arXiv:2601.00611v1 [cs.LG] 2 Jan 2026

# 非单调 $\gamma$-弱 DR 次模最大化的更强近似保证

Hareshkumar Jadav[1], Ranveer Singh[1],Vaneet Aggarwal[2*1]印度 IIT Indore, [2]美国普渡大学

摘要

在约束条件下最大化子模目标函数，是机器学习和优化中的一个基本问题。我们研究在向下封闭凸体上最大化非负、非单调的 $\gamma$-弱 DR 次模函数。我们的主要结果是一种近似算法，其近似保证随 $\gamma$ 平滑变化；特别地，当 $\gamma = 1$（即 DR-次模情形）时，我们的界达到 0.401 的近似因子；而当 $\gamma < 1$ 时，近似保证会平稳退化，并且优于此前针对相同约束下 $\gamma$-弱 DR 次模最大化所报告的界。我们的方法将 Frank–Wolfe 引导的连续贪心框架与一种考虑 $\gamma$ 的双重贪心步骤相结合，从而得到一种简单而有效的处理非单调性的方法。这为在向下闭凸体上进行非单调 $\gamma$-弱 DR 次模最大化提供了最先进的保证。

关键词：组合优化、弱DR子模函数、近似算法

## 1 引言

在各种约束下的子模最大化是优化与理论计算机科学中的核心问题。奠基性工作建立了现代工具箱的大部分内容，并催生了一系列丰富的研究[10, 14, 20, 23, 24]。非正式地说，若随着所选集合增大，边际增益减小（一种离散的“收益递减”性质），则集合函数是子模的。该问题之所以仍然重要，一个关键原因在于其适用范围广泛：许多经典组合任务都可以表述为最大化一个子模目标函数，包括最大割、指派问题、设施选址以及最大二分 [17, 19, 7, 12, 11, 1]。

在连续情形下，收益递减（DR）子模模型支持行列式点过程的最大后验（MAP）推断 [2, 4]。非正式地说，DR子模函数是子模性的连续类比，其中沿每个坐标的边际增益随当前点增大而减小（ $\mathbb{R}^n$中的“收益递减”性质）。我们在预备知识中给出形式定义。该问题也出现在在线分配与学习中 [9, 32]。

连续 DR-次模最大化的经典结果基于具有可证明保证的无投影一阶方法。奠基性工作为连续 DR-次模最大化发展了几何理论和算法 [2], ，也提出了连续非单调 DR-次模目标函数的最优算法 [25]。这一范式还扩展到在线和随机信息模型 [9, 33], ，近期进展则通过基于 DR 的分析改进了界限和约束处理 [5]。在这一研究脉络中，弱 DR 框架通过将收益递减放宽为一个因子 $\gamma \in (0, 1]$：粗略来说，边际增益仍在递减，但允许一定程度的受控乘性松弛 $\gamma$），拓展了建模范围。由此得以构建统一算法并提供相应保证 [16, 29, 28, 27, 26]。这些进展促使我们关注向下闭凸体上的连续 DR 和弱 DR 目标函数。

*电子邮箱：hareshjadavcse@gmail.com，ranveer@iiti.ac.in，vaneet@purdue.edu。本研究的部分内容将在2026年5月举行的国际自主智能体与多智能体系统会议（AAMAS）上发表。

已知有多种一阶方法，可用于向下闭凸体上的连续 DR 和弱 DR 最大化。 投影梯度方法是最基础的方案：沿梯度方向迈出一步，然后投影回可行域 $P$ [16,33]。Frank–Wolfe（无投影）方法也称为条件梯度方法，它们通过求解线性子问题 $\max_{y\in P}\langle y,\ \nabla F(x)\rangle$ 来避免投影，并朝该方向移动；它们利用 DR/弱 DR 限制凹性来证明算法取得进展 [18, 21, 2, 25, 29]。双重贪心类方法将离散夹逼思想应用于连续情形。它们维护两个解，即一个“下界”和一个“上界”，并反复沿相反方向调整其坐标，使两个解逐渐靠拢[6, 25, 29]。这些技术也已扩展到在线、赌博机和随机模型 [9, 33, 29]。

从整体上看，我们的算法包含两个相互作用的阶段。我们将一个考虑 $\gamma$ 的 Frank–Wolfe 引导型 Measured Continuous Greedy 步骤，与一个考虑 $\gamma$的双重贪心步骤结合起来。直观而言，Frank–Wolfe 部分通过沿着凸体内有希望的方向前进来推动全局进展，而双重贪心部分则解决每个坐标上纳入或排除权重所产生的局部冲突。这两个阶段共同生成一个解，其质量由综合性能保证加以认证。在我们的方法中，我们设计了一种考虑 $\gamma$ 的 Frank-Wolfe 引导型 Measured Continuous Greedy（$\gamma$-FWG）算法，采用依赖于 $\gamma$ 的阈值和进展证书；将其与一个考虑 $\gamma$ 的双重贪心算法配对；随后对这些证书的凸组合进行优化，从而得到性能曲线 $\Phi_\gamma$：对于所有 $\gamma\in(0,1)$，其性能都严格优于基准[2, 30]，同时在 $\gamma=1$ [5] 处达到 DR 边界，进而获得依赖于 $\gamma$ 的当前最优保证。

除了建立新的近似保证并改进已知的最佳界限外，我们的方法还引入了以下技术创新：

- 我们为非单调的 $\gamma$-弱 DR 设置设计了一种新颖的、考虑 $\gamma$ 的 Frank–Wolfe 引导的测量连续贪心算法。该方法引入了依赖于 $\gamma$ 的阈值安排和进展证书，在沿 Frank–Wolfe 方向上升与测量更新之间取得平衡，同时保持可行性并确保残差间隙单调衰减。

- 弱 DR 函数具有非对称性。在 DR 情形下，有如下“朴素”不等式

$$F(\mathbf{x}) \ \geq\ \frac{F(\mathbf{x}\vee\mathbf{y})+F(\mathbf{x}\wedge\mathbf{y})}{2}. \tag{1}$$

  相比之下，在 $\gamma$-弱 DR 设置中，引理 3.1 得出

$$F(\mathbf{x}) \ \geq\ \frac{\gamma^2\,F(\mathbf{x}\vee\mathbf{y})+F(\mathbf{x}\wedge\mathbf{y})}{1+\gamma^2}. \tag{2}$$

  此处只有 $F(\mathbf{x}\vee\mathbf{y})$ 项乘以 $\gamma^2$，因此不等式两侧不再受到对称对待。引理 2.1、2.2 以及附录中的若干辅助结果也呈现出类似的单侧 $\gamma$ 依赖性。这种对称性的丧失破坏了经典 DR 分析中使用的标准基于势函数的论证。因此，我们采用基于分类的进展分析，明确追踪单侧边际衰减，并引入考虑 $\gamma$ 的阈值，将 Frank–Wolfe 步骤与测量更新结合起来，从而即使存在这种非对称性，也能证明算法取得进展。

- 与此同时，我们将经典的双贪心势函数改造为 $\gamma$ 加权变体，明确平衡非对称收益与损失，从而在弱 DR 区间内给出紧确的进展保证。

### 1.1 我们的贡献

本文研究向下闭凸体 $P\subseteq[0,1]^n$ 上的非单调 $\gamma$-弱 DR 次模目标函数，其中 $0<\gamma\leq 1$。通俗来说，$\gamma$-弱 DR 意味着边际

增益会随着坐标增大而下降，但下降幅度最多受因子 $\gamma \in$ 限制 $(0,1]$，用于刻画仅表现出部分而非完全收益递减的目标函数）。在这种情形下，规范的近似包络为 $\kappa(\gamma) = \gamma e^{-\gamma}$（当 $\gamma = 1$ 时可还原为 $e^{-1}$）[2, 30]。在 [2]，中，该近似保证可通过时间复杂度为 $\mathcal{O}(1/\varepsilon)$ 的算法实现，其中 $\varepsilon > 0$ 为精度参数；而 [30] 则以运行时间 $\mathcal{O}(1/\varepsilon^3)$ 达到相同的保证。相比之下，我们的算法可在 Poly$(n, \delta^{-1})$ 时间内实现 $\Phi_\gamma$-近似。最近，Buchbinder和Feldman 提出了一种新颖的技术，为（完全）DR子模情形带来了 0.401-近似 [5]；他们的算法运行时间同样为 Poly$(n, \delta^{-1})$。

本文旨在弥合弱 DR 和完全 DR 之间的差距。我们提出了兼顾 $\gamma$ 的算法与分析，(i) 在 $\gamma = 1$ [5] 恢复经典 DR 常数，并且 (ii) 在整个弱 DR 区间 [2, 30] 内严格优于基准 $\kappa(\gamma) = \gamma e^{-\gamma}$。我们的方法将兼顾 $\gamma$ 的 Frank–Wolfe 引导的测量连续贪心算法子程序与兼顾 $\gamma$ 的双重贪心算法相结合，然后对其证书的凸组合进行优化。由此得到的保证 $\Phi_\gamma$ 是由三个可调参数（$\alpha$、$r$、$t_s$）决定的目标值——其中 $\alpha$ 是证书之间的混合权重，而（$r$、$t_s$）控制 FW 引导部分和双重贪心部分的调度与调参——我们针对每个给定的 $\gamma$，选择这些参数以最大化 $\Phi_\gamma$。主要结果的正式表述见定理 4.5。为便于理解我们的保证，图 1 绘出了优化后的曲线 $\Phi_\gamma$，表 1 列出了具有代表性的数值及相应的参数选择（$\alpha$、$r$、$t_s$）。

在这一方向上，我们的主要贡献概述如下——每一项都阐述了我们算法框架的一个独特组成部分及其保证。

- 我们提出了一种考虑 $\gamma$ 的 Frank–Wolfe 引导的测量连续贪心算法，以及一种考虑 $\gamma$ 的双重贪心算法；二者都针对向下闭凸体上的 $\gamma$-弱 DR 目标函数给出了明确的常数因子保证。

- 我们推导出两种证书的参数优化凸组合，得到一条性能曲线 $\Phi_\gamma$，对于所有 $\gamma \in$ (0,1)，该曲线都严格优于先前的基准 $\kappa(\gamma) = \gamma e^{-\gamma}$，并且在 $\gamma = 1$ 处达到经典 DR 常数。

- 我们的证明具有模块化结构，且不依赖曲率假设；它们将 DR 边界作为一个特例重新导出，并平滑地扩展至整个弱 DR 谱。

- 这些方法仅使用一阶信息，并在 $P$ 上进行线性优化（Frank–Wolfe 预言机），因此无需投影，适用于大规模实例。

当完全 DR-次模模型（例如连续预算分配、基于 DPP 的多样性目标，以及概率子模模型的平均场推断）加入过度曝光成本、风险或方差正则化、强度约束等实际惩罚项时，自然会产生非单调的 $\gamma$-弱 DR-次模目标函数。这些增强措施在保留弱 DR 结构的同时引入了非单调性，而这正是我们的算法所针对的情形。

## 1.2 相关工作

对于向下闭凸体上的非单调 $\gamma$-弱 DR 目标函数，统一分析确立了 基准包络 $\kappa(\gamma) = \gamma e^{-\gamma}$ [30]。早期研究通过将连续／“测量”贪心算法与针对单侧弱 DR 梯度调整的无投影（Frank–Wolfe）论证相结合，推导出了相同的包络 [2]。另一项 驻点基准 表明，通过投影上升或镜像上升得到的任一一阶驻点，其目标值均至少达到 OPT 的 $\gamma^2/(1+\gamma^2)$（在 $1/2$ 处达到 $\gamma = 1$），而随机／在线改进则采用非无知代理函数和无偏梯度 [16, 9, 33]。

当 $\gamma = 1$ 时，该问题可简化为在向下闭凸体上进行连续非单调 DR-次模最大化。这一研究方向始于多线性/连续贪心框架 [8] 以及 Measured Continuous Greedy 的 $1/e \approx 0.367$ [13] 保证，后续改进最终达到了当前最优常数 0.401 [5]。困难性结果仍然

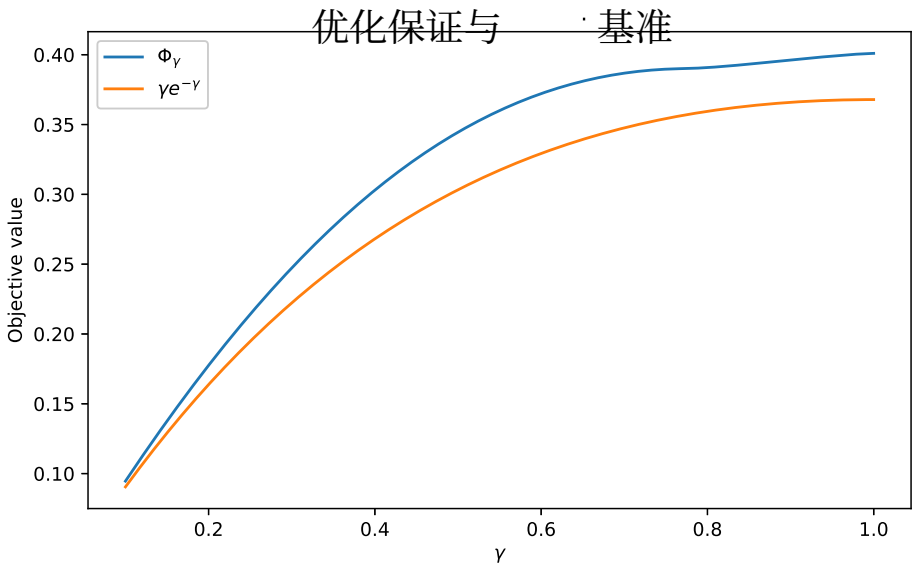



图 1：近似保证与弱 DR 参数的关系。横轴为弱 DR 参数 $\gamma \in (0,1]$，纵轴为近似因子。我们将优化保证 $\Phi_\gamma$（蓝色曲线）与非单调弱 DR 基准 $\kappa(\gamma) = \gamma e^{-\gamma}$（橙色曲线）进行比较。在整个范围 $\gamma \in (0, 1)$ 内，$\Phi_\gamma$ 严格超过 $\kappa(\gamma)$，并且在 $\gamma = 1$（完全 DR）处，我们的曲线达到 0。401，与当前最佳界相符。用于构造 $\alpha$、$r$、$t_s$的选定参数取值 $\Phi_\gamma$列于表 1中。

为常见约束下的非单调目标函数留下了显著差距（保证不优于 $\approx 0.478$ [15]），凸显了在 $\gamma = 1$处设计更优算法的重要性。我们的保证达到了这一 DR 界，同时对每个 $\gamma \in (0,1)$。均严格优于 $\kappa(\gamma)$。

