构件正是 Daskalakis 等人 [DSZ21] 和 Bernasconi 等人 [Ber+24] 已经使用过的从变分不等式到极小极大均衡的归约；为使归约尽可能简单，我们采用了与 Bernasconi 等人 [Ber+24] 相同的做法，使用线性变分不等式。不过在这里，我们不必显式加入约束 $x_i^q \simeq y_i^q$，因为所有满足 x_i^q 远离 y_i^q 的均衡都满足 Pure-Circuit 的一致性约束。这就确立了我们证明的核心二分。我们构造的 GDA 实例的任何解都必须满足以下两种情况之一：(i) 一致，从而满足 Pure-Circuit 的约束并给出 Pure-Circuit 的一个解；或 (ii) 至少在一个节点上不一致（即 $x_i^q \simeq y_i^q$ 即使在 $s_q = 1$ 时也如此），这就要求在 H_q 中找到一个对称均衡，进而给出 LinVI 的一个解。请注意，LinVI 实例必须隐藏在任意变量 q 和任意副本 i 中，而且变分不等式表述要求每个变量 x_i^q 都是 m 维的，其中 m 是 LinVI 实例的维数。由于 Pure-Circuit 和 LinVI 都是 PPAD 困难问题，我们得出结论：GDA 问题也必定是 PPAD 困难的。

注记 2.1。在 Fearnley 等人 [Fea+21] 的突破性成果中，归约也为一个 PPAD-困难问题生成了不需要的解。不过，在该情形下，情况在结构上有所不同，也比我们的设定深入得多。事实上，他们采用了一个 PLS-困难问题来隐藏额外的解，从而得到一个 PPAD-困难问题或 PLS-困难问题的解。在我们的情形中，我们认为使用"隐藏构件"的原因更多是出于实用考虑，而非根本性原因。最终，我们从 End-of-the-Line 出发，并由此导出两个问题（Pure-Circuit 和 LinVI）。

3 预备知识

对于 $x, y \in \mathfrak{R}^d$，我们用 x, y 表示标准欧几里得内积，并将相应的范数记为 $\|x\| = \sqrt{x, x}$。投影 Π_D 在凸闭集 D 上的定义依据

[3]我们在此简要给出该陈述的直观证明。任取一个 $x^q = y^q = z$。那么，由于 $H_q(z,\ z) = 0$，对于所有 $i \in [n]$，都有 $\partial_{x_i^q} H_q(z,\ z) = -\partial_{y_i^q} H_q(z,\ z) := S_i(z)$。因此，映射 $z \mapsto \Pi_{[0,1]^n}(z + S(z))$ 的任何不动点都满足均衡条件，而根据布劳威尔不动点定理，这些不动点必然存在。

于 ℓ_2-范数。我们关注利普希茨函数与光滑函数。一个连续可微函数在 $[0,1]^d$ 上是 G-利普希茨的，如果对所有 $x, x' \in [0,1]^d$ 都有 $|f(x) - f(x')| \le G\|x - x'\|$，并且当 $\|\nabla f(x) - \nabla f(x')\| \le L\|x - x'\|$ 时是 L-光滑的。我们用 $\mathfrak{R}_+$ 表示正实数集。

我们考虑以下与寻找梯度下降—上升动力学的不动点相关的全搜索问题。

问题 1 (GDA)。给定 ϵ, L, G, $B \in \mathfrak{R}_+$，两个实现 G-利普希茨且 L-光滑函数 $f : [0,1]^d \times [0,1]^d \to [-B, B]$ 及其梯度 $\nabla f : [0,1]^d \times [0,1]^d \to \mathfrak{R}^{2d}$ 的电路，求 $(x^\star, y^\star) \in [0,1]^{2d}$ 使得

$$\frac{\partial f(x^\star, y^\star)}{\partial x_i}(x_i - x_i^\star) \le \epsilon \quad \forall i \in [d], x_i \in [0,1],$$

and

$$-\frac{\partial f(x^\star, y^\star)}{\partial y_i}(y_i - y_i^\star) \le \epsilon \quad \forall i \in [d], y_i \in [0,1].$$

注 3.1。在证明中，按单个分量定义 GDA 问题会更方便。不过，更常见的定义要求求出 $(x^\star, y^\star) \in [0,1]^{2d}$ 使得 $\nabla_x f(x^\star, y^\star)\ x' - x^\star \le \epsilon$ 且 $-\nabla_y f(x^\star, y^\star)\ y' - y^\star \le \epsilon$ 对所有 $(x', y') \in [0,1]^{2d}$ 均成立。这些定义在近似精度上仅相差多项式因子。此外，正如 Daskalakis 等人 [DSZ21], 所证明的，该问题也等价于寻找映射 $(x, y) \mapsto (\Pi_{[0,1]^d}(x + \nabla_x f(x, y))\ \Pi_{[0,1]^d}(y - \nabla_y f(x, y)))$ 的（近似）不动点，这也说明了该名称的由来。

我们将从以下 PPAD 完全问题进行归约。

问题 2. (End-of-the-Line) 给定两个电路 $S, P : \{0,1\}^N \to \{0,1\}^N$（分别称为后继电路和前驱电路），满足 $S(0) \ne P(0) = 0$，求一个节点 $v \in \{0,1\}^N$，使得 $P(S(v)) \ne v$ 或 $S(P(v)) \ne v \ne 0$。

End-of-the-Line 是 PPAD 中的典型问题，因为 PPAD 是所有可归约到 End-of-the-Line 的问题所构成的类。反过来，如果一个问题至少与 PPAD 中的任何其他问题一样难，即存在从 End-of-the-Line 到该问题的多项式时间归约，那么该问题就是 PPAD 困难。End-of-the-Line 实例的规模是其前驱电路和后继电路表示的规模。

我们将把 End-of-the-Line 归约为两个不同的中间 PPAD 完全问题。第一个是 Pure-Circuit [Del+22]。

问题 3. (Pure-Circuit [Del+22]) Pure-Circuit 的一个实例由一个顶点集 $V = [\kappa]$ 以及两个门集 G_{NOR} 和 G_{PURIFY} 给出。每个门的形式为 (u, v, w)，其中 $u, v, w \in V$ 是互不相同的节点，其含义如下：

- 如果 $(u, v, w) \in \mathsf{G}_{NOR}$，则 u 和 v 是该门的输入，而 w 是其输出。
- 如果 $(u, v, w) \in \mathsf{G}_{PURIFY}$，则 u 是该门的输入，而 v 和 w 是其输出 s.

每个节点恰好是一个门的输出。Pure-Circuit 实例的一个解是赋值 $b : V \to \{0, 1, \perp\}$，它满足所有门，也就是说，对于每个门 $(u, v, w) \in \mathsf{G}$，有：

- 若 $(u, v, w) \in \mathsf{G}_{NOR}$，则 b 满足

$$b(u) = b(v) = 0 \implies b(w) = 1$$
$$(b(u) = 1) \text{ or } (b(v) = 1) \implies b(w) = 0$$

- 若 $(u,v,w) \in G_{PURIFY}$，则 b 满足

$$\{b(v), b(w)\} \cap \{0,1\} \neq \varnothing$$
$$b(u) \in \{0,1\} \implies b(v) = b(w) = b(u)$$

定理 3.2 ([Del+22])。 Pure-Circuit是 PPAD完全的。

我们将在归约中使用的第二个问题是 Bernasconi 等人 [Ber+24] 提出的 LinVI 问题。

问题 4 (LinVI [Ber+24])。 给定一个矩阵 $D \in [-1,1]^{m\times m}$、一个向量 $c \in [-1,1]^m$ 和一个近似参数 $\epsilon > 0$，寻找一个点 $z \in [0,1]^m$ 使得:

$$\langle Dz + c, z' - z\rangle \geq -\epsilon \quad \forall z' \in [0,1]^m.$$

在我们的证明中，使用 LinVI 更强的逐分量版本会更加方便，以下定理证明了该版本是困难的。

定理 3.3 ([Ber+24，定理 4。4])。 存在一个常数 ρ，使得寻找一个 $z \in [0,1]^m$ 满足下式是 PPAD-完全的

$$(Dz + c)_j \cdot (z'_j - z_j) \geq -\rho \quad \forall j \in [m], z'_j \in [0,1].$$

4 构造

我们从一个 End-of-the-Line 实例出发，然后将其归约为一个 LinVI 实例和一个 Pure-Circuit 实例。我们指出，没有结构性理由必须从 End-of-the-Line 归约；同样，我们也可以直接从 LinVI 或 Pure-Circuit 归约（例如，在后一种情况下，从一个 Pure-Circuit 实例构造一个 LinVI 实例）。不过，为了清晰起见，我们决定从典型的 PPAD 完全问题 End-of-the-Line 进行归约。