## 2 预备知识与符号

本节介绍贯穿全文所使用的基本符号、定义和假设。我们使用粗体字母（例如 $\mathbf{x}$、$\mathbf{y}$）表示 $\mathbb{R}^n$ 中的向量，并将向量写作 $\mathbf{x} = (x_1, \cdots, x_n)$。全一向量和全零向量分别记为 $\mathbf{1}$ 和 $\mathbf{0}$。我们用 $\mathbf{e}_i$ 表示 $\mathbb{R}^n$ 中的第 $i$ 个标准基向量。令 $N$ 为具有 $|N| = n$个元素的基集。离散超立方体和连续超立方体定义为

$$\{0,1\}^n = \{\mathbf{x} \in \mathbb{R}^n : x_i \in \{0,1\}\ \forall i\},$$

and

$$[0,1]^n = \{\mathbf{x} \in \mathbb{R}^n : x_i \in [0,1]\ \forall i\},$$

分别。对于正整数 $n$，我们记作 $[n] := \{1, 2, \ldots, n\}$。

对于 $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$，我们使用分量序：当且仅当 $x_i \le y_i$ 对所有 $i$ 成立时，$\mathbf{x} \le \mathbf{y}$（并且，当且仅当 $x_i < y_i$ 对所有 $i$成立时，$\mathbf{x} < \mathbf{y}$）。并与交定义为

$$\mathbf{x} \vee \mathbf{y} := (\max\{x_i, y_i\})_{i=1}^n, \quad \mathbf{x} \wedge \mathbf{y} := (\min\{x_i, y_i\})_{i=1}^n.$$

我们还使用逐元素（Hadamard）乘积 $\mathbf{x} \odot \mathbf{y} \in \mathbb{R}^n$，其定义为对每个 $i \in [n]$，$(\mathbf{x} \odot \mathbf{y})_i := x_i\, y_i$；以及标准内积 $\langle \mathbf{x}, \mathbf{y} \rangle := \sum_{i=1}^n x_i y_i$。对于元素属于$[0,1]$, 的向量，逐坐标概率和定义为

$$\mathbf{x} \oplus \mathbf{y} := \mathbf{1} - (\mathbf{1} - \mathbf{x}) \odot (\mathbf{1} - \mathbf{y}). \tag{3}$$

对于向量 $\mathbf{x}^{(1)}$、$\ldots$、$\mathbf{x}^{(m)} \in [0,1]^n$，我们记作

$$\bigoplus_{j=1}^{m} \mathbf{x}^{(j)} = \mathbf{1} - \bigodot_{j=1}^{m} (\mathbf{1} - \mathbf{x}^{(j)}) = \mathbf{1} - \left((\mathbf{1} - \mathbf{x}^{(1)}) \odot \cdots \odot (\mathbf{1} - \mathbf{x}^{(m)})\right). \tag{4}$$

| $\gamma$ | $\Phi_\gamma$ | $\gamma e^{-\gamma}$ | $\alpha$ | $r$ | $t_s$ |
|---|---|---|---|---|---|
| 0.1 | 0.095 | 0.090 | 0.001 | 3.750 | 0.000 |
| 0.2 | 0.178 | 0.164 | 0.000 | 3.083 | 0.000 |
| 0.3 | 0.247 | 0.222 | 0.000 | 2.833 | 0.000 |
| 0.4 | 0.303 | 0.268 | 0.000 | 2.667 | 0.000 |
| 0.5 | 0.345 | 0.303 | 0.000 | 2.583 | 0.000 |
| 0.6 | 0.372 | 0.329 | 0.000 | 2.417 | 0.000 |
| 0.7 | 0.387 | 0.348 | 0.000 | 2.333 | 0.000 |
| 0.8 | 0.391 | 0.359 | 0.054 | 2.250 | 0.075 |
| 0.9 | 0.396 | 0.366 | 0.160 | 2.083 | 0.267 |
| 1.0 | 0.401 | 0.368 | 0.197 | 2.220 | 0.368 |

表 1：我们优化后的保证 $\Phi_\gamma$ 与非单调弱 DR 基准 $\kappa(\gamma) = \gamma e^{-\gamma}$的数值比较。每一行对应一个弱 DR 参数$\gamma \in (0,1]$的取值；第二列和第三列分别报告所得近似因子 $\Phi_\gamma$ 和基准。最后三列列出了用于构造 $\Phi_\gamma$的代表性内部参数 $(\alpha, r, t_s)$： $\alpha$ 是两个证书之间的凸混合权重，而 $r$ 和$t_s$ 分别是感知 $\gamma$ 的 FW 引导 Measured Continuous Greedy 组件和感知$\gamma$的双重贪心组件的调度/调优参数（见算法描述）。所有数值均四舍五入到小数点后三位。

算子 $\odot$ 和 $\oplus$ 的优先级高于向量加法或减法，因此 $\mathbf{x} + \mathbf{y} \odot \mathbf{z}$ 表示 $\mathbf{x} + (\mathbf{y} \odot \mathbf{z})$。将标量函数作用于向量时，按逐元素方式解释；例如，对于 $\mathbf{x} \in [0,1]^n$，向量 $e^{\mathbf{x}}$ 的各个分量为 $\left(e^{\mathbf{x}}\right)_i = e^{x_i}$。

如果对于所有 $\mathbf{x}, \mathbf{y} \in P$和 $\lambda \in [0,1]$，都有 $\lambda\mathbf{x} + (1-\lambda)\mathbf{y} \in P$，则集合 $P \subseteq \mathbb{R}^n$ 是凸的。凸体是具有非空内部的紧凸集。如果对于每个满足 $\mathbf{0} \le \mathbf{x} \le \mathbf{y}$ 的 $\mathbf{x} \in \mathbb{R}^n$，都有 $\mathbf{y} \in P$蕴含 $\mathbf{x} \in P$，则多面体 $P \subseteq [0,1]^n$ 是向下封闭的。如果可以在多项式时间内对 $P$ 进行线性优化，则称 $P$ 是可求解的。集合 $P \subseteq \mathbb{R}^n$ 的（欧几里得）直径为

$$D := \sup\{\, \|\mathbf{x} - \mathbf{y}\|_2 : \mathbf{x}, \mathbf{y} \in P \,\}.$$

非负集合函数 $f\colon \{0,1\}^n \to \mathbb{R}_{\ge 0}$是子模的，当且仅当对所有满足 $\mathbf{x} \le \mathbf{y}$ 的 $\mathbf{x}$、$\mathbf{y}$ $\in \{0,1\}^n$，以及所有 $\mathbf{a} \in \{0,1\}^n$，

$$f(\mathbf{x} \vee \mathbf{a}) - f(\mathbf{x}) \;\ge\; f(\mathbf{y} \vee \mathbf{a}) - f(\mathbf{y}).$$

在连续情形下，非负函数 $F\colon [0,1]^n \to \mathbb{R}_{\ge 0}$ 是收益递减（DR）子模的，当且仅当对所有满足 $\mathbf{x} \le \mathbf{y}$ 的 $\mathbf{x}$、$\mathbf{y} \in [0,1]^n$，任意坐标 $i \in [n]$，以及任意满足 $\mathbf{x} + c\mathbf{e}_i$、$\mathbf{y} + c\mathbf{e}_i \in [0,1]^n$ 的 $c > 0$，

$$F(\mathbf{x} + c\,\mathbf{e}_i) - F(\mathbf{x}) \;\ge\; F(\mathbf{y} + c\,\mathbf{e}_i) - F(\mathbf{y}).$$

若 $F$ 可微，则这等价于 $\nabla F(\mathbf{x}) \ge \nabla F(\mathbf{y})$ 对所有 $\mathbf{x} \le \mathbf{y}$ [4]。

非负函数 $F\colon [0,1]^n \to \mathbb{R}_{\ge 0}$是 $\gamma$-弱 DR 次模的，当且仅当对所有满足 $\mathbf{x} \le \mathbf{y}$ 的 $\mathbf{x}$、$\mathbf{y} \in [0,1]^n$，任意 $i \in [n]$，以及任意满足 $\mathbf{x} + c\mathbf{e}_i$、$\mathbf{y} + c\mathbf{e}_i \in [0,1]^n$ 的 $c > 0$，

$$F(\mathbf{x} + c\,\mathbf{e}_i) - F(\mathbf{x}) \;\ge\; \gamma\big(F(\mathbf{y} + c\,\mathbf{e}_i) - F(\mathbf{y})\big). \tag{5}$$

当 $F$ 可微时，这等价于对所有 $\mathbf{x} \le \mathbf{y}$，$\nabla F(\mathbf{x}) \ge \gamma\, \nabla F(\mathbf{y})$。当且仅当 $F$ 为常数时，该条件对某个 $\gamma > 1$ 成立（且常数 $F$ 对任意 $\gamma$ 都满足该条件）；当 $F$ 逐坐标单调时，该条件恰好对某个 $\gamma \le 0$ 成立。因此，我们关注非平凡范围 $0 < \gamma \le 1$。

可微函数 $F\colon P \to \mathbb{R}$ 是 $L$-光滑的，如果对于所有 $\mathbf{x}, \mathbf{y} \in P$，它满足

$$\|\nabla F(\mathbf{x}) - \nabla F(\mathbf{y})\|_2 \le L\, \|\mathbf{x} - \mathbf{y}\|_2. \tag{6}$$

现在我们讨论 $\gamma$-弱 DR 次模函数的一些性质。在下面的三个引理中，设 $F\colon [0,1]^n \to \mathbb{R}_{\ge 0}$ 为可微且 $\gamma$-弱 DR 次模。

引理 2.1. 对所有 $\mathbf{x}$、$\mathbf{y} \in [0,1]^n$ 和 $\lambda \in [0,1]$，以下结论成立：

1. 若 $\mathbf{x} \leq \mathbf{y}$，则

$$F\big(\lambda\mathbf{x} + (1-\lambda)\mathbf{y}\big) \;\geq\; \frac{\lambda\, F(\mathbf{x}) + \gamma^2(1-\lambda)\, F(\mathbf{y})}{\lambda + \gamma^2(1-\lambda)}. \tag{7}$$

等价地，

$$F\big((1-\lambda)\mathbf{x} + \lambda\mathbf{y}\big) \;\geq\; \frac{(1-\lambda)\, F(\mathbf{x}) + \gamma^2\lambda\, F(\mathbf{y})}{(1-\lambda) + \gamma^2\lambda}. \tag{8}$$

2. 若 $\mathbf{x} + \mathbf{y} \in [0,1]^n$，则

$$F(\mathbf{x} + \lambda\mathbf{y}) - F(\mathbf{x}) \;\geq\; \frac{\gamma^2\lambda}{1 - \lambda + \gamma^2\lambda}\big(F(\mathbf{x}+\mathbf{y}) - F(\mathbf{x})\big). \tag{9}$$

引理 2.2. 对于任意 $\mathbf{x}, \mathbf{y} \in [0,1]^n$，以下不等式成立：

1. 若 $\mathbf{x} + \mathbf{y} \leq \mathbf{1}$，则

$$\langle \nabla F(\mathbf{x}),\, \mathbf{y} \rangle \;\geq\; \gamma\big(F(\mathbf{x}+\mathbf{y}) - F(\mathbf{x})\big). \tag{10}$$

2. 若 $\mathbf{x} - \mathbf{y} \geq \mathbf{0}$，则

$$\langle \nabla F(\mathbf{x}),\, \mathbf{y} \rangle \;\leq\; \frac{1}{\gamma}\big(F(\mathbf{x}) - F(\mathbf{x}-\mathbf{y})\big). \tag{11}$$

引理 2.3。对于任意固定的 $\mathbf{y} \in [0,1]^n$，定义

$$G_{\oplus}(\mathbf{x}) \;:=\; F(\mathbf{x} \oplus \mathbf{y}) \quad \textit{and} \quad G_{\odot}(\mathbf{x}) \;:=\; F(\mathbf{x} \odot \mathbf{y}). \tag{12}$$

则 $G_{\oplus}$ 和 $G_{\odot}$ 均为非负且为 $\gamma$–弱 DR 次模。

这些引理的证明见附录A。这些陈述推广了 [33, 16] 到 $\gamma$–弱 DR 情形，并且对所有 $\gamma \in (0,1]$ 均成立；特别地，当 $\gamma = 1$ 时，它们恰好与经典 DR 陈述一致。

## 3 支持性结果

在本节中，我们将两个标准算法推广到 $\gamma$–弱 DR 情形，并给出相应分析。首先，我们证明在可求解凸体上成立的 $\gamma$-加权 Frank–Wolfe 证书（定理 3.2）。其次，我们构造一种 $\gamma$-感知的 Double–Greedy 算法，并建立一个关于 $\gamma$ 的非平衡下界，该下界可平滑插值（定理 3.3）。

我们将经典 DR 框架推广到 $\gamma$–弱 DR 情形，并获得了能够在弱 DR 与完全 DR 之间连续插值的保证。我们的第一个要素（引理 3.1）表明，在 $\gamma$–弱 DR-次模性下，局部最优性证书 $\langle \mathbf{y} - \mathbf{x}\ \nabla F(\mathbf{x}) \rangle \leq 0$ 可转化为 $\gamma$–加权的值比较，即比较 $F(\mathbf{x})$ 与并/交值 $F(\mathbf{x} \vee \mathbf{y})$ 和 $F(\mathbf{x} \wedge \mathbf{y})$，从而在 $\gamma = 1$ 处恢复经典的 $\frac{1}{2}\big(F(\mathbf{x} \vee \mathbf{y}) + F(\mathbf{x} \wedge \mathbf{y})\big)$ 界（参见 [8]；另见 [2, 3, 9]）。我们的第二个要素（定理 3.2）在可求解凸体 $P$ 上“全局化”了这一比较：一种 Frank–Wolfe 型例程针对每个 $\mathbf{y} \in P$ 给出一致一阶证书，而无需曲率信息；这一做法是近期连续子模求解器中的标准工具（例如 [5]），并且与统一的弱 DR 分析（例如 [29]）一致。将该证书与弱 DR 性质结合，可得到一个随 $\gamma$ 平滑退化、并在 $\gamma = 1$ 处恰好匹配 DR 情形的值界；这一点在引理 3.1 和定理 3.2 中得到了形式化，相关结果的证明见附录 B

引理 3.1。设 $F : [0,1]^n \to \mathbb{R}_{\geq 0}$ be $\gamma$–弱 DR–子模。若 $\mathbf{x}$ 是关于向量 $\mathbf{y}$ 的局部最优解，即，

$$\langle \mathbf{y} - \mathbf{x}, \nabla F(\mathbf{x}) \rangle \leq 0, \tag{13}$$