回顾一下，我们令 κ 表示 Pure-Circuit 实例的顶点数，m 表示 LinVI 实例的变量数，ρ 表示其近似精度。接着，我们定义以下参数:

$$n = \frac{2^{64} m^{14} \kappa^2}{\rho^8}, \quad \epsilon = \frac{\rho^{18}}{2^{140} m^{28} \kappa^4} \quad \text{and} \quad \delta = \frac{\rho^2}{2^{10} m^2}.$$

基于 LinVI 和 Pure-Circuit 实例，我们按如下方式构造一个 GDA 实例。对每个 $i \in [n]$，令$M_i = \delta(-n/2 + i)$。我们设定 $d = mn\kappa$，并将 GDA 实例的变量集定义为 $x^v_{i,j}$，其中对每个 $v \in V$, $i \in [n]$, $j \in [m]$ 均如此。为方便起见，我们用 x^v_i 表示变量向量 $(x^v_{i,j})_{j\in[m]}$，并用 x^v 表示变量向量 $(x^v_{i,j})_{i\in[n],j\in[m]}$。我们定义

$$H_v(x,y) = \sum_{i\in[n]} \langle Dx^v_i + c, y^v_i - x^v_i\rangle,$$

其中，D 和 c 分别是 LinVI 实例中的 $m \times m$ 维矩阵和 m 维向量。

4.1 门的实现

为了定义效用函数，我们将使用以下连续且光滑的函数。我们定义函数 $g : \Re \to [0,1]$ 如下：

- $g(z) = 1$ 若 $z \leq 1/4$
- $g(z) = 0$ 若 $z \geq 1/2$
- $g(z) = 128(z - 1/4)^3 - 48(z - 1/4)^2 + 1$ 如果 $z \in (1/4,\ 1/2)$。

直观地说，这个函数可用于确定或非门的输出。具体而言，给定两个变量 $x,\ y \in [0,1]$，如果至少一个变量为 1，则 $g(x+y)$ 为 0，即假；如果两个变量都为 0，则 $g(x+y)$ 为 1，即真。

接下来，我们定义与净化门相关的函数 $\ell : \Re \to [0,1]$，如下所示：

- $\ell(z) = 0$ 当 $z \leq 5/12$
- $\ell(z) = 1$ 当 $z \geq 7/12$
- $\ell(z) = 144(z - 5/12)^2(2 - 3z)$ 如果 $z \in (5/12, 7/12)$.

给定一个值 $z \in [0,1]$，回想一下，我们将每个值 $z \in (0,\ 1)$ 解释为 $\perp$。然后，考虑函数 $\ell(z + 1/4)$ 和 $\ell(z - 1/4)$，可得：

i) 如果 $z = 1$，则 $\ell(z + 1/4)$ 和 $\ell(z - 1/4)$ 均为 1

ii) 如果 $z = 0$，则 $\ell(z + 1/4)$ 和 $\ell(z - 1/4)$ 均为 0

iii) 如果 $z \in (0,1)$，那么 $\ell(z + 1/4)$ 和 $\ell(z - 1/4)$ 中至少有一个属于 $\{0,1\}$。

此外，我们还需要以下阈值函数 $\lambda : \Re \to [0,1]$，其主要性质是：当输入至少为 $3m + 1$ 时取值为 1，当输入至多为 $3m$ 时取值为 0。这里回顾一下，m 是 LinVI 实例的变量数量。

- $\lambda(z) = 0$ 当 $z \leq 3m$
- $\lambda(z) = 1$ 当 $z \geq 3m + 1$
- $\lambda(z) = -2(z - 3m)^3 + 3(z - 3m)^2$ 如果 $z \in (3m, 3m + 1)$

直观而言，函数 λ 将用于构造 Pure-Circuit 的赋值。给定一个值 z，函数 $\lambda(z)$ 在 $z \geq 3m + 1$ 时赋值为 1，在 $z \leq 3m$ 时赋值为 0，并在 $z \in (3m, 3m + 1)$ 时赋值为 $\perp$。

最后，我们指出，所有函数都是连续的、$O(1)$-利普希茨且光滑的。接下来，我们将使用以下显式界：$\sup_{z \in \Re} |g'(z)| = 6$、$\sup_{z \in \Re} |\ell'(z)| = 9$ 和 $\sup_{m \in \mathbb{N},\ z \in \Re} |\lambda'(z)| = 3/2$。这三个函数如图 2 所示。

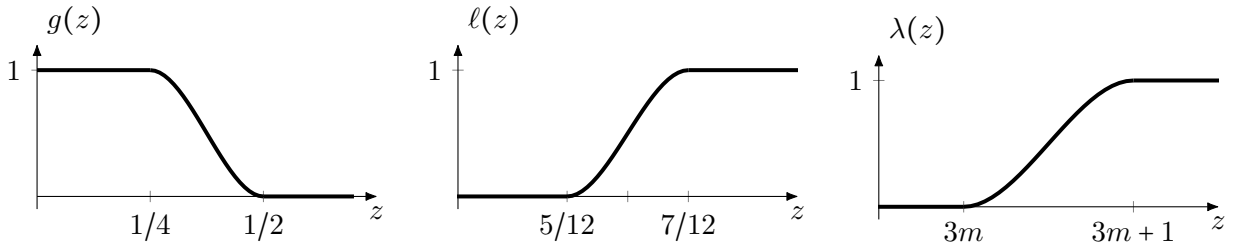


图 2：光滑函数 g、ℓ 和 λ 的示意图。

4.2 效用函数

效用函数由三个不同的分量组成。

- 第一个与 NOR 门有关

$$f_1(x,y) = \sum_{(u,v,w)\in\mathcal{G}_{NOR}} g\left(\lambda(\|x^u - y^u\|^2) + \lambda(\|x^v - y^v\|^2)\right) \cdot H_w(x,y).$$

- 第二个与 $PURIFY$ 门有关

$$f_2(x,y) = \sum_{(u,v,w)\in\mathcal{G}_{PURIFY}} \left[\ell\Big(\lambda(\|x^u - y^u\|^2) + \frac{1}{4}\Big) \cdot H_v(x,y) + \ell\Big(\lambda(\|x^u - y^u\|^2) - \frac{1}{4}\Big) \cdot H_w(x,y)\right].$$

- 最后一个组成部分是一个（加权）正则化项

$$\varphi(x,y) = \sum_{v\in V} \sum_{i\in[n]} M_i \|x_i^v - y_i^v\|^2.$$

请注意，权重 M_i 可以为正，也可以为负。

最后，GDA 的效用函数是三个分量之和

$$f(x,y) = f_1(x,y) + f_2(x,y) + \varphi(x,y). \tag{4}$$

显然，f 是光滑的。此外，GDA 实例的所有参数——具体而言，利普希茨常数 G、范围 B 和光滑性 L——都以多项式形式受 Pure-Circuit 和 LinVI 实例的规模所界定，因此也受原始 End-of-the-Line 实例的规模所界定。最后，近似 ϵ 与 End-of-the-Line 实例的规模成反多项式关系。

该构造背后的直觉如下。给定 GDA 的一个解 (x,y)，我们构造出 LinVI 和 Pure-Circuit 的多个可能解。首先，我们检查 x_i^v、$v \in V$、$i \in [n]$ 中是否有任何一个是 LinVI 的解。如果它们都不是 LinVI 的解，那么我们构造 Pure-Circuit 的一个赋值 $b(\cdot)$，如下所示：

$$\lambda(\|x^v - y^v\|^2) = 1 \implies b(v) = 1 \tag{5a}$$
$$\lambda(\|x^v - y^v\|^2) = 0 \implies b(v) = 0 \tag{5b}$$
$$\lambda(\|x^v - y^v\|^2) \in (0,1) \implies b(v) = \perp \tag{5c}$$

我们的构造无法保证该赋值对于 Pure-Circuit 实例是可行的；不过，如果 $\{x_i^v\}_{v\in V, i\in[n]}$ 都不是 LinVI 的解，那么该赋值就是可行的。

5 梯度的分解

本节将计算公式（4）中定义的效用函数 $f(x,\ y)$ 的偏导数。首先需要定义两个重要的量： $\Delta_q(x,\ y)$ 和 $s_q(x,\ y)$。它们有助于分解电路结构对导数的影响。

对于给定的顶点 $q \in V$，我们定义一个量 $\Delta_q(x,\ y)$，它与由以 q 为输入的门所产生的导数分量有关。形式化地，对于任意给定的点 $(x,\ y) \in [0,1]^{2d}$，令

$$\begin{aligned}\Delta_q(x,y) = &\sum_{(q,v,w)\in\mathcal{G}_{NOR}} g'(\lambda(\|x^q-y^q\|^2)+\lambda(\|x^v-y^v\|^2))\lambda'(\|x^q-y^q\|^2)H_w(x,y)\\ &+\sum_{(u,q,w)\in\mathcal{G}_{NOR}} g'(\lambda(\|x^u-y^u\|^2)+\lambda(\|x^q-y^q\|^2))\lambda'(\|x^q-y^q\|^2)H_w(x,y)\\ &+\sum_{(q,v,w)\in\mathcal{G}_{PURIFY}} \ell'(\lambda(\|x^q-y^q\|^2)+1/4)\lambda'(\|x^q-y^q\|^2)H_v(x,y)\\ &+\sum_{(q,v,w)\in\mathcal{G}_{PURIFY}} \ell'(\lambda(\|x^q-y^q\|^2)-1/4)\lambda'(\|x^q-y^q\|^2)H_w(x,y).\end{aligned}$$

此外，给定一个顶点 q，我们定义一个量 $s_q(x,\ y)$。s_q 表示 Pure-Circuit 的一个解中，输出为 q 的门所需的值。由于每个节点恰好是一个门的输出，对于任意给定的 q，下面的求和中只有一项非零：

$$\begin{aligned}s_q(x,y) = &\sum_{(u,v,q)\in\mathcal{G}_{NOR}} g(\lambda(\|x^u-y^u\|^2)+\lambda(\|x^v-y^v\|^2))\\ &+\sum_{(u,q,w)\in\mathcal{G}_{PURIFY}} \ell(\lambda(\|x^u-y^u\|^2)+1/4)\\ &+\sum_{(u,v,q)\in\mathcal{G}_{PURIFY}} \ell(\lambda(\|x^u-y^u\|^2)-1/4).\end{aligned}$$

例如，如果 q 是一个门 $(u,\ v,\ q) \in \mathsf{G}_{NOR}$ 的输出，那么根据公式（5）中的解释，s_q 就是后续 NOR 门的输出。

Δ_q 和 s_q 的定义使我们能够简洁地表示完整效用函数 $f(x,\ y)$ 的梯度。以下引理的证明推迟至附录 A。

引理 5.1。 f （根据公式 (4) 定义）的偏导数为

$$\frac{\partial f(x,y)}{\partial x_{i,j}^q} = s_q(x,y)\left[(D^\top(y_i^q-x_i^q))_j-(Dx_i^q+c)_j\right]+2(M_i+\Delta_q(x,y))(x_{i,j}^q-y_{i,j}^q), \tag{6a}$$

$$\frac{\partial f(x,y)}{\partial y_{i,j}^q} = s_q(x,y)(Dx_i^q+c)_j-2(M_i+\Delta_q(x,y))(x_{i,j}^q-y_{i,j}^q). \tag{6b}$$

直观地说，$s_q(x,y) = 1$ “激活”了梯度中的 LinVI 部分，打破了 x 与 y 的伪梯度之间的对称性。这将迫使 x_i^q 至少对一个副本 i 成为 LinVI 的近似解，或者迫使距离 $\|x^q-y^q\|^2$ 变大，从而满足 Pure-Circuit 所要求的一致性。

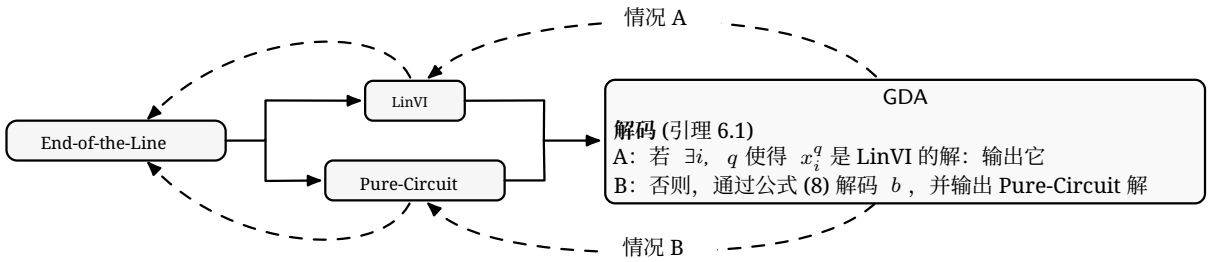


图 3：给定一个 End-of-the-Line 实例 I_{EOTL}，我们构造一个 Pure-Circuit 实例 I_{PC} 和一个 LinVI 实例 I_{VI}，并将它们组合成一个 GDA 实例 I_{GDA}。对于 GDA 的一个解 (x, y)，我们检查其中是否有任何一个 x_i^q 是 LinVI 的解；若没有，则根据方程 (8) 构造一个赋值 $b : V \to \{0, 1, \perp\}$，赋给 Pure-Circuit。引理 6.1 保证，x_i^q 中要么有一个是 LinVI 实例 I_{VI} 的解（情况 A），要么 b 是 Pure-Circuit 实例 I_{PC} 的一个有效赋值（情况 B）。无论哪种情况，我们都可以据此构造原始 End-of-the-Line 实例 I_{EOTL} 的一个解。

项 $\Delta_q(x, y)$ 是 $f(x, y)$ 的梯度中一个不希望出现的副产物。事实上，它依赖于输入包含 q 的那些门。理想情况下，这类门不应影响变量 x^q和 y^q 的取值。加权正则项 φ 包含项 M_i，其目的是确保至少对某些索引 i, $M_i + \Delta_q \simeq 0$，使 Δ_q 对梯度的影响可以忽略不计。

备注 5.2。效用函数 $f(x, y)$ 可以表示为 $f(x, y) = \sum_{q \in V} s_q(x, y) H_q(x, y) + \varphi(x, y)$，正如第 2 节所预期的那样。请注意，$s_q$ 的含义是明确的，因为每个节点恰好是一个门的输出。

6 GDA 的 PPAD 困难性

本节的目标是说明，对于给定的任意 GDA 实例的解，我们都能恢复出两个 PPAD 完全问题之一的解：Pure-Circuit 或 LinVI。这意味着 GDA 的 PPAD 困难性，因为任一问题的解都可以映射回原始 End-of-the-Line 实例的一个解。

主要的技术难点在于证明一致性：说明如果对于 $v \in V, i \in [n]$，不存在能解决 LinVI 的解 x_i^v，那么所有值 s_q 都与值 $\lambda(\|x^q - y^q\|^2)$ 一致。形式化地，我们证明以下二分结论。图 3 概述了这一论证。

引理 6.1 (二分引理)。考虑一个 GDA 实例的解 $(x, y) \in [0, 1]^{2d}$。以下两个条件至少有一个成立：

- 存在一个 $q \in V$和 $i \in [n]$，使得 x_i^q 是LinVI的一个解
- 对于所有顶点 $q \in V$，都有

$$s_q(x, y) = 1 \implies \lambda(\|x^q - y^q\|^2) = 1$$
$$s_q(x, y) = 0 \implies \lambda(\|x^q - y^q\|^2) = 0.$$

前一个引理（其证明推迟到第 7 节），结合$s_q(x, y)$的定义，完成了证明。具体而言，我们结合以下结果：函数 g 和 ℓ

正确实现了NOR和$PURIFY$门、 $s_q(x,\ y)$的性质，以及公式（5）中对 GDA 解的解释。

定理 6.2。 GDA 即使在PPAD-困难，且$1/\epsilon$, L, G和 B 关于 d为多项式时也是如此。

证明。 考虑 GDA 的任意一个解 (x, y)。若存在一个 $q \in V$和 $i \in [n]$，使得 x_i^q 是 LinVI 的一个解，那么我们就能恢复一个 LinVI 的解，从而完成证明。

因此，接下来我们假设不存在这样的 x_i^q，并构造一个 Pure-Circuit 的解。根据引理 6.1，对于所有顶点 $q \in V$，有：

$$s_q(x, y) = 1 \implies \lambda(\|x^q - y^q\|^2) = 1 \tag{7a}$$
$$s_q(x, y) = 0 \implies \lambda(\|x^q - y^q\|^2) = 0. \tag{7b}$$

我们按如下方式构造赋值 $b(\cdot)$。

$$b(v) = \begin{cases} \lambda(\|x^v - y^v\|^2) & \text{if } \lambda(\|x^v - y^v\|^2) \in \{0, 1\} \\ \perp & \text{otherwise} \end{cases} \tag{8}$$

接下来，我们证明该赋值满足所有门约束。首先考虑 NOR门。考虑一个门 $(u, v, w) \in \mathsf{G}_{NOR}$。由于 w是该NOR门的输出， $s_w(x, y)$可化简为 $s_w(x, y) = g(\lambda(\|x^u - y^u\|^2) + \lambda(\|x^v - y^v\|^2))$。于是：