*then*

$$F(\mathbf{x}) \;\geq\; \frac{\gamma^2\, F(\mathbf{x} \vee \mathbf{y}) + F(\mathbf{x} \wedge \mathbf{y})}{1 + \gamma^2}. \tag{14}$$

定理 3.2。设 $F\colon [0,1]^n \to \mathbb{R}_{\ge 0}$ be 一个非负、 $L$-光滑且 $\gamma$–弱 DR–子模的函数。设 $P \subseteq [0,1]^n$ be 一个直径为 $D$ 的可求解凸体，且 $\delta \in (0,1)$。存在一个多项式时间算法，输出 $\mathbf{x} \in P$，使得对于每个 $\mathbf{y} \in P$,

$$F(\mathbf{x}) \ge \frac{\gamma^2 F(\mathbf{x}\vee\mathbf{y}) + F(\mathbf{x}\wedge\mathbf{y})}{1+\gamma^2} - \frac{\delta\gamma}{1+\gamma^2}\left[\max_{\mathbf{z}\in P} F(\mathbf{z}) + \frac{LD^2}{2}\right]. \tag{15}$$

在连续 DR-次模领域，双重贪心型过程已由若干工作推广到DR-次模目标函数 $[0,1]^n$ （ $\gamma$=1 情形）， [4, 9, 25], 但这些分析仍给出了统一的 1/2 界，且早于非平衡改进。我们将这一研究方向推广到 $\gamma$–弱 DR-次模情形（$0 < \gamma \le 1$），并利用为连续次模最大化引入的弱 DR 结构 [3, 2] 以及近期提出的统一弱 DR 视角[29]。

我们的算法保留了双重贪心结构，同时加入 $\gamma$感知的平滑/阈值处理，使其能够稳健地处理任意给定的 $\gamma \in (0,1]$；所得保证随 $\gamma$ 连续变化，并在 $\gamma = 1$时退化为经典 DR 界。相应的下界保证见定理3.3，详细证明见附录C。

定理 3.3。设 $F\colon [0,1]^n \to \mathbb{R}_{\ge 0}$ 为非负且 $\gamma$–弱 DR 次模函数，其中 $\gamma \in$ 为某个 $0,1]$，（ $\varepsilon \in$ 并固定参数 $\varepsilon \in$ （ 0, 1 ）。存在一个多项式时间算法，输出满足 $\mathbf{x} \in [0,1]^n$ 的解，使得

$$F(\mathbf{x}) \;\ge\; \max_{r\ge 0} \frac{\left(2\gamma^{3/2} - 4\varepsilon\gamma^{9/2}\right) r\, F(\mathbf{o}) \;+\; F(\mathbf{0}) \;+\; r^2 F(\mathbf{1})}{r^2 \;+\; 2\gamma^{3/2} r \;+\; 1}. \tag{16}$$

当 $\gamma = 1$ 且 $r = 1$ 时，这将恢复规范的 1/2 近似保证；而对于许多实例，可以选择 $r \ne 1$，从而获得严格更优的保证，这与集合函数的非平衡界 [22, 31]直接类似。

非平衡双重贪心保证可直接扩展到轴对齐盒。固定上界 $\mathbf{x} \in [0,1]^n$，并考虑在盒域 $[\mathbf{0},\mathbf{x}]$ 上最大化 $F$。定义 $G\colon [0,1]^n \to \mathbb{R}_{\ge 0}$，其中 $G(\mathbf{a}) := F(\mathbf{x}\odot\mathbf{a})$。根据引理 2.3， $G$ 仍为非负且 $\gamma$–弱 DR 次模函数，因此定理 3.3 可应用于 $G$。通过 $\mathbf{y} := \mathbf{x}\odot\mathbf{a}' \le \mathbf{x}$ 将输出转换回来，可得到以下推论。我们在算法 2 中将以下推论称为盒最大化。

推论 3.4（盒最大化）。设 $F\colon [0,1]^n \to \mathbb{R}_{\ge 0}$ 为非负且 $\gamma$–弱 DR 次模，满足某个 $\gamma \in (0,1]$, 令 $\mathbf{x} \in [0,1]^n$，并固定 $\varepsilon \in (0,1)$。存在一个多项式时间算法，输出向量 $\mathbf{y} \in [0,1]^n$，使得 $\mathbf{y} \le \mathbf{x}$，并且对于每个固定的 $\mathbf{o} \in [0,1]^n$,

$$F(\mathbf{y}) \ge \max_{r\ge 0} \frac{\left(2\gamma^{3/2} - 4\varepsilon\,\gamma^{9/2}\right) r\, F(\mathbf{x}\odot\mathbf{o}) + F(\mathbf{0}) + r^2 F(\mathbf{x})}{r^2 \;+\; 2\gamma^{3/2} r \;+\; 1}. \tag{17}$$

证明。固定 $\mathbf{x} \in [0,1]^n$，并定义受限目标函数

$$G(\mathbf{a}) \;:=\; F(\mathbf{x}\odot\mathbf{a}) \qquad \text{for all } \mathbf{a} \in [0,1]^n. \tag{18}$$

根据引理 2.3， $G$ 是非负的，且 $\gamma$–弱 DR 次模。将定理 3.3 应用于 $G$（使用相同的 $\varepsilon \in (0,\ 1)$）可得某个 $\mathbf{a}' \in [0,1]^n$ 满足

$$G(\mathbf{a}') \ge \max_{r\ge 0} \frac{\left(2\gamma^{3/2} - 4\varepsilon\,\gamma^{9/2}\right) r\, G(\mathbf{o}) + G(\mathbf{0}) + r^2\, G(\mathbf{1})}{r^2 \;+\; 2\gamma^{3/2} r \;+\; 1}. \tag{19}$$

由（18），可得

$$G(\mathbf{o}) = F(\mathbf{x}\odot\mathbf{o}), \qquad G(\mathbf{0}) = F(\mathbf{0}), \qquad G(\mathbf{1}) = F(\mathbf{x}).$$

将这些恒等式代入（19）可得

$$G(\mathbf{a}') \ge \max_{r\ge 0} \frac{\left(2\gamma^{3/2} - 4\varepsilon\ \gamma^{9/2}\right) r\, F(\mathbf{x}\odot\mathbf{o}) + F(\mathbf{0}) + r^2\, F(\mathbf{x})}{r^2 \;+\; 2\gamma^{3/2} r \;+\; 1}. \tag{20}$$

定义 $\mathbf{y} := \mathbf{x} \odot \mathbf{a}'$。则 $\mathbf{y} \le \mathbf{x}$ 逐坐标地，并且根据（18），

$$F(\mathbf{y}) = F(\mathbf{x} \odot \mathbf{a}') = G(\mathbf{a}'). \tag{21}$$

结合（20）和（21）可得所述界（17）。 □

# 4 主算法与结果

本节介绍我们的主要结果，以及实现该结果的算法。我们的方法具有递归结构，其核心是在递归的每一层调用一个子程序：$\gamma$–Frank–Wolfe 引导的测度连续贪心算法（$\gamma$-**FWG**）。我们将在第 $\gamma$-**FWG** 节介绍该算法，并证明其保证。基于这一组件，第4.1节介绍完整的递归算法，第4.2 节则证明我们的主要定理。4.3

## 4.1 $\gamma$-FWG 算法

我们开发了一种测量连续贪心方法，由 Frank–Wolfe 方向引导，并通过 $\gamma$进行显式调节，以适用于 $\gamma$–弱 DR 情形。该算法显式地以 $\gamma$为参数，因此适用于任意 $\gamma \in (0, 1]$，且当 $\gamma = 1$ [5] 时退化为经典 DR 情形。为清晰起见，在算法 1 的描述中，我们假设 $\delta^{-1}$ 是一个整数，且 $\delta \le \varepsilon$（这使我们可以设定 $m = \delta^{-1}$）。若这些条件不成立，我们将 $\delta$缩减为 $1/\lceil 1/\min\{\delta, \varepsilon\}\rceil$，而不影响分析。此外，我们定义 $\beta := \frac{\gamma^2 \delta}{1-\delta+\gamma^2\delta}$。由于算法不知道 $F(\mathbf{o})$、$F(\mathbf{z} \odot \mathbf{o})$和 $F(\mathbf{z} \oplus \mathbf{o})$的值，我们采用以下猜测引理；其证明使用标准的猜测论证，完整证明见附录 D。

引理 4.1. 设 $F : [0,1]^n \to \mathbb{R}_{\ge 0}$ 为非负且是 $\gamma$-弱 DR 次模的，其中$0 < \gamma \le 1$为某个常数，并设 $P \subseteq [0,1]^n$ 为向下封闭。存在一个常数大小（仅取决于$\varepsilon$ 和 $\gamma$）的三元组集合 $\mathcal{G} \subseteq \mathbb{R}^3_{\ge 0}$，使得 $\mathcal{G}$ 中包含一个三元组 $(g, g_\odot, g_\oplus)$，满足

$$(1-\varepsilon)\, F(\mathbf{o}) \le g \le F(\mathbf{o}), \tag{22a}$$
$$F(\mathbf{z} \odot \mathbf{o}) - \varepsilon\, g \le g_\odot \le F(\mathbf{z} \odot \mathbf{o}), \tag{22b}$$
$$F(\mathbf{z} \oplus \mathbf{o}) - \varepsilon\, g \le g_\oplus \le F(\mathbf{z} \oplus \mathbf{o}). \tag{22c}$$

因此，通过尝试 $\mathcal{G}$中的所有三元组，我们可以假设算法 1获得了满足这些界的有效替代项$g$、$g_\odot$和 $g_\oplus$。为方便起见，我们先定义阈值函数。对于 $i \in \{0, 1, \ldots, \delta^{-1} - 1\}$，定义

$$v_1(i) := \left[(1-\beta)^i + \tfrac{1-(1-\beta)^i - 2\varepsilon}{\gamma}\right] g - \tfrac{1}{\gamma}\, g_\odot - \tfrac{1-(1-\beta)^i}{\gamma}\, g_\oplus \tag{23}$$

$$v_2(i) := (1-\beta)^i \left[\left(\tfrac{(1-\beta)^{-i_s}}{\gamma} - \left(1 + \tfrac{3}{\gamma}\right)\varepsilon + 1 - \tfrac{1}{\gamma}\right) g - \left(\tfrac{(1-\beta)^{-i_s}}{\gamma} - \tfrac{1}{\gamma} - \beta\,(i - i_s)\right) g_\oplus\right] \tag{24}$$

对于固定的 $\gamma$，参数 $t_s$ 和 $\varepsilon$ 均为常数，因此算法 1 的运行时间为 Poly$(n, \delta^{-1})$。因此，算法 1 对其输出保证以下性能。该定理的证明见附录D。

定理 4.2。 $\gamma$-FWG 的输入包括一个非负、$L$-光滑、$\gamma$–弱 DR 次模函数 $F: [0,1]^n \to \mathbb{R}_{\ge 0}$，一个直径为 $D$ 的元可解向下封闭凸体 $P \subseteq [0,1]^n$、一个向量 $\mathbf{z} \in P$，以及参数 $t_s \in (0, 1)$、$\varepsilon \in (0, 1/2)$ 和 $\delta \in (0, 1)$。给定这些输入，$\gamma$-FWG 会输出一个向量 $\mathbf{y} \in P$和向量 $\mathbf{x}(1) \ldots, \mathbf{x}(m) \in P$，其中某个 $m = O(\delta^{-1} + \varepsilon^{-1})$，使得以下条件至少有一个成立。

1.
$$F(\mathbf{y}) \ge A_\gamma(t_s)\, F(\mathbf{o}) + B_\gamma(t_s)\, F(\mathbf{z} \odot \mathbf{o}) + C_\gamma(t_s)\, F(\mathbf{z} \oplus \mathbf{o}) - \delta L D^2. \tag{25}$$

**Algorithm 1** $\gamma$-FWG$(F, P, \mathbf{z}, \gamma, t_s, \varepsilon, \delta)$

**Input:** nonnegative $L$-smooth $\gamma$-weakly DR-submodular $F : [0,1]^n \to \mathbb{R}_{\geq 0}$; meta-solvable down-closed $P \subseteq [0,1]^n$; $\mathbf{z} \in P$; $\gamma \in (0,1]$; parameters $t_s \in (0,1)$, $\varepsilon \in (0, 1/2)$, $\delta \in (0,1)$.
$i_s \leftarrow \lceil t_s/\delta \rceil$
**for** $i = 0$ **to** $\delta^{-1} - 1$ **do**
    $v(i) := \begin{cases} v_1(i), & \text{if } i \leq i_s, \\ v_2(i), & \text{if } i \geq i_s. \end{cases}$
    $\mathbf{z}(i) \leftarrow \begin{cases} \mathbf{z}, & \text{if } i < i_s, \\ \mathbf{0}, & \text{if } i \geq i_s \end{cases}$
**end for**
$\mathbf{y}(0) \leftarrow \mathbf{0}$
**for** $i = 1$ **to** $\delta^{-1}$ **do**
    $\mathbf{w}(i) \leftarrow \big(\mathbf{1} - \mathbf{y}(i-1) - \mathbf{z}(i-1)\big) \odot \nabla F\big(\mathbf{y}(i-1)\big)$
    $Q(i) \leftarrow \big\{ \mathbf{x} \in P \;\big|\; \langle \mathbf{w}(i), \mathbf{x} \rangle \geq \gamma\big(v(i-1) - F(\mathbf{y}(i-1))\big) \big\}$
    Use Theorem 3.2 to compute an approximate local maximum $\mathbf{x}(i)$ of $Q(i)$ (if $Q(i) = \varnothing$, set $\mathbf{x}(i)$ to an arbitrary vector in $P$)
    $\mathbf{y}(i) \leftarrow \mathbf{y}(i-1) + \delta \big(\mathbf{1} - \mathbf{y}(i-1) - \mathbf{z}(i-1)\big) \odot \mathbf{x}(i)$
**end for**
**return** $\mathbf{y}(\delta^{-1})$ and the sequence $\mathbf{x}(1), \mathbf{x}(2), \ldots, \mathbf{x}(\delta^{-1})$

2. 存在 $i \in [m]$，使得

$$F\big(\mathbf{x}(i) \oplus \mathbf{o}\big) \;\leq\; F(\mathbf{z} \oplus \mathbf{o}) \;-\; \varepsilon\, F(\mathbf{o}), \tag{26}$$

且点 $\mathbf{x}(i)$ 满足 $\gamma$–弱 DR 局部值界

$$F\big(\mathbf{x}(i)\big) \geq \frac{\gamma^2\, F\big(\mathbf{x}(i) \vee \mathbf{o}\big) + F\big(\mathbf{x}(i) \wedge \mathbf{o}\big)}{1+\gamma^2} - \frac{\delta\,\gamma}{1+\gamma^2}\left(\max_{\mathbf{y}' \in Q(i)} F(\mathbf{y}') + \tfrac{1}{2} L D^2\right). \tag{27}$$

即，在 $\gamma$–弱 DR 保证和针对 $Q(i)$ 的 Frank–Wolfe 证书下，$\mathbf{x}(i)$ 是关于 $\mathbf{o}$ 的近似局部最大值。

这里，依赖于 $\gamma$ 的系数为

$$A_\gamma(t_s) := -\frac{e^{\gamma t_s - \gamma}}{1-\gamma} + \frac{e^{-\gamma^2}}{\gamma(1-\gamma)}\Big(e^{\gamma^2 t_s} - (1-\gamma)\Big) \;-\; O(\varepsilon) \tag{28a}$$

$$B_\gamma(t_s) := \frac{e^{-\gamma} - e^{\gamma t_s - \gamma}}{\gamma} \tag{28b}$$

$$\begin{aligned} C_\gamma(t_s) := {} & \frac{e^{\gamma^2 t_s} - 1}{\gamma(1-\gamma)}\Big(e^{-\gamma(1-t_s) - \gamma^2 t_s} - e^{-\gamma^2}\Big) \\ & + \frac{e^{-\gamma(1-t_s)}}{\gamma}\left[\big(e^{-\gamma t_s} - 1\big) + \frac{e^{-\gamma^2 t_s} - e^{-\gamma t_s}}{1-\gamma}\right] \\ & + e^{-\gamma(1-t_s) - \gamma^2 t_s}\left[\frac{\gamma^2}{1-\gamma}(1-t_s)\, e^{\gamma(1-\gamma)(1-t_s)}\right. \\ & \left. + \frac{\gamma}{(1-\gamma)^2}\Big(1 - e^{\gamma(1-\gamma)(1-t_s)}\Big)\right]. \end{aligned} \tag{28c}$$

## 4.2 主算法

本节介绍我们的主算法，并分析用于证明主要结果（定理 4.5）的递归框架。其核心构件是我们新提出的过程 $\gamma$-弱

Frank–Wolfe 引导的测度连续贪心算法 ($\gamma$-FWG) 在第 4.1 节中提出。我们将 $\gamma$-FWG 的一次执行归类为 成功，当且仅当出现第一种结果（即 $F(\mathbf{y})$ 达到“高值”情形）时。否则，该次执行被视为 不成功。明确了这些术语后，我们现在介绍主递归驱动程序，即 算法 2，并用它证明定理 4.5。除定理 4.5 中的参数外，算法 2 还需要两个辅助输入： $\varepsilon \in (0, 1/2)$ 和 $t_s \in (0, 1)$。这两个输入会原样传递给每次对$\gamma$-FWG 的调用。

算法 2 运行 $L = 1 + \left\lceil \frac{1+\gamma}{\varepsilon\gamma} \right\rceil$ 个递归层级，由 $i$ 编号。在第 1 层，它找到一个近似局部最大化器 $\mathbf{z}(0) \in P$，并从该种子出发运行 Box-Maximization 和 $\gamma$-FWG，生成 $\mathbf{z}'$、$\mathbf{y}$，以及一批 $\{\mathbf{x}(1) \ldots, \mathbf{x}(m)\}$。对于后续各层 $i = 2, \ldots, L$，在第 $i-1$ 层生成的每个候选解 $\mathbf{x}(\cdot)$ 都会成为新的种子 $\mathbf{z}$；对每个种子应用相同的两个子程序，得到新的输出 $\mathbf{z}'$、$\mathbf{y}$，以及 $\{\mathbf{x}(1) \ldots, \mathbf{x}(m)\}$。所有层级完成后，算法返回任意层级产生的所有 $\mathbf{z}'$、$\mathbf{y}$ 和 $\mathbf{x}(\cdot)$ 中目标值最大的向量。

**算法 2** 主算法$(F, P, \gamma, t_s, \varepsilon, \delta)$

令 $\mathbf{z}(0)$ 为 $P$ 中通过定理 3.3.
执行 MAIN-RECURSIVE$(F, P, \gamma, t_s, \mathbf{z}(0)\ \varepsilon, \delta, 1)$。
函数 MAIN-RECURSIVE$(F, P, \gamma, t_s, \mathbf{z}\ \varepsilon, \delta, i)$
    $\mathbf{z}' =$ 盒约束最大化$(\mathbf{z})$
    令 $(\mathbf{y}, \mathbf{x}(1) \ldots, \mathbf{x}(m)) = \gamma$-FWG$(F, P, \mathbf{z}\ \gamma, t_s, \varepsilon)$ (算法 1)
    如果 $i < L$ 那么
        对于 $j = 1$ 到 $m$ 执行
            $\mathbf{y}(j)$ = MAIN-RECURSIVE$(F, P, \gamma, \mathbf{x}(j)\ t_s, \varepsilon, \delta, i+1)$
        结束循环
    结束条件判断
    返回在 $F$ 中使其最大化的向量：$\mathbf{z}'$、$\mathbf{y}$ 以及 $\{\mathbf{y}(j) \mid j \in [m]\}$中的向量。
结束函数

注意，对于固定的 $\gamma$，算法 1 执行的递归调用次数为 $O(m^L) = (\delta^{-1} + \varepsilon^{-1})^{O(1/\varepsilon)}$。对于任意常数 $\varepsilon$，该数量关于 $\delta^{-1}$是多项式级的。此外，每次单独的递归调用的运行时间关于 $\delta^{-1}$ 和 $n$是多项式级的。因此，算法 1 的总运行时间关于 $\delta^{-1}$ 和 $n$是多项式级的。

如果算法 2 的一次递归调用成功，即其 $\gamma$-FWG 内部运行成功，我们就称该调用为成功。第 2 节表明，算法 2 会执行足够多次递归调用，从而确保至少有一次调用成功；此外，它还会得到一个向量 $\mathbf{z}$，该向量是相对于 $\mathbf{o}'$ 的近似局部最大化器。分析进一步证明，对于这样一次调用，其相应向量 $\mathbf{z}$或向量 $\mathbf{y}$ 满足定理4.5中所述的性能保证。

## 4.3 主要结果

在算法 2的递归树中，我们聚焦于一条由继承者调用构成的指定路径。在这条路径上，-FWG 的“后备”保证会在每个层级向前传递。递归调用 $\mathcal{C}$ 满足以下任一条件时，称为继承者：

1. $\mathcal{C}$ 是唯一的第 1 层调用（即递归中的首次调用）；或者
2. $\mathcal{C}$ 由另一个继承者调用 $\mathcal{C}_p$ 调用，且该调用未成功；其输入种子 $\mathbf{z}$ 等于 $\mathbf{x}(j^\star)$，其中某个索引 $j^\star$ 满足定理 4.2 针对 $\gamma$-FWG 在 $\mathcal{C}_p$内的运行所述的第二种结果。

直观而言，如果一次继承调用 $\mathcal{C}_p$ 是 不成功的 （其 $\gamma$-FWG 的运行未返回一个较大的值 $\mathbf{y}$），那么定理 4.2 保证存在一个索引 $j^\star \in [m]$，使得 $\mathbf{x}(j^\star)$ 满足一阶

$\gamma$–弱 DR 证书。随后，我们将指定路径上下一位继承者的种子设为 $\mathbf{z} \leftarrow \mathbf{x}(j^\star)$。以下观察记录了该证书的不变量，我们随后将其结合起来以获得最终保证。

观察 1。 固定 $\gamma \in (0,1]$。算法 2 中的每次继承者递归调用都接收一个种子 $\mathbf{z} \in P$，满足

$$F(\mathbf{z}) \;\geq\; \frac{\gamma^2\,F(\mathbf{z}\vee\mathbf{o}) \;+\; F(\mathbf{z}\wedge\mathbf{o})}{1+\gamma^2} \;-\; O(\varepsilon)\,F(\mathbf{o}) \;-\; O(\delta\,LD^2). \tag{29}$$

证明。我们根据生成 $\mathbf{z}$ 的递归层级进行分类讨论。

情形 1：$\mathbf{z}$ 来自较后的层级。此处 $\mathbf{z} = \mathbf{x}(i)$ 对于某个索引 $i$ 由前一次$\gamma$-FWG 调用返回，其中“成功”的界（25）并不适用。因此（27）成立：

$$F(\mathbf{x}(i)) \;\geq\; \frac{\gamma^2 F(\mathbf{x}(i)\vee\mathbf{o}) + F(\mathbf{x}(i)\wedge\mathbf{o})}{1+\gamma^2} - \frac{\delta\,\gamma}{1+\gamma^2}\left(\max_{\mathbf{y}'\in Q(i)} F(\mathbf{y}') + \tfrac{1}{2}LD^2\right). \tag{30}$$

由于 $Q(i) \subseteq P$，我们有 $\max_{\mathbf{y}'\in Q(i)} F(\mathbf{y}') \leq \max_{\mathbf{y}'\in P} F(\mathbf{y}') \leq F(\mathbf{o})$，将其代入（30）并令 $\mathbf{z} = \mathbf{x}(i)$ 可得

$$F(\mathbf{z}) \;\geq\; \frac{\gamma^2 F(\mathbf{z}\vee\mathbf{o}) + F(\mathbf{z}\wedge\mathbf{o})}{1+\gamma^2} - \frac{\delta\,\gamma}{1+\gamma^2}\left(F(\mathbf{o}) + \tfrac{1}{2}LD^2\right), \tag{31}$$

这与不变量（29）一致，除了所述的 $O(\delta)\;F(\mathbf{o})$ 和 $O(\delta\,LD^2)$ 项之外。

情形 2：$\mathbf{z}$ 在第一个递归层级生成。此处 $\mathbf{z}$ 是定理 3.2 中的弱 DR 局部最大化例程的输出，其精度参数为 $\eta := \min\{\varepsilon,\delta\}$.应用定理 3.2 并令 $\mathbf{y} = \mathbf{o}$，再利用 $\max_{\mathbf{y}'\in P} F(\mathbf{y}') \leq F(\mathbf{o})$ 可得

$$F(\mathbf{z}) \;\geq\; \frac{\gamma^2 F(\mathbf{z}\vee\mathbf{o}) + F(\mathbf{z}\wedge\mathbf{o})}{1+\gamma^2} \;-\; \frac{\eta\,\gamma}{1+\gamma^2}\left(F(\mathbf{o}) + \tfrac{1}{2}LD^2\right). \tag{32}$$

由于 $\eta \leq \varepsilon$ 且 $\eta \leq \delta$ 根据 $\eta$的定义，(32) 中的误差项仍具有 $O(\varepsilon)\;F(\mathbf{o})$ $+O(\delta\,LD^2)$ 的形式，从而得到 (29)。

两种情况均证明了 (29)，证明得证。 □

在证明存在一个成功的继承者之前（推论 4.3），我们先指出一个简单的度量，它会在每个递推层级下降。

观察 2。 假设算法 2 的每个继承者递归调用均不成功（按定理 4.2的含义）。那么，对每个递归层级 $i \geq 1$，都存在一个位于层级 $i$ 的继承调用，它接收一个种子 $\mathbf{z}$，满足

$$F(\mathbf{z}\oplus\mathbf{o}) \;\leq\; F\big(\mathbf{z}(0)\oplus\mathbf{o}\big) \;-\; \varepsilon\,(i-1)\,F(\mathbf{o}). \tag{33}$$

证明。我们按层级 $i$ 进行归纳。

基例（$i = 1$）。唯一的第 1 层继承调用接收 $\mathbf{z} = \mathbf{z}(0)$。因此

$$F(\mathbf{z}\oplus\mathbf{o}) \;=\; F\big(\mathbf{z}(0)\oplus\mathbf{o}\big) \;\leq\; F\big(\mathbf{z}(0)\oplus\mathbf{o}\big) - \varepsilon\cdot 0\cdot F(\mathbf{o}), \tag{34}$$

这恰好是（33），其中 $i = 1$。

归纳步骤。假设该陈述对层级 $i-1 \geq 1$ 成立；也就是说，在层级 $i-1$ 上存在一个以 $\mathbf{z}$ 为种子的继承调用，使得

$$F(\mathbf{z}\oplus\mathbf{o}) \;\leq\; F\big(\mathbf{z}(0)\oplus\mathbf{o}\big) \;-\; \varepsilon\,(i-2)\,F(\mathbf{o}). \tag{35}$$

根据假设，该继承调用未成功。因此，定理 4.2中的“间隙”结果（26）适用于其 $\gamma$-FWG 内部运行，因此存在 $j^\star \in [m]$ 使得

$$F\big(\mathbf{x}(j^\star)\oplus\mathbf{o}\big) \;\leq\; F(\mathbf{z}\oplus\mathbf{o}) \;-\; \varepsilon\,F(\mathbf{o}). \tag{36}$$

结合（35）和（36）可得

$$F\big(\mathbf{x}(j^\star)\oplus\mathbf{o}\big) \;\leq\; F\big(\mathbf{z}(0)\oplus\mathbf{o}\big) \;-\; \varepsilon\,(i-1)\,F(\mathbf{o}). \tag{37}$$

根据继承者的定义，层级 $i$ 以 $\mathbf{z} \leftarrow \mathbf{x}(j^\star)$ 初始化的子调用本身就是一次继承调用，并满足条件 (37)，这恰好就是层级 33) 的条件 ( $i$。 □

弱 DR 细节（ $\gamma$-感知的局部值下界和 Frank–Wolfe 证书）只会影响种子 的质量 保证$\mathbf{x}(j^\star)$, 而不会影响 在 上的下降量 $F(\cdot \oplus o)$。因此$\varepsilon$逐层下降幅度与 DR 情形 [5], 保持一致，而 $\gamma$ 则会在后续用于得出分析结论的值下界中出现。

推论 4.3。 算法 2的某个递归调用是成功的继承调用。

证明。 反设不存在成功的继承调用。根据观察 2，在层级

$$i \;:=\; 1 + \left\lceil \tfrac{\gamma+1}{\gamma\,\varepsilon} \right\rceil \tag{38}$$

存在一个种子为 $\mathbf{z}$ 的继承调用，使得

$$\begin{aligned} F(\mathbf{z} \oplus \mathbf{o}) &\overset{(a)}{\leq}\; F(\mathbf{z}(0) \oplus \mathbf{o}) \;-\; \varepsilon\,(i-1)\,F(\mathbf{o}) \\ &\overset{(b)}{\leq}\; F(\mathbf{z}(0) \oplus \mathbf{o}) \;-\; \Big(1 + \tfrac{1}{\gamma}\Big)\,F(\mathbf{o}), \end{aligned} \tag{39}$$

其中，(a)由应用于层级 2 的观察 $i$得出，而 (b) 用到了 $i - 1 \geq (\gamma+1)/(\gamma\varepsilon)$，由 (38) 得出。