- 如果 $b(u) = b(v) = 0$，那么 $s_w(x,\ y) = 1$，因为 $g(0) = 1$。因此，根据方程（7），有 $b(w) = 1$。
- 如果 $b(u)$ 或 $b(v)$ 等于 1，那么 $s_w(x,\ y) = 0$，因为 $g(q) = 0$ 取 $q \geq 1/2$。因此，根据方程（7），有 $b(w) = 0$。
- 否则，任意满足门条件的 $b(w) \in \{0,\ 1,\ \perp\}$。

现在，考虑 纯化 门。考虑一个门 $(u,\ v,\ w) \in \mathsf{G}_{PURIFY}$。与 或非 门类似， $s_v(x,\ y)$ 和 $s_w(x,\ y)$（它们是 纯化 门的输出）的表达式可简化为$s_v(x,\ y) = \ell(\lambda(\|x^u - y^u\|^2) + 1/4)$ 和 $s_w(x,\ y) = \ell(\lambda(\|x^u - y^u\|^2) - 1/4)$。于是：

- 若 $b(u) = 1$，则根据 ℓ 的定义， $s_v(x, y) = 1$且 $s_w(x, y) = 1$。因此，由等式 (7) 可得 $b(v) = b(w) = 1$。
- 若 $b(u) = 0$，则根据 ℓ 的定义， $s_v(x, y) = 0$且 $s_w(x, y) = 0$。因此，由等式 (7) 可得 $b(v) = b(w) = 0$。
- 若 $b(u) = \perp$，则 $s_v(x, y) = 1$或 $s_w(x, y) = 0$。事实上，若 $s_v(x, y) \neq 1$，则 $\lambda(\|x^u - y^u\|^2) + 1/4 \leq 7/12$，且 $s_w(x, y) = 0$。另一方面，若 $s_w(x, y) \neq 0$，则 $\lambda(\|x^u - y^u\|^2) - 1/4 \geq 5/12$，因此 $s_v(x, y) = 1$。因此，由等式 (7) 可得 $b(v) = 1$ 或 $b(w) = 0$。

这表明 $b(\cdot)$ 是 Pure-Circuit 的有效赋值。 □

7 引理 6.1 的证明

为证明主要技术引理，我们首先给出关于任意 GDA 解的性质的一些辅助结果。在第 7.1 节中，我们论证，对于 GDA 的任意解 $(x,\ y)$，距离$|x^q_{i,\ j} - y^q_{i,\ j}|$可以由其他参数的函数控制。随后，在第 7.2 节中，我们利用这一点证明，在任意 GDA 解中，$|\Delta_q(x,\ y)|$具有关于 $O(\log\ n)$ 的次线性上界。在第 7.3 节中，我们证明，对于每个节点 q，存在许多索引 $i \in [n]$，使得 $|M_i + \Delta_q| \le 1$（即“猜测良好”的副本）。在第 7.4 节中，我们假设引理 6.1 的第一个条件不成立，并证明第二个条件（一致性）必然成立。最后，第 7.5 节完成证明。

7.1 GDA 解的 $|x^q_{i,j} - y^q_{i,j}|$ 刻画

考虑一个节点 $q \in V$、一个索引 $i \in [n]$和一个 $j \in [m]$。对于 GDA 的任意解，我们将 $x^q_{i,j}$ 与 $y^q_{i,j}$ 之间的距离界定为门输出 $s_q(x,y)$ 和项 $|M_i + \Delta_q|$的函数。

引理 7.1. 考虑任意 $q \in V$、 $i \in [n]$ 和一个 $j \in [m]$。那么，对于 *GDA* 的任意解 $(x,y) \in [0,1]^{2d}$，若 $M_i + \Delta_q(x,y) \neq 0$，则 (x,y) 满足

$$|x^q_{i,j} - y^q_{i,j}| \le \frac{3 s_q(x,y) m}{|M_i + \Delta_q(x,y)|} + \sqrt{\frac{\epsilon}{|M_i + \Delta_q(x,y)|}}.$$

证明. 由于 (x,y) 是 GDA 的一个解，我们得到

$$\frac{\partial f(x,y)}{\partial x^q_{i,j}}(y^q_{i,j} - x^q_{i,j}) \le \epsilon \quad \text{and} \quad -\frac{\partial f(x,y)}{\partial y^q_{i,j}}(x^q_{i,j} - y^q_{i,j}) \le \epsilon,$$

其中，我们分别针对第一位参与者的 $y^q_{i,j}$ 和第二位参与者的 $x^q_{i,j}$ 考虑了最优性条件。

根据方程 (6) 计算偏导数的值，可得

$$\left(s_q(x,y)\left[\left(D^\top(y^q_i - x^q_i)\right)_j - (Dx^q_i + c)_j\right] + 2\left(M_i + \Delta_q(x,y)\right)\left(x^q_{i,j} - y^q_{i,j}\right)\right)\left(y^q_{i,j} - x^q_{i,j}\right) \le \epsilon \tag{9a}$$

$$\left(-s_q(x,y)\left(Dx^q_i + c\right)_j + 2\left(M_i + \Delta_q(x,y)\right)\left(x^q_{i,j} - y^q_{i,j}\right)\right)\left(x^q_{i,j} - y^q_{i,j}\right) \le \epsilon. \tag{9b}$$

现在我们注意到，方程 (9a) 具有 $-sd_1 z - Mz^2 \le \epsilon$ 的形式，其中 $s = s_q(x,\ y) \ge 0$,$d_1 = (D^\top(y^q_i - x^q_i))_j - (Dx^q_i + c)_j$, $M = 2(M_i + \Delta_q(x,\ y))$ 和 $z = (x^q_{i,\ j} - y^q_{i,\ j})$。类似地，方程 (9b) 具有 $sd_2 z + Mz^2 \le \epsilon$ 的形式，其中 $d_2 = -(Dx^q_i + c)_j$。显然 $|d_1|,\ |d_2| \le 3m$，事实上，根据赫尔德不等式，

$$\begin{aligned} |d_1| &\le |(D^\top(y^q_i - x^q_i))_j| + |(Dx^q_i + c)_j| \\ &\le \|D^\top_j\|_1 \|y^q_i - x^q_i\|_\infty + \|D_j\|_1 \|x^q_i\|_\infty + |c_j| \\ &\le 3m, \end{aligned}$$

对于 d_2，也可以类似地得到。

现在我们证明 $|M||z|^2 \le \epsilon + 3ms|z|$。事实上，

- 如果 $M < 0$，那么第一个方程给出 $-sd_1 z + |M|z^2 \le \epsilon$，因此 $|M|z^2 \le \epsilon + zsd_1 \le \epsilon + 3ms|z|$.
- 如果 $M > 0$，那么第二个方程给出 $Mz^2 \le \epsilon - sd_2 z \le \epsilon + 3ms|z|$。

求解 $|z|$ 中的二次方程 $|M||z|^2 \le \epsilon + 3ms|z|$ 可得

$$\begin{aligned}|z| &\le \frac{3ms + \sqrt{9m^2s^2 + 4|M|\epsilon}}{2|M|} \\ &\le \frac{3ms + \sqrt{9m^2s^2} + \sqrt{4|M|\epsilon}}{2|M|} \\ &\le \frac{3ms}{|M|} + \sqrt{\frac{\epsilon}{|M|}},\end{aligned}$$

将 s 和 $|M|$ 的值代回，即可得到所需的界。请注意，二次方程的另一个根，即 $\frac{3ms-\sqrt{9m^2s^2+4|M|\epsilon}}{2|M|}$，是负数，因此无需考虑。

□

7.2 $|\Delta_q(x,\ y)|$ 在 n 中呈次线性增长

在本节中，我们证明，GDA 的任意解都能保证对于每个节点 $q \in V$，非期望分量 $|\Delta_q(x,\ y)|$ 足够小。具体而言，我们证明，对于足够小的 ϵ，它至多随 n 对数级增长。正如我们将要看到的，这意味着只需多项式级别的大 n，即可保证 $\max_{i\in[n]}|M_i| > |\Delta_q(x,\ y)|$。

我们首先给出 $\|x^q - y^q\|_1$ 的一个上界（引理 7.2），随后利用该上界给出 $|\Delta_q(x,\ y)|$ 的上界（引理 7.3）。

引理 7.2. 对于GDA的任意解$(x, y) \in [0,1]^{2d}$，成立：

$$\|x^q - y^q\|_1 \le 14m^2\frac{1}{\delta}\log(n) + mn\sqrt{\frac{\epsilon}{\delta}} \quad \forall q \in V.$$