由于该调用（按假设）也不成功，定理 26 的间隙备选条件（4.2）适用，从而得到某个 $j$ 满足

$$F\big(\mathbf{x}(j) \oplus \mathbf{o}\big) \;\leq\; F(\mathbf{z} \oplus \mathbf{o}) \;-\; \varepsilon\,F(\mathbf{o}). \tag{40}$$

结合（39）和（40）可得

$$F\big(\mathbf{x}(j) \oplus \mathbf{o}\big) \;\leq\; F(\mathbf{z}(0) \oplus \mathbf{o}) \;-\; \Big(1 + \tfrac{1}{\gamma} + \varepsilon\Big)\,F(\mathbf{o}). \tag{41}$$

根据 $F$ 的非负性，我们有 $F\big(\mathbf{x}(j) \oplus \mathbf{o}\big) \geq 0$，因此（41）意味着

$$F(\mathbf{z}(0) \oplus \mathbf{o}) - F(\mathbf{o}) \;\geq\; \Big(\tfrac{1}{\gamma} + \varepsilon\Big)\,F(\mathbf{o}) \;>\; \tfrac{1}{\gamma}\,F(\mathbf{o}). \tag{42}$$

另一方面，根据 $\gamma$–弱 DR 性质，我们有

$$\begin{aligned} F(\mathbf{z}(0) \oplus \mathbf{o}) - F(\mathbf{o}) &\overset{(c)}{\leq}\; \frac{1}{\gamma}\Big(F\big(\mathbf{z}(0) \odot (\mathbf{1} - \mathbf{o})\big) - F(\mathbf{0})\Big) \\ &\overset{(d)}{\leq}\; \frac{1}{\gamma}\,F\big(\mathbf{z}(0) \odot (\mathbf{1} - \mathbf{o})\big) \end{aligned} \tag{43}$$

其中，（c）由 $\gamma$–弱 DR 定义得出，而（d）使用了 $F(\mathbf{0}) \geq 0$ （非负性）。结合（42）和（43）可得

$$F(\mathbf{o}) \;<\; F\big(\mathbf{z}(0) \odot (\mathbf{1} - \mathbf{o})\big). \tag{44}$$

由于 $P$是向下封闭的，且 $\mathbf{z}(0) \in P$，我们有 $\mathbf{z}(0) \odot (\mathbf{1} - \mathbf{o}) \leq$ 逐坐标 $\mathbf{z}(0) \odot (\mathbf{1} - \mathbf{o}) \in P$。式（44）中的严格改进与 $\mathbf{o}$ 在 $P$上的最优性相矛盾。因此，我们的假设是错误的，某个递归调用必然是成功的继承调用。 □

考虑算法 2中的任意一次成功的继承调用。将其输入种子记为 $\mathbf{z}^\star$，并令$\mathbf{y}^\star$ 为 $\gamma$-FWG 返回的高价值解，而 $\mathbf{z}'^\star$ 是同一次调用期间生成的子种子。调用定理 4.2（1）可得

$$F(\mathbf{y}^\star) \geq A_\gamma(t_s)\,F(\mathbf{o}) + B_\gamma(t_s)\,F(\mathbf{z}^\star \odot \mathbf{o}) + C_\gamma(t_s)\,F(\mathbf{z}^\star \oplus \mathbf{o}) - \delta\,LD^2. \tag{45}$$

为了对成功的继承调用进行后续分析，我们还需要关于 $F(\mathbf{z}'^\star)$ 的一个配套下界；下一个引理将给出该下界。

引理 4.4。 设 $F\colon [0,1]^n \to \mathbb{R}_{\geq 0}$ 为非负函数，且 $L$-光滑，并且 $\gamma$-弱 DR 次模，其中 $\gamma \in (0,1]$。设 $\mathbf{z}^* \in [0,1]^n$ 为提供给继承者递归调用的当前向量，并设 $\mathbf{z}'^* \leq \mathbf{z}^*$ 为推论 3.4 （在盒域 $[\mathbf{0},\mathbf{z}^*]$ 上以误差 $\varepsilon$运行）的输出。那么

$$F(\mathbf{z}'^*) \geq \max_{r\geq 0} \frac{\left(2\gamma^{3/2}r + \frac{\gamma}{1+\gamma^2}r^2\right)F(\mathbf{z}^* \odot \mathbf{o}) + \frac{\gamma^2}{1+\gamma^2}r^2 F(\mathbf{z}^* \oplus \mathbf{o})}{r^2 + 2\gamma^{3/2}r + 1} - O(\varepsilon)F(\mathbf{o}) - O(\delta L D^2). \quad (46)$$

证明。 将推论 3.4 应用于盒域 $[\mathbf{0},\mathbf{z}^*]$ （误差为 $\varepsilon$）可得

$$F(\mathbf{z}'^*) \geq \max_{r\geq 0} \frac{\left(2\gamma^{3/2} - 4\varepsilon\gamma^{9/2}\right)rF(\mathbf{z}^* \odot \mathbf{o}) + F(\mathbf{0}) + r^2 F(\mathbf{z}^*)}{r^2 + 2\gamma^{3/2}r + 1}. \quad (47)$$

由于 $F(\mathbf{0}) \geq 0$，去掉非负的 $F(\mathbf{0})$ 只会使右侧减小，因此得到 (a)；

$$\begin{aligned} F(\mathbf{z}'^*) &\overset{(a)}{\geq} \max_{r\geq 0} \frac{\left(2\gamma^{3/2} - 4\varepsilon\gamma^{9/2}\right) r\, F(\mathbf{z}^* \odot \mathbf{o}) \;+\; r^2\, F(\mathbf{z}^*)}{r^2 + 2\gamma^{3/2}r + 1} \\ &\overset{(b)}{\geq} \max_{r\geq 0} \frac{2\gamma^{3/2}\, r\, F(\mathbf{z}^* \odot \mathbf{o}) \;+\; r^2\, F(\mathbf{z}^*)}{r^2 + 2\gamma^{3/2}r + 1} \;-\; O(\varepsilon)\, F(\mathbf{o}), \end{aligned} \quad (48)$$

其中 (b) 用到了 $F(\mathbf{z}^* \odot \mathbf{o}) \leq F(\mathbf{o})$ 和界 $\dfrac{r}{r^2 + 2\gamma^{3/2}r + 1} \leq 1$，因此对 $r \geq 0$ 取最大值后，负扰动项 $-4\varepsilon\gamma^{9/2}\frac{r}{r^2+2\gamma^{3/2}r+1}F(\mathbf{z}^* \odot \mathbf{o})$ 至多为 $-O(\varepsilon)\ F(\mathbf{o})$。

接下来，由于 $\mathbf{z}^*$ 是一次继承者递归调用的种子，观察结果 1 给出

$$F(\mathbf{z}^*) \geq \frac{\gamma^2 F(\mathbf{z}^* \vee \mathbf{o}) + F(\mathbf{z}^* \wedge \mathbf{o})}{1+\gamma^2} - O(\varepsilon)F(\mathbf{o}) - O(\delta L D^2). \quad (49)$$

由于 $F \geq 0$和 $\gamma \leq 1$，我们还有 $F(\mathbf{z}^* \wedge \mathbf{o}) \geq \gamma\, F(\mathbf{z}^* \wedge \mathbf{o})$；将其与弱 DR 的“交换”不等式结合起来

$$\gamma\, F(\mathbf{x} \vee \mathbf{y}) + F(\mathbf{x} \wedge \mathbf{y}) \;\geq\; \gamma\, F(\mathbf{x} \oplus \mathbf{y}) + F(\mathbf{x} \odot \mathbf{y}) \quad (\forall\, \mathbf{x}, \mathbf{y} \in [0,1]^n), \quad (50)$$

我们得到精细下界

$$F(\mathbf{z}^*) \geq \frac{\gamma^2 F(\mathbf{z}^* \oplus \mathbf{o}) + \gamma\, F(\mathbf{z}^* \odot \mathbf{o})}{1+\gamma^2} - O(\varepsilon)F(\mathbf{o}) - O(\delta L D^2). \quad (51)$$

将（51）代入（48），并注意到 $\frac{r^2}{r^2+2\gamma^{3/2}r+1} \leq 1$，保留加性误差$-O$（$\varepsilon$）$F$（$\mathbf{o}$）$-O$（$\delta L D^2$），并得到系数

$$\begin{aligned} &\frac{2\gamma^{3/2}r}{r^2 + 2\gamma^{3/2}r + 1} + \frac{\gamma}{1+\gamma^2} \cdot \frac{r^2}{r^2 + 2\gamma^{3/2}r + 1} \quad \text{for } F(\mathbf{z}^* \odot \mathbf{o}), \\ &\qquad\qquad \frac{\gamma^2}{1+\gamma^2} \cdot \frac{r^2}{r^2 + 2\gamma^{3/2}r + 1} \quad \text{for } F(\mathbf{z}^* \oplus \mathbf{o}), \end{aligned}$$

这恰好就是所声称的形式。 □

我们已经为任意成功的继承调用确立了两个经认证的下界：一个针对 Frank–Wolfe 引导的输出$\mathbf{y}^\star$（定理 4.2(1)），另一个针对盒域限制子调用 $\mathbf{z}'^*$（引理 4.4）。由于算法返回这两个值中较大的一个，这两个下界的任意凸组合仍然是算法输出的有效下界。令 $\alpha \in [0,1]$ 为混合参数。将这两个下界写成统一形式并加以组合，可得，对于任意$r \geq 0$ 以及 $t_s \in$ （0，1），

$$\begin{aligned} F(\mathrm{ALG}) \;\geq\; &(1-\alpha)\, A_\gamma(t_s)\, F(\mathbf{o}) + \Big[(1-\alpha)\, B_\gamma(t_s) \;+\; \alpha\, D_\gamma(r)\Big]\, F(\mathbf{z}^\star \odot \mathbf{o}) \\ &+ \Big[(1-\alpha)\, C_\gamma(t_s) \;+\; \alpha\, E_\gamma(r)\Big]\, F(\mathbf{z}^\star \oplus \mathbf{o}) - \; O(\varepsilon)\, F(\mathbf{o}) \;-\; O(\delta\, L D^2), \end{aligned} \quad (52)$$

其中

$$D_\gamma(r) \;:=\; \frac{2\gamma^{3/2}\,r + \frac{\gamma}{1+\gamma^2}\,r^2}{r^2 + 2\gamma^{3/2} r + 1}, \qquad E_\gamma(r) \;:=\; \frac{\frac{\gamma^2}{1+\gamma^2}\,r^2}{r^2 + 2\gamma^{3/2} r + 1}. \tag{53}$$

为了提取出位于 $F(\mathbf{o})$ 前的纯乘法因子，我们选择参数，使得乘在 $F(\mathbf{z}^\star \odot \mathbf{o})$ 和 $F(\mathbf{z}^\star \oplus \mathbf{o})$ 前的系数非负，这样便可利用 $F$ 的非负性将这些项舍去。定义可行集

$$\mathcal{F}_\gamma := \Big\{(\alpha, r, t_s) \in [0,1] \times [0,\infty) \times (0,1) \;:\; (1-\alpha)B_\gamma(t_s) + \alpha D_\gamma(r) \geq 0, (1-\alpha)C_\gamma(t_s) + \alpha E_\gamma(r) \geq 0\Big\}. \tag{54}$$

对于任意 $(\alpha, r, t_s) \in \mathcal{F}_\gamma$，不等式 (52) 给出

$$F(\mathrm{ALG}) \;\geq\; \Big[(1-\alpha)\,A_\gamma(t_s) \;-\; O(\varepsilon)\Big]\,F(\mathbf{o}) \;-\; O(\delta\, LD^2). \tag{55}$$

这促使我们对首项因子进行优化：

$$\Phi_\gamma \;:=\; \max_{(\alpha,r,t_s)\in\mathcal{F}_\gamma} (1-\alpha)\,A_\gamma(t_s). \tag{56}$$

我们的近似保证 $\Phi_\gamma$ 定义为 (56) 中最大化问题的最优值。这一优化仅涉及三个标量参数 $(\alpha, r, t_s)$，并受 (54) 中简单的线性可行性约束，因此对于任意固定的 $\gamma$，该问题都很容易通过数值方法求解。特别地，尽管 $\Phi_\gamma$ 似乎无法表示为 $\gamma$的闭式表达式，但通过直接网格搜索，可以在多项式时间内（关于离散步长的倒数）将其计算到任意所需精度。

对于每个固定的 $\gamma$，我们通过在有界区域 $r \in [0, r_{\max}]$, $t_s \in (0, 1)$ 上对 $(r, t_s)$ 进行显式网格搜索来求解 (56)，然后利用 (54) 中的线性约束以闭式形式优化 $\alpha$。对于每个网格点 $(r, t_s)$，我们确定使 (54) 中两个不等式均成立的可行 $\alpha \in [0,1]$ 区间，再选择使 $(1-\alpha)A_\gamma(t_s)$ 最大化的区间端点。由此得到一个候选三元组 $(\alpha, r, t_s)$ 及其对应的候选值 $\Phi_\gamma$，并在所有网格点中保留最优者。运行时间关于网格步长的倒数为多项式。这正是我们用于生成图1和参数表的 Python代码所实现的过程（我们采用了 $r_{\max} = 10$，因为更大的 $r_{\max}$ 值并未改变结果）。

边界情形 $\gamma = 1$。若干系数（例如 $A_\gamma$、$B_\gamma$、$C_\gamma$）的分母中包含 $(1-\gamma)$ 的因子；因此，(52) 中的表达式可能呈现表面上的 $0/0$ 形式，即 $\gamma \to 1$。对于所有此类项，我们均取其连续极限，并在需要时通过洛必达法则求值。将这些极限代入（52）可得到我们混合界的 DR（即 $\gamma = 1$）特例，并且在 56处进行优化，$\gamma = 1$ 可重现 Buchbinder 和 Feldman 当前最佳的 DR 保证（0。401） [5]。

优化所实现的结果。优化保证 $\Phi_\gamma$ （a）在 $\gamma = 1$ 处恰好与当前最佳 DR 常数一致，并且（b）严格优于非单调弱 DR 基准 $\kappa(\gamma) = \gamma e^{-\gamma}$ 对所有 $\gamma \in (0, 1)$。直观而言，该混合方法通过（$\alpha$、$r$、$t_s$）平衡 Frank–Wolfe 引导的进展与盒约束改进，从而为每个 $\gamma$ 确定最佳因子。

定理 4.5。固定 $\gamma \in (0,1]$ 和 $\delta \in (0,1)$。设 $F\colon [0,1]^n \to \mathbb{R}_{\geq 0}$ 是一个非负、$L$-光滑、$\gamma$-弱 DR 次模函数，且 $P \subseteq [0,1]^n$ 是一个直径为 $D$的向下封闭元可解凸体。存在一种多项式时间算法，返回一个点 $\mathbf{x} \in P$，使得

$$F(\mathbf{x}) \;\geq\; \Phi_\gamma \cdot \max_{\mathbf{y}\in P} F(\mathbf{y}) \;-\; O\big(\delta\, D^2 L\big)\,, \tag{57}$$

其中 $\Phi_\gamma$ 是 (56) 的最优值。

证明。 令 $\mathbf{o} \in \arg\max_{\mathbf{y}\in P} F(\mathbf{y})$。根据推论 4.3，算法 2 至少有一次成功的继承调用，其种子为 $\mathbf{z}^\star$。定理 4.2 (1) 和引理4.4 为返回的候选解 $\mathbf{y}^\star$ 和 $\mathbf{z}'^\star$提供了两个经认证的下界。对这两个下界取任意凸组合，即可得到（52），其中 $D_\gamma$， $E_\gamma$ 由（53）给出。

选择 $(\alpha, r, t_s) \in \mathcal{F}_\gamma$(参见 (54))，使得 $F(\mathbf{z}^\star \odot \mathbf{o})$ 和 $F(\mathbf{z}^\star \oplus \mathbf{o})$ 中的系数均非负。去掉这些 (52) 非负项后，得到 (55):

$$F(\mathrm{ALG}) \;\geq\; \big[(1-\alpha)A_\gamma(t_s) - O(\varepsilon)\big]\, F(\mathbf{o}) \;-\; O(\delta L D^2). \tag{58}$$

在可行的 $(\alpha, r, t_s)$ 上取最大值，由 ( $\Phi_\gamma$56) 可得 ，因此，对于最优选择，

$$F(\mathrm{ALG}) \;\geq\; \big[\Phi_\gamma - O(\varepsilon)\big]\, F(\mathbf{o}) \;-\; O(\delta L D^2). \tag{59}$$

最后，令 $\varepsilon = \Theta(\delta)$ 吸收 $-O(\varepsilon)F(\mathbf{o})$ 项，并入 $-O(\delta L D^2)$ 平滑误差。 □

## 5 结论

本文提出了一个统一的无投影框架，用于在向下闭凸体上最大化连续非单调的 $\gamma$-弱 DR 次模函数。我们的方法将一种 $\gamma$-感知的 Frank–Wolfe 引导的测量连续贪心算法与一种 $\gamma$-感知的双重贪心算法相结合，并通过三个可调参数（$\alpha, r, t_s$）对它们的证书进行凸组合优化。在整个弱 DR 谱范围内，所得保证 $\Phi_\gamma$ 严格 优于经典的非单调基准 $\kappa(\gamma) = \gamma e^{-\gamma}$，且在 DR 边界（$\gamma = 1$）处 达到当前最佳常数 0.401。我们在图 1 和表 1 中展示了相较于先前工作的改进。

## 参考文献

[1] Per Austrin、Siavosh Benabbas 和 Konstantinos Georgiou。偏置带来更好的平衡：一种用于最大二分的 0.8776 近似算法。《ACM 算法汇刊》（TALG），13(1):1–27，2016年。[2] An Bian、Kfir Y. Levy、Andreas Krause 和 Joachim M. Buhmann。连续 DR-次模最大化：结构与算法。载于《神经信息处理系统进展》（NeurIPS），2017年。[3] Andrew An Bian、Baharan Mirzasoleiman、Joachim Buhmann 和 Andreas Krause。保证的非凸优化：连续域上的子模最大化。载于《第 20 届国际人工智能与统计学会议》论文集，2017年。[4]Yatao Bian、Joachim M. Buhmann 和 Andreas Krause。最优连续 DR-次模最大化及其在可证明的平均场推断中的应用。载于《第 36 届国际机器学习会议》（ICML）论文集，2019年。[5] Niv Buchbinder 和 Moran Feldman。利用 DR-次模函数的新界进行约束子模最大化。载于第 56 届年度 ACM 理论计算研讨会论文集，2024年。[6] Niv Buchbinder、Moran Feldman、Joseph Naor 和 Roy Schwartz。一种紧确的线性时间 (1/2)-近似算法，用于无约束子模最大化。SIAM 计算杂志，44(5):1384–1402，2015年。[7] Chandra Chekuri 和 Sanjeev Khanna。多重背包问题的多项式时间近似方案。SIAM 计算杂志，35(3):713–728，2005年。[8] Chandra Chekuri、Jan Vondr´ak 和 Rico Zenklusen。通过多线性扩展最大化子模函数。载于第二十二届 ACM-SIAM 年度离散算法研讨会（SODA）论文集，第 114–125 页。SIAM，2011年。

[9] Lin Chen、Hamed Hassani 和 Amin Karbasi。在线连续子模最大化。载于第 35 届国际机器学习会议论文集，2018年。[10] Michele Conforti 和 G´erard Cornu´ejols。子模集合函数、拟阵与贪心算法：紧确的最坏情况界及 Rado-Edmonds 定理的一些推广。《离散应用数学》，7(3):251–274，1984年。[11] Gerard Cornuejols、Marshall Fisher 和 George L Nemhauser。关于无容量限制选址问题。载于《离散数学年刊》，第 1 卷，第 163–177 页。爱思唯尔，1977年。[12] Uriel Feige 和 Jan Vondr´ak。分配问题的近似算法：改进 1-1/e 的近似因子。载于 2006 年第 47 届 IEEE 计算机科学基础年度研讨会（FOCS'06）论文集，第 667–676 页。IEEE，2006年。[13] Moran Feldman、Joseph Naor、Roy Schwartz 和 Justin Ward。针对 $k$-交换系统的改进近似算法。载于 Camil Demetrescu 和 Magn´us M. Halld´orsson（编），《欧洲算法研讨会》（ESA），《计算机科学讲义》第 6942 卷，第 784–798 页。施普林格，2011年。[14] Marshall L Fisher、George L Nemhauser 和 Laurence A Wolsey。最大化子模集合函数的近似分析——II。载于《多面体组合学：献给 DR Fulkerson 的纪念文集》，第 73–87 页。施普林格，2009年。[15] Shayan Oveis Gharan 和 Jan Vondr´ak。通过模拟退火实现子模最大化。载于第二十二届 ACM-SIAM 离散算法年度研讨会论文集，第 1098–1116 页。SIAM，2011年。[16] Hamed Hassani、Mahdi Soltanolkotabi 和 Amin Karbasi。用于子模最大化的梯度方法。《神经信息处理系统进展》，第 30 卷，2017年。[17] Johan H˚astad。一些最优不可近似性结果。《美国计算机协会会刊》（JACM），48(4):798–859，2001年。[18] Martin Jaggi。重新审视 Frank–Wolfe：无投影稀疏凸优化。载于第 30 届国际机器学习会议论文集，第 427–435 页，2013年。[19] Richard M Karp。组合问题之间的可约性。载于《整数规划 50 年（1958–2008）：从早期到最新进展》，第 219–241 页。施普林格，2009年。[20] Bernhard Korte 和 Dirk Hausmann。独立系统贪心启发式的分析。载于《离散数学年刊》，第 2 卷，第 65–74 页。爱思唯尔，1978年。[21] Simon Lacoste-Julien 和 Martin Jaggi。关于 Frank–Wolfe 算法的线性收敛。《神经信息处理系统进展》，2015年。[22] Loay Mualem 和 Moran Feldman。在子模最大化中利用部分单调性。《神经信息处理系统进展》，2022年。[23] George L Nemhauser 和 Laurence A Wolsey。用于近似次模集合函数最大值的最佳算法。《运筹学数学》，3(3):177–188，1978年。[24] George L Nemhauser、Laurence A Wolsey 和 Marshall L Fisher。最大化子模集合函数的近似分析——I。《数学规划》，14(1):265–294，1978年。

[25] Rad Niazadeh、Tim Roughgarden 和 Joshua R. Wang。《连续非单调子模与 DR-次模最大化的最优算法》。机器学习研究杂志，21:125:1–125:31，2020年。[26] Guanyu Nie、Vaneet Aggarwal 和 Christopher Quinn。《具有随机长期约束的在线 DR-次模最大化梯度方法》。神经信息处理系统进展，37:20510–20539，2024年。[27] Mohammad Pedramfar 和 Vaneet Aggarwal。《从线性优化到可线性化优化：一种应用于平稳和非平稳 DR-次模优化的新框架》。神经信息处理系统进展，37:37626–37664，2024年。[28] Mohammad Pedramfar、Yididiya Y. Nadew、Christopher John Quinn 和 Vaneet Aggarwal。《对抗性DR-次模优化的统一无投影算法》。载于第十二届国际学习表征会议，2024年。

[29] Mohammad Pedramfar、Christopher Quinn 和 Vaneet Aggarwal。最大化连续 DR-次模函数的统一方法。《神经信息处理系统进展》，36：61103–61114，2023年。

[30]Mohammad Pedramfar、Christopher John Quinn 和 Vaneet Aggarwal。最大化连续 $\gamma$ 弱 DR 次模函数的方法。https://optimization-online.org/?p=25915，2024年3月。

[31] Benjamin Qi。关于最大化非单调次模函数与线性函数之和。在《国际算法与计算研讨会》（ISAAC）上，LIPIcs 第248卷，第41:1–41:16页，2022年。

[32] Mingrui Zhang、Lin Chen、Hamed Hassani 和 Amin Karbasi。在线连续子模最大化：从全信息到老虎机反馈。在《神经信息处理系统进展》（NeurIPS）上，2019年。

[33] Qixin Zhang、Zengde Deng、Zaiyi Chen、Haoyuan Hu 和 Yu Yang。《随机连续次模最大化：通过非盲函数提升》。载于《第39届国际机器学习会议论文集》，第26116–26134页，2022年。

# A 第2节引理的证明

在本附录中，我们详细证明 $\gamma$-弱 DR 次模函数的基本性质，并将 DR-次模函数的经典结果扩展到 [5] -弱 DR 次模的广义 $\gamma$情形。为清晰起见，我们先重述引理，再给出证明。

引理 2.1。设 $F:\ [0,1]^n \to \mathbb{R}_{\geq 0}$ 可微且为 $\gamma$-弱 DR-次模函数。那么对于所有的 $\mathbf{x}$、$\mathbf{y} \in [0,1]^n$ 和 $\lambda \in [0,1]$，以下结论成立：

（1）若 $\mathbf{x} \leq \mathbf{y}$，则

$$F\big(\lambda \mathbf{x} + (1-\lambda)\mathbf{y}\big) \ \geq \ \frac{\lambda\, F(\mathbf{x}) + \gamma^2(1-\lambda)\, F(\mathbf{y})}{\lambda + \gamma^2(1-\lambda)}. \tag{60}$$

等价地，

$$F\big((1-\lambda)\mathbf{x} + \lambda \mathbf{y}\big) \ \geq \ \frac{(1-\lambda)\, F(\mathbf{x}) + \gamma^2\lambda\, F(\mathbf{y})}{(1-\lambda) + \gamma^2\lambda}. \tag{61}$$

(2) 若 $\mathbf{y} \geq \mathbf{0}$ 且 $\mathbf{x} + \mathbf{y} \in [0,1]^n$，则

$$F(\mathbf{x} + \lambda \mathbf{y}) - F(\mathbf{x}) \ \geq \ \frac{\gamma^2\lambda}{1 - \lambda + \gamma^2\lambda}\big(F(\mathbf{x}+\mathbf{y}) - F(\mathbf{x})\big). \tag{62}$$

证明。固定 $\mathbf{x} \in [0,1]^n$ 以及一个方向 $\mathbf{v} \geq 0$，使得对所有 $t \in [0,1]$，$\mathbf{x} + t\mathbf{v} \in [0,1]^n$。定义单变量函数

$$\phi(t) \;:=\; F(\mathbf{x} + t\mathbf{v}), \qquad t \in [0,1]. \tag{63}$$

根据链式法则，$\phi$ 可微，且

$$\phi'(t) \;=\; \big\langle \nabla F(\mathbf{x} + t\mathbf{v}),\, \mathbf{v} \big\rangle. \tag{64}$$

现在固定 $0 \leq s \leq t \leq 1$。由于 $\mathbf{v} \geq 0$，我们有 $\mathbf{x} + s\mathbf{v} \leq \mathbf{x} + t\mathbf{v}$，根据 $\gamma$-弱 DR-子模性可得

$$\nabla F(\mathbf{x} + s\mathbf{v}) \;\geq\; \gamma\, \nabla F(\mathbf{x} + t\mathbf{v}). \tag{65}$$

对 (65) 两边取与 $\mathbf{v} \geq 0$ 的内积，并利用 (64，得到

$$\phi'(s) \;=\; \big\langle \nabla F(\mathbf{x} + s\mathbf{v}), \mathbf{v} \big\rangle \;\geq\; \gamma \big\langle \nabla F(\mathbf{x} + t\mathbf{v}), \mathbf{v} \big\rangle \;=\; \gamma\, \phi'(t), \qquad 0 \leq s \leq t \leq 1. \tag{66}$$

接下来，固定 $\lambda \in (0,1)$。对于 $t \in [\lambda, 1]$，将 66) 与 $s = \lambda$ 结合应用 ( $\phi'(\lambda) \geq \gamma\, \phi'(t)$,so

$$\phi'(t) \;\leq\; \frac{1}{\gamma}\, \phi'(\lambda).$$

对 $t \in [\lambda, 1]$ 上的这个上界积分，并利用微积分基本定理，可得

$$\phi(1) - \phi(\lambda) \;=\; \int_{\lambda}^{1} \phi'(t)\, dt \;\leq\; \int_{\lambda}^{1} \frac{1}{\gamma}\, \phi'(\lambda)\, dt \;=\; \frac{1-\lambda}{\gamma}\, \phi'(\lambda). \tag{67}$$

类似地，对于 $s \in [0, \lambda]$，将 (66) 与 $t = \lambda$ 结合应用，可得 $\phi'(s) \geq \gamma\, \phi'(\lambda)$。对 $s \in [0, \lambda]$ 上的这个下界积分，得到

$$\phi(\lambda) - \phi(0) \;=\; \int_{0}^{\lambda} \phi'(s)\, ds \;\geq\; \int_{0}^{\lambda} \gamma\, \phi'(\lambda)\, ds \;=\; \gamma\, \lambda\, \phi'(\lambda). \tag{68}$$

由（67）可得

$$\phi'(\lambda) \;\geq\; \frac{\gamma}{1-\lambda} \big(\phi(1) - \phi(\lambda)\big),$$