证明. 回顾对所有 $i \in [n]$ 均有 $M_i = (-\frac{n}{2} + i)\delta$，并定义 $h_0 = \frac{n}{2} - \frac{\Delta_q(x,y)}{\delta}$。于是可得$|M_i + \Delta_q(x,y)| = \delta|h_0 - i|$。现在注意到，$|M_i + \Delta_q(x,y)| < \delta$当且仅当 $|h_0 - i| < 1$，并且满足 $|h_0 - i| < 1$ 的 i至多有2个。我们还定义

$$S_{\ge\delta} = \sum_{i\in[n]:|M_i+\Delta_q(x,y)|\ge\delta} \frac{1}{|M_i + \Delta_q(x,y)|} = \frac{1}{\delta}\sum_{i\in[n]:|h_0-i|\ge 1}\frac{1}{|h_0 - i|}.$$

最后，我们对所有 $k \in \mathbb{N}$定义一个桶 $B_k = \{i \in [n] : |i - h_0| \in [k, k+1)\}$。于是，对所有 $k \in \mathbb{N}$，我们有 $|B_k| \le 2$，并且对每个 $i \in B_k$，成立

$$\frac{1}{|M_i + \Delta_q(x,y)|} = \frac{1}{\delta|h_0 - i|} \le \frac{1}{\delta k} \quad \forall i \in B_k.$$

显然，至多有 n 个非空桶覆盖 $[n]$。更形式化地说，存在一个 $j \ge 0$，使得所有满足 $|i - h_0| \ge 1$ 的 i都包含在 $\cup_{k\in[n]} B_{k+j}$中。因此，下列不等式成立

$$S_{\ge\delta} \le \frac{1}{\delta}\sum_{k=1}^{n}\sum_{i\in B_{k+j}}\frac{1}{|h_0 - i|} \le \frac{2}{\delta}\sum_{k=1}^{n}\frac{1}{k+j} \le \frac{2}{\delta}\sum_{k=1}^{n}\frac{1}{k}.$$

因此， $S_{\geq\delta} \leq \frac{4}{\delta}\log(n)$，其中我们使用了调和级数的上界$2\log(n)$。现在，我们将 $\|x^q - y^q\|_1$ 界定如下

$$
\begin{aligned}
\|x^q - y^q\|_1 &= \sum_{i\in[n]} \|x_i^q - y_i^q\|_1 \\
&= \sum_{i\in[n]:|M_i+\Delta_q(x,y)|\geq\delta} \|x_i^q - y_i^q\|_1 + \sum_{i\in[n]:|M_i+\Delta_q(x,y)|<\delta} \|x_i^q - y_i^q\|_1 \\
&\leq \sum_{i\in[n]:|M_i+\Delta_q(x,y)|\geq\delta} \|x_i^q - y_i^q\|_1 + 2m && (\|x_i^q - y_i^q\|_1 \leq m) \\
&\leq \sum_{i\in[n]:|M_i+\Delta_q(x,y)|\geq\delta} \sum_{j\in[m]} \left[\frac{3s_q(x,y)m}{|M_i+\Delta_q(x,y)|} + \sqrt{\frac{\epsilon}{|M_i+\Delta_q(x,y)|}}\right] + 2m && \text{(Lemma 7.1)} \\
&\leq 3s_q(x,y)m^2 S_{\geq\delta} + mn\sqrt{\frac{\epsilon}{\delta}} + 2m \\
&\leq 12s_q(x,y)m^2\frac{1}{\delta}\log(n) + mn\sqrt{\frac{\epsilon}{\delta}} + 2m \\
&\leq 14m^2\frac{1}{\delta}\log(n) + mn\sqrt{\frac{\epsilon}{\delta}}
\end{aligned}
$$

完成证明。 □

前一个引理意味着，当 ϵ 足够小时， $|\Delta_q(x,\ y)|$ 随 n 至多呈对数级增长。

引理 7.3。 对于 $(x,y) \in [0,1]^{2d}$ 的任意解（其中 $\in [0,1]^{2d}$ 满足 GDA），均有：

$$|\Delta_q(x,y)| \leq 2^9\frac{m^3\kappa}{\delta}\log(n) + 2^5m^2n\kappa\sqrt{\frac{\epsilon}{\delta}} \quad \forall q \in V.$$

证明。 我们注意到， q 至多是 κ 个门的输入，并且 g、 λ 和 ℓ 的导数均为常数，其绝对值分别不超过 6、3/2 和 9。此外，根据赫尔德不等式，$|Dx_i^w + c,\ y_i^w - x_i^w|$ $\leq \|Dx_i^w + c\|_\infty\|y_i^w - x_i^w\|_1 \leq 2m\|x_i^w - y_i^w\|_1$。这意味着 $|H_w(x,\ y)| \leq 2m\|x^w - y^w\|_1$。因此

$$|\Delta_q(x,y)| \leq 14\kappa\max_{w\in V}|H_w(x,y)| \leq 28m\kappa\left(14m^2\frac{1}{\delta}\log(n) + mn\sqrt{\frac{\epsilon}{\delta}}\right) \leq 2^9\frac{m^3\kappa}{\delta}\log(n) + 2^5m^2n\kappa\sqrt{\frac{\epsilon}{\delta}},$$

其中，在第一个不等式中，我们用 $9\cdot\frac{3}{2} \leq 14$ 对两个导数（g 和 λ 或 ℓ 和 λ）的乘积进行了界定，并利用门的数量至多为 κ；而在第二个不等式中，我们使用了引理 7.2。

□

7.3 $|\Delta_q(x,y) + M_i| \leq 1$ 对于 $\Omega(1/\delta)$ 多个 i

考虑一个节点 $q \in V$。引理 7.3 的结果表明，非期望项 $|\Delta_q(x,\ y)|$ 随 n 的增长至多呈对数级增长。于是，通过将 n 设为实例规模的多项式级别的大，我们可以确保正则化项 M_i 的范围（大致跨越 δn）超过 $|\Delta_q|$。这保证了至少存在 $\Omega(1/\delta)$ 个索引 $i \in [n]$ 使得 $|\Delta_q + M_i| \leq 1$。这些就是“猜测正确”的副本，其中正则化项 M_i 抵消了非期望分量 Δ_q。

引理 7.4。 对于 *GDA* 的任意解 $(x、\ y) \in [0,1]^{2d}$，至少存在 $1/\delta$ 个索引 $i \in [n]$ 使得 $|\Delta_q(x、\ y) + M_i| \leq 1$。

证明。 首先，我们证明，所选参数 $n、\ \delta$ 和 ϵ 意味着 $|\Delta_q(x,y)| \leq n\delta/4$。具体而言，很容易验证，我们对 δ, n, ϵ 的选择意味着 $2^{24}\frac{m^6\kappa^2}{\delta^4} \leq n$ 和 $\epsilon \leq \frac{\delta^3}{2^{16}m^4\kappa^2}$。重新整理第一个式子，可得 $\frac{2^9m^3\kappa}{\delta} \leq \frac{\sqrt{n}\delta}{8}$，因此：

$$\frac{2^9m^3\kappa}{\delta}\sqrt{n} \leq \frac{n\delta}{8}, \tag{A}$$

而第二个式子给出：

$$2^5m^2n\kappa\sqrt{\frac{\epsilon}{\delta}} \leq \frac{2^5m^2n\kappa}{\sqrt{\delta}}\frac{\delta^{3/2}}{2^8m^2\kappa} = \frac{\delta n}{8}. \tag{B}$$

接下来，我们可以利用引理 7.3 对 $|\Delta_q(x,y)|$ 进行上界估计，如下所示：

$$\begin{aligned}
|\Delta_q(x,y)| &\leq 2^9\frac{m^3\kappa}{\delta}\log(n) + 2^5m^2n\kappa\sqrt{\frac{\epsilon}{\delta}} && \text{(from Lemma 7.3)}\\
&\leq 2^9\frac{m^3\kappa}{\delta}\sqrt{n} + 2^5m^2n\kappa\sqrt{\frac{\epsilon}{\delta}}\\
&\leq \frac{n\delta}{4}. && \text{(by } (A) \text{ and (B))}
\end{aligned}$$

现在，定义 $h_0 = \frac{n}{2} - \frac{\Delta_q(x,y)}{\delta}$，则有 $\frac{1}{4}n \leq h_0 \leq \frac{3}{4}n$，并且当 $|i - h_0| \leq \frac{1}{\delta}$ 时，$|M_i + \Delta_q(x,y)| = \delta|h_0 - i| \leq 1$。此外，我们还得到，存在一个 $j \in [n/4, 3n/4]$ 使得 $|j - h_0| \leq 1$。由于按照我们对参数的选择，有 $\frac{1}{\delta} \leq n/4 - 1$，因此我们也得到索引集合

$$\mathcal{I} = \left\{i \in \mathbb{N} : i = j + k, |k| \leq \frac{1}{\delta}, k \in \mathbb{Z}\right\}$$