将此下界代入（68）可得

$$\phi(\lambda) - \phi(0) \;\geq\; \gamma\, \lambda \cdot \frac{\gamma}{1-\lambda} \big(\phi(1) - \phi(\lambda)\big) \;=\; \frac{\gamma^2 \lambda}{1-\lambda} \big(\phi(1) - \phi(\lambda)\big). \tag{69}$$

将（69）两边乘以 $1 - \lambda$ 并展开，可得

$$(1-\lambda)\, \phi(\lambda) - (1-\lambda)\, \phi(0) \;\geq\; \gamma^2 \lambda\, \phi(1) - \gamma^2 \lambda\, \phi(\lambda).$$

重新整理含有 $\phi(\lambda)$ 的项至左侧，其余项移至右侧，得到 s

$$\big(1 - \lambda + \gamma^2 \lambda\big)\, \phi(\lambda) \;\geq\; (1-\lambda)\, \phi(0) + \gamma^2 \lambda\, \phi(1).$$

两边除以 $1 - \lambda + \gamma^2 \lambda > 0$，得到

$$\phi(\lambda) \;\geq\; \frac{(1-\lambda)\, \phi(0) + \gamma^2 \lambda\, \phi(1)}{(1-\lambda) + \gamma^2 \lambda}. \tag{70}$$

为证明第 (2) 部分，取 $\mathbf{v} = \mathbf{y} \geq 0$ 和 $\phi(t) = F(\mathbf{x} + t\mathbf{y})$，如式 (63) 所示。由于 $\mathbf{x} + \mathbf{y} \in [0,1]^n$，整个线段 $\{\mathbf{x} + t\mathbf{y} : \; t \in [0,1]\}$ 都位于 $[0,1]^n$中，因此上述论证适用。在这种情况下，

$$\phi(0) = F(\mathbf{x}), \qquad \phi(1) = F(\mathbf{x} + \mathbf{y}), \qquad \phi(\lambda) = F(\mathbf{x} + \lambda \mathbf{y}).$$

将这些表达式代入式（70）可得

$$F(\mathbf{x} + \lambda \mathbf{y}) \;\geq\; \frac{(1-\lambda) F(\mathbf{x}) + \gamma^2 \lambda F(\mathbf{x} + \mathbf{y})}{(1-\lambda) + \gamma^2 \lambda}. \tag{71}$$

将 $F$（$\mathbf{x}$）从式（71）两边同时减去，可得

$$F(\mathbf{x}+\lambda\mathbf{y})-F(\mathbf{x}) \;\geq\; \frac{(1-\lambda)F(\mathbf{x})+\gamma^2\lambda F(\mathbf{x}+\mathbf{y})}{(1-\lambda)+\gamma^2\lambda} \;-\; F(\mathbf{x}).$$

将 $F(\mathbf{x})$ 写成 $\frac{(1-\lambda)+\gamma^2\lambda}{(1-\lambda)+\gamma^2\lambda}F(\mathbf{x})$，并化简分子，得到

$$F(\mathbf{x}+\lambda\mathbf{y})-F(\mathbf{x}) \;\geq\; \frac{\gamma^2\lambda}{1-\lambda+\gamma^2\lambda}\big(F(\mathbf{x}+\mathbf{y})-F(\mathbf{x})\big),$$

这恰好就是 (62)。这就证明了第 (2) 部分。

对于第 (1) 部分，假设 $\mathbf{x}\leq\mathbf{y}$，并定义 $\mathbf{v}:=\mathbf{y}-\mathbf{x}\geq 0$。对于 $t\in[0,1]$ 集合

$$\phi(t) \;:=\; F\big(\mathbf{x}+t(\mathbf{y}-\mathbf{x})\big) \;=\; F\big((1-t)\mathbf{x}+t\mathbf{y}\big). \tag{72}$$

同样，$\mathbf{x}$ 与 $\mathbf{y}$ 之间的线段 $[0,1]^n$包含于其中，因此适用 (70)。这里

$$\phi(0)=F(\mathbf{x}), \qquad \phi(1)=F(\mathbf{y}), \qquad \phi(\lambda)=F\big((1-\lambda)\mathbf{x}+\lambda\mathbf{y}\big).$$

代入 (70)，得到

$$F\big((1-\lambda)\mathbf{x}+\lambda\mathbf{y}\big) \;\geq\; \frac{(1-\lambda)\,F(\mathbf{x})+\gamma^2\lambda\,F(\mathbf{y})}{(1-\lambda)+\gamma^2\lambda}, \tag{73}$$

这恰好就是 (61)。最后，在 (73) 中用 $1-\lambda$ 替换 $\lambda$，得到

$$F\big(\lambda\mathbf{x}+(1-\lambda)\mathbf{y}\big) \;\geq\; \frac{\lambda\,F(\mathbf{x})+\gamma^2(1-\lambda)\,F(\mathbf{y})}{\lambda+\gamma^2(1-\lambda)},$$

即（60）。这证明了第（1）部分。 □

引理 2.2. 设 $F:\;[0,1]^n\to\mathbb{R}_{\geq 0}$ 是一个可微的 $\gamma$-弱 DR 次模函数。那么，对于任意 $\mathbf{x}$、$\mathbf{y}\in[0,1]^n$ 且 $\mathbf{y}\geq\mathbf{0}$，以下不等式成立：

1. 若 $\mathbf{x}+\mathbf{y}\leq\mathbf{1}$，则
$$\big\langle\nabla F(\mathbf{x}),\,\mathbf{y}\big\rangle \;\geq\; \gamma\big(F(\mathbf{x}+\mathbf{y})-F(\mathbf{x})\big). \tag{74}$$

2. 若 $\mathbf{x}-\mathbf{y}\geq\mathbf{0}$，则
$$\big\langle\nabla F(\mathbf{x}),\,\mathbf{y}\big\rangle \;\leq\; \frac{1}{\gamma}\big(F(\mathbf{x})-F(\mathbf{x}-\mathbf{y})\big). \tag{75}$$

证明。我们先证明第（1）部分。假设 $\mathbf{x}+\mathbf{y}\leq\mathbf{1}$，并定义

$$g(t) \;:=\; F(\mathbf{x}+t\mathbf{y}), \qquad t\in[0,1]. \tag{76}$$

条件 $\mathbf{x}+\mathbf{y}\leq\mathbf{1}$ 且 $\mathbf{y}\geq\mathbf{0}$ 意味着对所有 $t\in[0,1]$，都有 $\mathbf{x}+t\mathbf{y}\in[0,1]^n$。根据链式法则，

$$g'(t) \;=\; \big\langle\nabla F(\mathbf{x}+t\mathbf{y}),\,\mathbf{y}\big\rangle. \tag{77}$$

对于每个 $t\in[0,1]$，我们有 $\mathbf{x}\leq\mathbf{x}+t\mathbf{y}$，因此根据 $\gamma$-弱 DR-次模性，

$$\nabla F(\mathbf{x}) \;\geq\; \gamma\,\nabla F(\mathbf{x}+t\mathbf{y}), \qquad 0\leq t\leq 1. \tag{78}$$

对（78）两边取与 $\mathbf{y}\geq\mathbf{0}$ 的内积，并利用（77），可得

$$\big\langle\nabla F(\mathbf{x}),\,\mathbf{y}\big\rangle \;\geq\; \gamma\,\big\langle\nabla F(\mathbf{x}+t\mathbf{y}),\,\mathbf{y}\big\rangle \;=\; \gamma\,g'(t), \qquad 0\leq t\leq 1. \tag{79}$$

等价地，

$$g'(t) \;\leq\; \frac{1}{\gamma}\,\big\langle\nabla F(\mathbf{x}),\,\mathbf{y}\big\rangle, \qquad 0\leq t\leq 1. \tag{80}$$

对界限（80）在 $t \in [0,1]$ 上积分，并利用微积分基本定理可得

$$F(\mathbf{x}+\mathbf{y}) - F(\mathbf{x}) \;=\; \int_0^1 g'(t)\,dt \;\le\; \int_0^1 \frac{1}{\gamma}\,\big\langle \nabla F(\mathbf{x}),\, \mathbf{y}\big\rangle\,dt \;=\; \frac{1}{\gamma}\,\big\langle \nabla F(\mathbf{x}),\, \mathbf{y}\big\rangle. \tag{81}$$

重新整理（81）得

$$\big\langle \nabla F(\mathbf{x}),\, \mathbf{y}\big\rangle \;\ge\; \gamma\big(F(\mathbf{x}+\mathbf{y}) - F(\mathbf{x})\big), \tag{82}$$

这恰好就是（74）。

现在证明第（2）部分。假设 $\mathbf{x}-\mathbf{y} \ge \mathbf{0}$ 并定义

$$h(t) \;:=\; F(\mathbf{x}-t\mathbf{y}), \qquad t \in [0,1]. \tag{83}$$

条件 $\mathbf{x}-\mathbf{y} \ge \mathbf{0}$ 且 $\mathbf{y} \ge \mathbf{0}$ 意味着，对于所有 $t \in [0,1]$，都有 $\mathbf{x}-t\mathbf{y} \in [0,1]^n$。再次利用链式法则，

$$h'(t) \;=\; \big\langle \nabla F(\mathbf{x}-t\mathbf{y}),\, -\mathbf{y}\big\rangle \;=\; -\big\langle \nabla F(\mathbf{x}-t\mathbf{y}),\, \mathbf{y}\big\rangle. \tag{84}$$

对每个 $t \in [0,1]$，我们都有 $\mathbf{x}-t\mathbf{y} \le \mathbf{x}$，因此由 $\gamma$-弱 DR-次模性可得

$$\nabla F(\mathbf{x}-t\mathbf{y}) \;\ge\; \gamma\,\nabla F(\mathbf{x}), \qquad 0 \le t \le 1. \tag{85}$$

与 $\mathbf{y} \ge \mathbf{0}$ 取内积，并利用（84），得到

$$\big\langle \nabla F(\mathbf{x}-t\mathbf{y}),\, \mathbf{y}\big\rangle \;\ge\; \gamma\,\big\langle \nabla F(\mathbf{x}),\, \mathbf{y}\big\rangle, \qquad 0 \le t \le 1, \tag{86}$$

因此

$$h'(t) \;=\; -\big\langle \nabla F(\mathbf{x}-t\mathbf{y}),\, \mathbf{y}\big\rangle \;\le\; -\gamma\,\big\langle \nabla F(\mathbf{x}),\, \mathbf{y}\big\rangle, \qquad 0 \le t \le 1. \tag{87}$$

对 87) 在 $t \in [0,1]$ 上积分，并应用微积分基本定理，得到

$$F(\mathbf{x}-\mathbf{y}) - F(\mathbf{x}) \;=\; \int_0^1 h'(t)\,dt \;\le\; \int_0^1 -\gamma\,\big\langle \nabla F(\mathbf{x}),\, \mathbf{y}\big\rangle\,dt \;=\; -\gamma\,\big\langle \nabla F(\mathbf{x}),\, \mathbf{y}\big\rangle. \tag{88}$$

将 88) 乘以 $-1$可得

$$F(\mathbf{x}) - F(\mathbf{x}-\mathbf{y}) \;\ge\; \gamma\,\big\langle \nabla F(\mathbf{x}),\, \mathbf{y}\big\rangle. \tag{89}$$

重新整理 89)，得到

$$\big\langle \nabla F(\mathbf{x}),\, \mathbf{y}\big\rangle \;\le\; \frac{1}{\gamma}\big(F(\mathbf{x}) - F(\mathbf{x}-\mathbf{y})\big), \tag{90}$$

这正是 75)。证明完成。 □

引理 2.3。设 $F:\ [0,1]^n \to \mathbb{R}_{\ge 0}$ 非负且为 $\gamma$–弱 DR 次模，其中某个$\gamma \in (0,1]$。对于任意固定的 $\mathbf{y} \in [0,1]^n$，定义

$$G_{\oplus}(\mathbf{x}) \;:=\; F(\mathbf{x} \oplus \mathbf{y}) \tag{91}$$

and

$$G_{\odot}(\mathbf{x}) \;:=\; F(\mathbf{x} \odot \mathbf{y}). \tag{92}$$

那么， $G_{\oplus}$ 和 $G_{\odot}$ 均为非负且是 $\gamma$–弱 DR 次模的；也就是说，对于所有满足 $\mathbf{x}^{(1)}$、$\mathbf{x}^{(2)} \in [0,1]^n$ 且 $\mathbf{x}^{(1)} \le \mathbf{x}^{(2)}$ 的情形，任意坐标 $u \in [n]$，以及任意步长 $p \in [0, 1-\mathbf{x}^{(2)}_u]$（使更新后的点仍处于 $[0,1]^n$ 中），我们有

$$G_{\oplus}\big(\mathbf{x}^{(1)} + p\,\mathbf{e}_u\big) - G_{\oplus}(\mathbf{x}^{(1)}) \;\ge\; \gamma\Big(G_{\oplus}\big(\mathbf{x}^{(2)} + p\,\mathbf{e}_u\big) - G_{\oplus}(\mathbf{x}^{(2)})\Big) \tag{93}$$

and

$$G_{\odot}\big(\mathbf{x}^{(1)} + p\,\mathbf{e}_u\big) - G_{\odot}(\mathbf{x}^{(1)}) \;\ge\; \gamma\Big(G_{\odot}\big(\mathbf{x}^{(2)} + p\,\mathbf{e}_u\big) - G_{\odot}(\mathbf{x}^{(2)})\Big). \tag{94}$$

证明。 非负性 $G_{\oplus}$ 和 $G_{\odot}$ 可直接由（91）、（92）以及 $F$的非负性得出。取 $\mathbf{x}^{(1)}$，$\mathbf{x}^{(2)} \in [0,1]^n$，满足 $\mathbf{x}^{(1)} \le \mathbf{x}^{(2)}$，再取坐标 $u \in [n]$和步长 $p \in [0, 1-\mathbf{x}_u^{(2)}]$，使得 $\mathbf{x}^{(j)} + p\,\mathbf{e}_u \in [0,1]^n$ 对于 $j = 1,\ 2$。

我们首先考虑 $\oplus$ 情形。设

$$\mathbf{z}^{(j)} \ := \ \mathbf{x}^{(j)} \oplus \mathbf{y}, \qquad j \in \{1,2\}. \tag{95}$$