是 $[n]$ 的子集。

回顾一下，对于每个 $i \in \mathbf{I}$，都有 $|i - h_0| \leq \frac{1}{\delta}$，因此 $|M_i + \Delta_q(x,\ y)| \leq 1$。注意到 $|\mathbf{I}| \geq 1/\delta$，我们可以得出，至少有 $1/\delta$ 个索引能够保证 $|\Delta_q(x,\ y) + M_i| \leq 1$。 □

7.4 一致性：将 s_q 与 $\lambda\left(\|x^q - y^q\|^2\right)$ 联系起来

现在，我们终于可以证明一些关于一致性的结果了。首先，我们证明，如果 $s_q(x,\ y) = 0$，那么 $\|x^q - y^q\|^2 \leq 3m$。注意，该结果与引理 6.1 的第一个条件以及近似 LinVI 解的存在性无关。该结果直接由引理 7.1 得出。事实上，由于对于满足 $|M_i + \Delta_q(x,\ y)| \gg 0$ 的每个 i，$s_q(x,\ y) = 0$，$\|x^q - y^q\|^2$ 很小。

引理 7.5. 考虑 $(x, y) \in [0,1]^{2d}$ 是 GDA 的一个解，以及满足 $s_q(x, y) = 0$ 的节点 q。那么，$\|x^q - y^q\|^2 \leq 3m$。

证明。 我们注意到，对于至多两个索引 i，有 $|M_i + \Delta_q| < \delta$，且对于这些索引 $\|x_i^q - y_i^q\|^2 \leq m$。对于所有其他索引 i 以及任意 $j \in [m]$，由引理 7.1 可得 $|x_{i,j}^q - y_{i,j}^q| \leq \sqrt{|M_i + \Delta_q(x,y)|} \leq \sqrt{\epsilon}$

因为 $s_q(x,y) = 0$。因此，

$$\|x_i^q - y_i^q\|^2 = \sum_{j\in[m]} |x_{i,j}^q - y_{i,j}^q|^2 \le \sum_{j\in[m]} \frac{\epsilon}{\delta} = m\frac{\epsilon}{\delta}.$$

结合满足 $|M_i + \Delta_q| < \delta$ 和 $|M_i + \Delta_q| \ge \delta$ 的索引对应的结果，我们得到

$$\|x^q - y^q\|^2 \le 2m + nm\frac{\epsilon}{\delta} \le 3m,$$

其中我们利用了根据参数的选择，$\epsilon \le \delta/n$ 。 □

为完成证明，我们考虑 $s_q(x,y) = 1$的情况。此处需要使用引理 6.1 的第一个条件。具体而言，我们证明 $\|x^q - y^q\|^2$ 很大（至少为 $3m+1$），或者可以找到变分不等式 LinVI 的一个解。

引理 7.6。考虑 GDA 的任意一个解$(x,y) \in [0,1]^{2d}$，以及满足 $s_q(x,y) = 1$的任意节点 $q \in V$。如果不存在这样的 $i \in [n]$，使得 x_i^q 是变分不等式LinVI的一个解，那么 $\|x^q - y^q\|^2 \ge 3m + 1$.

证明。根据引理 7.4，至少对 $1/\delta$ 个 $i \in [n]$，有 $|M_i + \Delta_q(x,y)| \le 1$。我们开始证明，对于这样的 i，有 $\|x_i^q - y_i^q\|_1 \ge \rho/10$。

反设对于某些这样的 i，有 $\|x_i^q - y_i^q\|_1 < \frac{\rho}{10}$。那么，鉴于 (x,y) 是 GDA 的一个解，且根据假设 $s_q(x,y) = 1$，第一位参与者的最优性条件要求对所有 $j \in [m]$ 和 $z \in [0,1]$:

$$\Big((D^\top(y_i^q - x_i^q))_j - (Dx_i^q + c)_j + 2(M_i + \Delta_q(x,y))(x_{i,j}^q - y_{i,j}^q)\Big)(z - x_{i,j}^q) \le \epsilon,$$

这意味着对所有 $j \in [m]$ 和 $z \in [0,1]$

$$\begin{aligned} -(Dx_i^q + c)_j(z - x_{i,j}^q) &\le \epsilon + (D^\top(x_i^q - y_i^q))_j(z - x_{i,j}^q) + 2(M_i + \Delta_q(x,y))(y_{i,j}^q - x_{i,j}^q)(z - x_{i,j}^q) \\ &\le \epsilon + \frac{\rho}{10} + 2\frac{\rho}{10} \le \rho, \end{aligned}$$

这里用到了 $\epsilon \le \rho/2$。这是一个矛盾，因为这表明 x_i^q 是 LinVI 的一个解。因此，至少存在 $1/\delta$ 个 i，使得 $\|x_i^q - y_i^q\|_1 \ge \rho/10$。因此，有

$$\|x^q - y^q\|^2 \ge \sum_{i\in[n]:|M_i+\Delta_q(x,y)|\le 1} \|x_i^q - y_i^q\|^2 \ge \frac{1}{m} \sum_{i\in[n]:|M_i+\Delta_q(x,y)|\le 1} \|x_i^q - y_i^q\|_1^2 \ge \frac{\rho^2}{100m\delta},$$

其中，在倒数第二个不等式中，我们利用了 $\|\cdot\|_1^2 \le m\|\cdot\|^2$。观察到 $\delta \le \frac{\rho^2}{2^9 m^2}$，因此 $\frac{\rho^2}{100m\delta} \ge 5m$，即可完成证明。 □

7.5 汇总全部结果

引理 6.1 可直接由引理 7.5、引理 7.6 以及函数 λ 的定义推出。特别地，若存在一个 $q \in V$ 和 $i \in [n]$，使得 x_i^q 是 LinVI 的一个解，则引理的第一个条件成立。

否则，以下结论成立:

- 若 $s_q(x,\ y) = 0$，根据引理 7.5，有 $\|x^q - y^q\|^2 \le 3m$ 并且，根据 λ 的定义，我们有$\lambda(\|x^q - y^q\|^2) = 0$。
- 若 $s_q(x,\ y) = 1$，根据引理 7.6，有 $\|x^q - y^q\|^2 \ge 3m + 1$；并且根据 λ 的定义，我们有$\lambda(\|x^q - y^q\|^2) = 1$。

引理 6.1 的证明至此完成。

8 结论与未解决问题

本文证明了，即使所有参数均为多项式有界，带乘积约束的极小极大优化仍然是 PPAD 困难问题。是否存在更强的不可近似性结果，仍是一个未解决问题；例如，除维数外，所有参数均为常数时是否如此。此外，我们的归约基于对 Pure-Circuit 和 LinVI 实例的白盒访问。另一方面，Daskalakis 等人 [DSZ21], 采用了一种仅需黑盒访问布劳威尔实例的归约，因此能够利用 Hirsch 等人[HPV89] 给出的无条件下界，证明梯度查询次数的下界。目前的技术无法排除这样一种算法：其运行时间效率低下，但查询效率很高（例如，Chen 等人 [CLY24]最近为 ℓ_∞ 压缩映射提出了一种查询高效算法）。若能改进我们的归约，使其仅利用对某个已知查询下界的 PPAD 完全问题的黑盒访问，将会很有意义。最后，在常数维数下（例如，$d = 2$）寻找一个解的复杂度，仍是一个引人关注的未解决问题。我们目前的归约依赖于高维构造，无法直接推广到低维情形。

附录

A 省略的证明

引理 5.1。 f 的偏导数（按方程(4)定义）为

$$\frac{\partial f(x,y)}{\partial x_{i,j}^q} = s_q(x,y)\left[(D^\top(y_i^q - x_i^q))_j - (Dx_i^q + c)_j\right] + 2(M_i + \Delta_q(x,y))(x_{i,j}^q - y_{i,j}^q), \tag{6a}$$

$$\frac{\partial f(x,y)}{\partial y_{i,j}^q} = s_q(x,y)(Dx_i^q + c)_j - 2(M_i + \Delta_q(x,y))(x_{i,j}^q - y_{i,j}^q). \tag{6b}$$

证明。 我们回顾一下，对于所有 $v \in V$， $H_v(x,y) = \sum_{i\in[n]} Dx_i^v + c, y_i^v - x_i^v$。因此，

$$\frac{\partial H_v(x,y)}{\partial x_{i,j}^v} = (D^\top(y_i^v - x_i^v))_j - (Dx_i^v + c)_j \quad \text{and} \quad \frac{\partial H_v(x,y)}{\partial y_{i,j}^v} = (Dx_i^v + c)_j.$$

我们将使用（复合）链式法则。形式上，给定三个函数 $t, r, u : \mathfrak{R} \to \mathfrak{R}$，有：

$$[t(r(u(x)))]' = t'(r(u(x)))r'(u(x))u'(x).$$

Player x**:** 首先，我们来计算 f_1 关于 $x_{i,j}^q$ 的偏导数：

$$\begin{aligned}
\frac{\partial f_1(x,y)}{\partial x_{i,j}^q} &= 2(x_{i,j}^q - y_{i,j}^q) \sum_{(q,v,w)\in\mathcal{G}_{NOR}} g'(\lambda(\|x^q - y^q\|^2) + \lambda(\|x^v - y^v\|^2))\lambda'(\|x^q - y^q\|^2)H_w(x,y) \\
&\quad + 2(x_{i,j}^q - y_{i,j}^q) \sum_{(u,q,w)\in\mathcal{G}_{NOR}} g'(\lambda(\|x^u - y^u\|^2) + \lambda(\|x^q - y^q\|^2))\lambda'(\|x^q - y^q\|^2)H_w(x,y) \\
&\quad + \left[(D^\top(y_i^q - x_i^q))_j - (Dx_i^q + c)_j\right] \sum_{(u,v,q)\in\mathcal{G}_{NOR}} g(\lambda(\|x^u - y^u\|^2) + \lambda(\|x^v - y^v\|^2)).
\end{aligned}$$

然后，我们计算 f_2 关于 $x_{i,j}^q$ 的偏导数：

$$\begin{aligned}
\frac{\partial f_2(x,y)}{\partial x_{i,j}^q} &= 2(x_{i,j}^q - y_{i,j}^q) \sum_{(q,v,w)\in\mathcal{G}_{PURIFY}} \ell'(\lambda(\|x^q - y^q\|^2) + 1/4)\lambda'(\|x^q - y^q\|^2)H_v(x,y) \\
&\quad + 2(x_{i,j}^q - y_{i,j}^q) \sum_{(q,v,w)\in\mathcal{G}_{PURIFY}} \ell'(\lambda(\|x^q - y^q\|^2) - 1/4)\lambda'(\|x^q - y^q\|^2)H_w(x,y) \\
&\quad + \left[(D^\top(y_i^q - x_i^q))_j - (Dx_i^q + c)_j\right] \sum_{(u,q,w)\in\mathcal{G}_{PURIFY}} \ell(\lambda(\|x^u - y^u\|^2) + 1/4) \\
&\quad + \left[(D^\top(y_i^q - x_i^q))_j - (Dx_i^q + c)_j\right] \sum_{(u,v,q)\in\mathcal{G}_{PURIFY}} \ell(\lambda(\|x^u - y^u\|^2) - 1/4),
\end{aligned}$$

最后， φ 关于 $x_{i,j}^q$ 的偏导数：

$$\frac{\partial \varphi(x,y)}{\partial x_{i,j}^q} = 2M_i(x_{i,j}^q - y_{i,j}^q).$$

玩家 y： 类似地， f_1 关于$y_{i,j}^q$ 的偏导数为

$$\begin{aligned}\frac{\partial f_1(x,y)}{\partial y_{i,j}^q} &= 2(y_{i,j}^q - x_{i,j}^q) \sum_{(q,v,w)\in\mathcal{G}_{NOR}} g'(\lambda(\|x^q - y^q\|^2) + \lambda(\|x^v - y^v\|^2))\lambda'(\|x^q - y^q\|^2)H_w(x,y) \\ &+ 2(y_{i,j}^q - x_{i,j}^q) \sum_{(u,q,w)\in\mathcal{G}_{NOR}} g'(\lambda(\|x^u - y^u\|^2) + \lambda(\|x^q - y^q\|^2))\lambda'(\|x^q - y^q\|^2)H_w(x,y) \\ &+ (Dx_i^q + c)_j \sum_{(u,v,q)\in\mathcal{G}_{NOR}} g(\lambda(\|x^u - y^u\|^2) + \lambda(\|x^v - y^v\|^2)).\end{aligned}$$

f_2 关于 $y_{i,j}^q$ 的偏导数为

$$\begin{aligned}\frac{\partial f_2(x,y)}{\partial y_{i,j}^q} &= 2(y_{i,j}^q - x_{i,j}^q) \sum_{(q,v,w)\in\mathcal{G}_{PURIFY}} \ell'(\lambda(\|x^q - y^q\|^2) + 1/4)\lambda'(\|x^q - y^q\|^2)H_v(x,y) \\ &+ 2(y_{i,j}^q - x_{i,j}^q) \sum_{(q,v,w)\in\mathcal{G}_{PURIFY}} \ell'(\lambda(\|x^q - y^q\|^2) - 1/4)\lambda'(\|x^q - y^q\|^2)H_w(x,y) \\ &+ (Dx_i^q + c)_j \sum_{(u,q,w)\in\mathcal{G}_{PURIFY}} \ell(\lambda(\|x^u - y^u\|^2) + 1/4) \\ &+ (Dx_i^q + c)_j \sum_{(u,v,q)\in\mathcal{G}_{PURIFY}} \ell(\lambda(\|x^u - y^u\|^2) - 1/4).\end{aligned}$$

最后， ϕ 关于$y_{i,j}^q$ 的偏导数为

$$\frac{\partial \varphi(x,y)}{\partial y_{i,j}^q} = 2M_i(y_{i,j}^q - x_{i,j}^q).$$

合并各项： 定义

$$\begin{aligned}\Delta_q(x,y) = &\sum_{(q,v,w)\in\mathcal{G}_{NOR}} g'(\lambda(\|x^q - y^q\|^2) + \lambda(\|x^v - y^v\|^2))\lambda'(\|x^q - y^q\|^2)H_w(x,y) \\ &+ \sum_{(u,q,w)\in\mathcal{G}_{NOR}} g'(\lambda(\|x^u - y^u\|^2) + \lambda(\|x^q - y^q\|^2))\lambda'(\|x^q - y^q\|^2)H_w(x,y) \\ &+ \sum_{(q,v,w)\in\mathcal{G}_{PURIFY}} \ell'(\lambda(\|x^q - y^q\|^2) + 1/4)\lambda'(\|x^q - y^q\|^2)H_v(x,y) \\ &+ \sum_{(q,v,w)\in\mathcal{G}_{PURIFY}} \ell'(\lambda(\|x^q - y^q\|^2) - 1/4)\lambda'(\|x^q - y^q\|^2)H_w(x,y),\end{aligned}$$

and

$$\begin{aligned}s_q(x,y) = &\sum_{(u,v,q)\in\mathcal{G}_{NOR}} g(\lambda(\|x^u - y^u\|^2) + \lambda(\|x^v - y^v\|^2)) \\ &+ \sum_{(u,q,w)\in\mathcal{G}_{PURIFY}} \ell(\lambda(\|x^u - y^u\|^2) + 1/4) \\ &+ \sum_{(u,v,q)\in\mathcal{G}_{PURIFY}} \ell(\lambda(\|x^u - y^u\|^2) - 1/4),\end{aligned}$$

我们可以将所有适当的项收集到 $\partial_{x_i^q,j} f$ 和 $\partial_{y_i^q,j} f$ 中，从而得到

$$\frac{\partial f(x,y)}{\partial x_{i,j}^q} = 2(x_{i,j}^q - y_{i,j}^q)(M_i + \Delta_q(x,y)) + \left[(D^\top(y_i^q - x_i^q))_j - (Dx_i^q + c)_j\right] s_q(x,y)$$

and

$$\frac{\partial f(x,y)}{\partial y^q_{i,j}} = 2(y^q_{i,j} - x^q_{i,j})(M_i + \Delta_q(x,y)) + (Dx^q_i + c)_j s_q(x,y)$$

如所需。 □

致谢

MB 的工作部分由欧洲联盟资助。但文中表达的观点和意见仅代表作者本人，并不一定反映欧洲联盟或欧洲研究理事会执行局的立场。欧洲联盟和资助机构均不对此承担责任。

本研究获得欧洲研究理事会（ERC）资助（项目 101165466 — PLA-STEER）、下一代欧盟计划框架下 PNRRPE-AI 计划（M4C2，投资 1.3，人工智能专项）资助的 FAIR（未来人工智能研究）项目 PE0000013 资助，以及欧盟地平线项目 ELIAS（欧洲可持续人工智能灯塔，编号 101120237）资助。

感谢 Andrea Celli、Gabriele Farina 以及匿名评审提出的宝贵意见并参与讨论。

参考文献

[ACB17] Martin Arjovsky、Soumith Chintala 和 Léon Bottou。“Wasserstein 生成对抗网络”。载于：国际机器学习会议。PMLR。2017 年，第 214–223 页。[ALW21] Jacob Abernethy、Kevin A Lai 和 Andre Wibisono。“极小极大优化的最后迭代收敛率：哈密顿梯度下降与共识优化的收敛性”。载于：算法学习理论。PMLR。2021 年，第 3–47 页。[Ana+25] Ioannis Anagnostides、Ioannis Panageas、Tuomas Sandholm 和 Jingming Yan。“极小极大优化与团队零和博弈中对称均衡的复杂度”。载于：arXiv 预印本 arXiv:2502.08519（2025 年）。[Bab16] Yakov Babichenko。“近似纳什均衡的查询复杂度”。载于：ACM期刊（JACM）63.4（2016 年），第 1–24 页。[Bal+18] David Balduzzi、Sebastien Racaniere、James Martens、Jakob Foerster、Karl Tuyls 和 Thore Graepel。“n 人可微博弈的机制”。载于：国际机器学习会议。PMLR。2018 年，第 354–363 页。[Ber+24] Martino Bernasconi、Matteo Castiglioni、Andrea Celli 和 Gabriele Farina。《约束在极小极大优化复杂度中的作用》。2024 年。arXiv：2411.03248[cs.GT]。网址：https://arxiv.org/abs/2411.03248。[BR21] Yakov Babichenko 和 Aviad Rubinstein。“确定拥塞博弈中纳什均衡的复杂度”。载于：第 53 届 ACM SIGACT计算理论研讨会论文集。2021 年，第 1426–1437 页。[CDT09] Xi Chen、Xiaotie Deng 和 Shang-Hua Teng。“确定计算双人纳什均衡的复杂度”。载于：ACM期刊（JACM）56.3（2009 年），第 1–57 页。[CLY24] Xi Chen 、 Yuhao Li 和 Mihalis Yannakakis。“ 通过多项式查询计算压缩映射的不动点”。载于：第 56 届 ACM计算理论年会论文集。2024 年，第 1364–1373 页。[DDJ21] Jelena Diakonikolas、Constantinos Daskalakis 和 Michael I Jordan。“结构化非凸非凹极小极大优化的高效方法”。载于：国际人工智能与统计学会议。PMLR。2021 年，第 2746–2754 页。[Del+22] Argyrios Deligkas、John Fearnley、Alexandros Hollender 和 Themistoklis Melissourgos。“Pure-Circuit：PPAD 的强不可近似性”。载于：2022 年 IEEE 第 63 届计算机科学基础年度研讨会（FOCS）。IEEE。2022 年，第 159–170 页。[DGP09] Constantinos Daskalakis、Paul W Goldberg 和 Christos H Papadimitriou。“计算纳什均衡的复杂度”。载于：《ACM通讯》52.2（2009 年），第 89–97 页。[DP11] Constantinos Daskalakis 和 Christos Papadimitriou。“连续局部搜索”。载于：第二十二届 ACM-SIAM离散算法研讨会论文集。SIAM。2011 年，第 790–804 页。[DSZ21] Constantinos Daskalakis、Stratis Skoulakis 和 Manolis Zampetakis。“约束极小极大优化的复杂度”。载于：第 53 届 ACM SIGACT计算理论研讨会论文集。2021 年，第 1466–1478 页。

[Fea+21] John Fearnley、Paul W Goldberg、Alexandros Hollender和Rahul Savani。“梯度下降的复杂度：CLS=PPAD∩PLS”。载于：第53届年度ACM SIGACT计算理论研讨会论文集。2021，第46–59页。[Fea+22] John Fearnley、Paul Goldberg、Alexandros Hollender和Rahul Savani。“梯度下降的复杂度：CLS=PPAD∩PLS”。载于：ACM期刊 70.1（2022），第1–74页。[Goo+14] Ian J Goodfellow、Jean Pouget-Abadie、Mehdi Mirza、Bing Xu、David Warde-Farley、Sherjil Ozair、Aaron Courville和约书亚·本吉奥。“生成对抗网络”。载于：神经信息处理系统进展 27（2014）。[HMC21] Ya-Ping Hsieh、Panayotis Mertikopoulos和Volkan Cevher。“极小极大优化算法的极限：收敛到伪非临界集”。载于：国际机器学习会议。PMLR。2021，第4337–4348页。[HMN24] Alexandros Hollender、Gilbert Maystre和Sai Ganesh Nagarajan。“具有独立对手的两团队多矩阵博弈的复杂度”。载于：arXiv预印本 arXiv:2409.07398（2024）。[HPV89] Michael D Hirsch、Christos H Papadimitriou和Stephen A Vavasis。“寻找布劳威尔不动点的指数下界”。载于：复杂性期刊 5.4（1989），第379–416页。[Jin+21] Chi Jin、Praneeth Netrapalli、Rong Ge、Sham M Kakade和Michael I Jordan。“论机器学习中的非凸优化：梯度、随机性与鞍点”。载于：ACM期刊（JACM）68.2（2021），第1–29页。[Kod+17] Naveen Kodali、Jacob Abernethy、James Hays和Zsolt Kira。“论生成对抗网络的收敛性与稳定性”。载于：arXiv预印本 arXiv:1705.07215（2017）。[KYP24] Fivos Kalogiannis、Jingming Yan和Ioannis Panageas。“在对抗团队马尔可夫博弈中学习均衡：一个非凸隐凹极小极大优化问题”。载于：神经信息处理系统进展 37（2024），第92832–92890页。[LJJ20] Tianyi Lin、Chi Jin和Michael Jordan。“论非凸-凹极小极大问题的梯度下降上升”。载于：国际机器学习会议。PMLR。2020，第6083–6093页。[LJJ25] Tianyi Lin、Chi Jin和Michael I Jordan。“用于非凸极小极大优化的双时间尺度梯度下降上升算法”。载于：机器学习研究期刊 26.11（2025），第1–45页。[Mad+17] Aleksander Madry、Aleksandar Makelov、Ludwig Schmidt、Dimitris Tsipras和Adrian Vladu。“迈向能够抵御对抗攻击的深度学习模型”。载于：arXiv预印本 arXiv:1706.06083（2017）。[Mer+18] Panayotis Mertikopoulos、Bruno Lecouat、Houssam Zenati、Chuan-Sheng Foo、Vijay Chandrasekhar和Georgios Piliouras。“鞍点问题中的乐观镜像下降：再多走一（梯度）步”。载于：arXiv预印本 arXiv:1807.02629（2018）。[MP91] Nimrod Megiddo和Christos H Papadimitriou。“论全函数、存在性定理与计算复杂性”。载于：理论计算机科学 81.2（1991），第317–324页。

[MSJ25] Eric Mazumdar、S Shankar Sastry和Michael I Jordan。“论在零和博弈中寻找局部纳什均衡（且仅寻找局部纳什均衡）”。载于：*ACM/IMS*数据科学期刊 2.2（2025），第1–26页。[Mun+24] Rémi Munos、Michal Valko、Daniele Calandriello、Mohammad Gheshlaghi Azar、Mark Rowland、Zhaohan Daniel Guo、Yunhao Tang、Matthieu Geist、Thomas Mesnard、Côme Fiegel等。“基于人类反馈的纳什学习”。载于：第四十一届国际机器学习会议。2024。[OLR21] Dmitrii M Ostrovskii、Andrew Lowy和Meisam Razaviyayn。“在非凸-凹光滑极小极大问题中高效搜索一阶纳什均衡”。载于：*SIAM*优化期刊 31.4（2021），第2508–2538页。[Pap94] ChristosH Papadimitriou。“论奇偶性论证的复杂度及其他低效的存在性证明”。载于：计算机与系统科学期刊 48.3（1994），第498–532页。[Rub15] Aviad Rubinstein。“纳什均衡的不可近似性”。载于：第*47*届*ACM*计算理论年度研讨会论文集。2015，第409–418页。[Rub17] Aviad Rubinstein。“确定计算近似双人纳什均衡的复杂度”。载于：*ACM SIGecom*交流 15.2（2017），第45–49页。[SND18] Aman Sinha、Hongseok Namkoong和John Duchi。“以原则性对抗训练认证若干分布鲁棒性”。载于：国际学习表征会议。2018。[The+19] Kiran K Thekumparampil、Prateek Jain、Praneeth Netrapalli和Sewoong Oh。“光滑极小极大优化的高效算法”。载于：神经信息处理系统进展 32（2019）。[VGFP19] Emmanouil-Vasileios Vlatakis-Gkaragkounis、Lampros Flokas和Georgios Piliouras。“非凸非凹零和博弈中梯度下降-上升的庞加莱回归、循环与伪均衡”。载于：神经信息处理系统进展 32（2019）。[WZB19] Yuanhao Wang、Guodong Zhang和Jimmy Ba。“论局部求解极小极大优化：一种沿脊方法”。载于：国际学习表征会议。2019。[XSC21] Yue Xing、Qifan Song和Guang Cheng。“论对抗训练的算法稳定性”。载于：神经信息处理系统进展 34（2021），第26523–26535页。