由于 $\mathbf{x}^{(1)} \le \mathbf{x}^{(2)}$，且映射 $x \mapsto x \oplus y$ 关于 $x$ 单调不减，因此有

$$\mathbf{z}^{(1)} \ \le \ \mathbf{z}^{(2)}. \tag{96}$$

利用逐坐标恒等式

$$(\mathbf{x} \oplus \mathbf{y})_i \ = \ x_i + y_i - x_i y_i, \tag{97}$$

我们可以用 $\mathbf{z}^{(j)}$ 表示更新 $(\mathbf{x}^{(j)} + p\mathbf{e}_u) \oplus \mathbf{y}$。事实上，$\mathbf{x}^{(j)}$ 只有第 $u$ 个坐标发生变化，因此对每个 $j \in \{1,2\}$，有

$$(\mathbf{x}^{(j)} + p\,\mathbf{e}_u) \oplus \mathbf{y} \ = \ \mathbf{z}^{(j)} + \alpha\,\mathbf{e}_u, \tag{98}$$

其中

$$\alpha \ := \ p\,(1 - \mathbf{y}_u). \tag{99}$$

这是将 $x_u^{(j)} + p$ 代入表达式 (97) 并化简后得到的。

接下来，我们验证更新后的点在 $[0,1]^n$ 的坐标 $u$处仍然属于该集合。利用（97），结合 $x_u^{(2)}$可得

$$1 - \mathbf{z}_u^{(2)} \ = \ 1 - \left(x_u^{(2)} + y_u - x_u^{(2)} y_u\right) \ = \ (1 - \mathbf{x}_u^{(2)})(1 - \mathbf{y}_u). \tag{100}$$

由假设可知 $p \le 1 - \mathbf{x}_u^{(2)}$，且 $1 - \mathbf{y}_u \ge 0$，因此可得

$$\alpha \ = \ p(1 - \mathbf{y}_u) \ \le \ (1 - \mathbf{x}_u^{(2)})(1 - \mathbf{y}_u) \ = \ 1 - \mathbf{z}_u^{(2)}. \tag{101}$$

因此，$\mathbf{z}^{(2)} + \alpha\,\mathbf{e}_u$ 仍属于 $[0,1]^n$。

现在我们利用 $\gamma$的 $F$弱 DR-次模性。由（96）和（101），我们可以将 $\gamma$弱 DR-次模性的定义应用于点对（$\mathbf{z}^{(1)}$,$\mathbf{z}^{(2)}$），步长为 $\alpha$，沿坐标 $u$ 进行，从而得到

$$F\big(\mathbf{z}^{(1)} + \alpha\,\mathbf{e}_u\big) - F(\mathbf{z}^{(1)}) \ \ge \ \gamma\Big(F\big(\mathbf{z}^{(2)} + \alpha\,\mathbf{e}_u\big) - F(\mathbf{z}^{(2)})\Big). \tag{102}$$

利用（91）和（98），我们可以将（102）的左端和右端分别改写为关于 $G_{\oplus}$ 的形式：

$$F\big(\mathbf{z}^{(j)} + \alpha\,\mathbf{e}_u\big) \ = \ G_{\oplus}\big(\mathbf{x}^{(j)} + p\,\mathbf{e}_u\big), \qquad F(\mathbf{z}^{(j)}) = G_{\oplus}(\mathbf{x}^{(j)}), \quad j = 1,2. \tag{103}$$

将（103）代入（102）可得

$$G_{\oplus}\big(\mathbf{x}^{(1)} + p\,\mathbf{e}_u\big) - G_{\oplus}(\mathbf{x}^{(1)}) \ \ge \ \gamma\Big(G_{\oplus}\big(\mathbf{x}^{(2)} + p\,\mathbf{e}_u\big) - G_{\oplus}(\mathbf{x}^{(2)})\Big), \tag{104}$$

这恰好是（93）。这表明 $G_{\oplus}$ 是 $\gamma$–弱 DR 次模。

现在考虑 $\odot$ 情形。令

$$\mathbf{w}^{(j)} \ := \ \mathbf{x}^{(j)} \odot \mathbf{y}, \qquad j \in \{1,2\}. \tag{105}$$

由于映射 $x \mapsto x \odot y$ 关于 $x$ 也是单调不减的，因此我们再次有

$$\mathbf{w}^{(1)} \ \le \ \mathbf{w}^{(2)}. \tag{106}$$

根据 $\odot$的定义，

$$(\mathbf{x} \odot \mathbf{y})_i \ = \ x_i y_i, \tag{107}$$

因此，将坐标 $u$ 中的 $\mathbf{x}^{(j)}$ 更新 $p$ 后得到

$$(\mathbf{x}^{(j)} + p\,\mathbf{e}_u) \odot \mathbf{y} \;=\; \mathbf{w}^{(j)} + \beta\,\mathbf{e}_u, \tag{108}$$

其中

$$\beta \;:=\; p\,\mathbf{y}_u. \tag{109}$$

再次 我们检查更新后的点仍然位于 $[0,1]^n$ 的坐标 $u$处。由（107）可得 e

$$\mathbf{w}_u^{(2)} \;=\; \mathbf{x}_u^{(2)}\mathbf{y}_u, \tag{110}$$

so

$$1 - \mathbf{w}_u^{(2)} \;=\; 1 - \mathbf{x}_u^{(2)}\mathbf{y}_u. \tag{111}$$

利用 $p \le 1 - \mathbf{x}_u^{(2)}$和 $\mathbf{y}_u \le 1$，我们得到

$$\beta \;=\; p\,\mathbf{y}_u \;\le\; (1 - \mathbf{x}_u^{(2)})\mathbf{y}_u \;\le\; 1 - \mathbf{x}_u^{(2)}\mathbf{y}_u \;=\; 1 - \mathbf{w}_u^{(2)}. \tag{112}$$

因此 $\mathbf{w}^{(2)} + \beta\,\mathbf{e}_u \in [0,1]^n$。

现在我们再次使用 $\gamma$的–弱 DR-次模性 $F$。由（106）和（112），将定义应用于（$\mathbf{w}^{(1)}$, $\mathbf{w}^{(2)}$），在坐标 $\beta$ 上采用步长 $u$ 可得

$$F\big(\mathbf{w}^{(1)} + \beta\,\mathbf{e}_u\big) - F(\mathbf{w}^{(1)}) \;\ge\; \gamma\Big(F\big(\mathbf{w}^{(2)} + \beta\,\mathbf{e}_u\big) - F(\mathbf{w}^{(2)})\Big). \tag{113}$$

利用 (92) 和 (108)，我们可以将 (113) 改写为

$$G_\odot\big(\mathbf{x}^{(1)} + p\,\mathbf{e}_u\big) - G_\odot(\mathbf{x}^{(1)}) \;\ge\; \gamma\Big(G_\odot\big(\mathbf{x}^{(2)} + p\,\mathbf{e}_u\big) - G_\odot(\mathbf{x}^{(2)})\Big), \tag{114}$$

which 恰好等于 (94)。因此 $G_\odot$ 也具有 $\gamma$–弱 DR 次模性，从而完成证明。 □

# B Frank–Wolfe 算法与定理 3.2的证明

本节针对 $\gamma$–弱 DR 次模情形构造一阶证书，并用它证明我们的主要结果——定理 3.2。论证遵循 从局部到整体 的框架：(i) 在 $\mathbf{x}$ 处的局部最优性条件给出 $\gamma$–加权比较，比较 $F(\mathbf{x})$ 与任意比较对象 $\mathbf{y}$ 的并/交值（引理 3.1）；(ii) 一个 Frank–Wolfe 变体生成点 $\mathbf{x} \in P$，它满足针对 每个 一阶证书，且该证书 对所有 $\mathbf{y} \in P$ 均成立（引理 B.1）；(iii) 将两者结合，可得整体值界，其随 $\gamma$ 平滑退化，并在 $\gamma = 1$处达到经典 DR 保证。

算法设置。 我们将调用以下 [5]中的 Frank–Wolfe 型例程。为便于表述，我们假设 $\delta^{-1} \in \mathbb{N}$；若不满足，则将 $\delta$ 替换为 $1/\lceil \delta^{-1} \rceil$，而不影响渐近行为。

**算法 3** Frank–Wolfe 变体$(F, P, \delta)$

令 $\mathbf{x}(0)$ 为 $P$ 中任意一个向量。对于 $i = 1$ 到 $\delta^{-2}$，
执行令 $\mathbf{z}(i) \in \arg\max_{\mathbf{y} \in P} \langle \mathbf{y},\ \nabla F(\mathbf{x}(i-1)) \rangle$。令 $\mathbf{x}(i) \leftarrow$
$(1-\delta)\,\mathbf{x}(i-1) + \delta\,\mathbf{z}(i)$。结束循环令 $i^* \in \arg\min$
$_{1 \le i \le \delta^{-2}} \{ \langle \mathbf{z}(i) - \mathbf{x}(i-1),\ \nabla F(\mathbf{x}(i-1)) \rangle \}$。返回 $\mathbf{x}(i^* - 1)$。

如 [5], 所观察到的，更新规则

$$\mathbf{x}(i) \;=\; (1-\delta)\,\mathbf{x}(i-1) + \delta\,\mathbf{z}(i), \qquad \mathbf{z}(i) \in P,$$

使所有迭代点都保持在 $P$中；特别是，对每个 $0 \le i \le \delta^{-2}$，$\mathbf{x}(i) \in P$。

下一个引理将局部最优性条件转化为一种基于格的比较，并在 $\gamma$ 中进行插值；当 $\gamma = 1$ 时，它与经典的 $\frac{1}{2}\big(F(\mathbf{x} \vee \mathbf{y}) + F(\mathbf{x} \wedge \mathbf{y})\big)$ 界一致。

引理 3.1. 设 $F:\ [0,1]^n \to \mathbb{R}_{\geq 0}$ 可微且为 $\gamma$–弱 DR 次模。若 $\mathbf{x}$ 相对于 $\mathbf{y}$ 是局部最优解，则意味着;

$$\langle \mathbf{y}-\mathbf{x}, \nabla F(\mathbf{x})\rangle \ \leq\ 0, \tag{115}$$

*then*

$$F(\mathbf{x}) \ \geq\ \frac{\gamma^2\, F(\mathbf{x}\vee\mathbf{y}) + F(\mathbf{x}\wedge\mathbf{y})}{1+\gamma^2}. \tag{116}$$

证明。 从局部最优性条件（115）出发，我们将 $\mathbf{y}\ -\mathbf{x}$ 分解为

$$\mathbf{y}-\mathbf{x}\ =\ (\mathbf{y}\vee\mathbf{x}-\mathbf{x})\ -\ (\mathbf{x}-\mathbf{y}\wedge\mathbf{x}). \tag{117}$$

将（117）代入（115）可得

$$0\ \geq\ \langle \mathbf{y}-\mathbf{x}, \nabla F(\mathbf{x})\rangle\ =\ \langle \mathbf{y}\vee\mathbf{x}-\mathbf{x}, \nabla F(\mathbf{x})\rangle\ -\ \langle \mathbf{x}-\mathbf{y}\wedge\mathbf{x}, \nabla F(\mathbf{x})\rangle. \tag{118}$$

现在，我们利用引理 2.2 对每个内积进行界定。首先，注意到 $\mathbf{y}\ \vee\mathbf{x}-\mathbf{x}\ \geq\mathbf{0}$ 且

$$\mathbf{x}+(\mathbf{y}\vee\mathbf{x}-\mathbf{x})\ =\ \mathbf{y}\vee\mathbf{x}\ \in\ [0,1]^n,$$

因此，引理 2.2(1) 适用于方向 $\mathbf{y}\ \vee\mathbf{x}\ -\mathbf{x}$。我们得到

$$\langle \nabla F(\mathbf{x}),\, \mathbf{y}\vee\mathbf{x}-\mathbf{x}\rangle\ \geq\ \gamma\big(F(\mathbf{y}\vee\mathbf{x})-F(\mathbf{x})\big). \tag{119}$$

同理，$\mathbf{x}-\mathbf{y}\ \wedge\mathbf{x}\ \geq\mathbf{0}$ 且

$$\mathbf{x}-(\mathbf{x}-\mathbf{y}\wedge\mathbf{x})\ =\ \mathbf{y}\wedge\mathbf{x}\ \in\ [0,1]^n,$$

因此，引理 2.2(2) 适用于方向 $\mathbf{x}\ -\mathbf{y}\ \wedge\mathbf{x}$。这给出

$$\langle \nabla F(\mathbf{x}),\, \mathbf{x}-\mathbf{y}\wedge\mathbf{x}\rangle\ \leq\ \frac{1}{\gamma}\big(F(\mathbf{x})-F(\mathbf{y}\wedge\mathbf{x})\big). \tag{120}$$

将界（119）和（120代入（118）可得

$$0\ \geq\ \gamma\big(F(\mathbf{y}\vee\mathbf{x})-F(\mathbf{x})\big)\ -\ \frac{1}{\gamma}\big(F(\mathbf{x})-F(\mathbf{y}\wedge\mathbf{x})\big). \tag{121}$$

展开（121），可得

$$0\ \geq\ \gamma F(\mathbf{y}\vee\mathbf{x})\ -\ \gamma F(\mathbf{x})\ -\ \frac{1}{\gamma}F(\mathbf{x})\ +\ \frac{1}{\gamma}F(\mathbf{y}\wedge\mathbf{x}). \tag{122}$$

重新整理（122），将涉及 $F(\mathbf{x})$ 的项移至右侧，可得

$$\gamma F(\mathbf{y}\vee\mathbf{x})\ +\ \frac{1}{\gamma}F(\mathbf{y}\wedge\mathbf{x})\ \leq\ \left(\gamma+\frac{1}{\gamma}\right)F(\mathbf{x}). \tag{123}$$

将（123）两边同时除以 $\gamma+1/\gamma=(1+\gamma^2)/\gamma$ 可得

$$F(\mathbf{x})\ \geq\ \frac{\gamma F(\mathbf{y}\vee\mathbf{x})+\frac{1}{\gamma}F(\mathbf{y}\wedge\mathbf{x})}{\gamma+\frac{1}{\gamma}}\ =\ \frac{\gamma^2 F(\mathbf{y}\vee\mathbf{x})+F(\mathbf{y}\wedge\mathbf{x})}{1+\gamma^2}. \tag{124}$$

注意到 $\mathbf{y}\ \vee\mathbf{x}\ =\mathbf{x}\ \vee\mathbf{y}$ 且 $\mathbf{y}\ \wedge\mathbf{x}\ =\mathbf{x}\ \wedge\mathbf{y}$ 后，这正是（116） □

下一个引理是算法 3 给出的标准基于光滑性的保证（参见 [5]中的定理 2.4）。

引理 B.1 （ [5]中的定理 2.4）。 设 $F:\ [0,1]^n \to \mathbb{R}_{\geq 0}$ 为非负且 $L$-光滑，设$P \subseteq [0,1]^n$为直径为 $D$的可求解凸体，并设 $\delta \in (0,1)$。存在一个多项式时间算法，返回满足以下条件的 $\mathbf{x} \in P$：

$$\langle \mathbf{y}-\mathbf{x}, \nabla F(\mathbf{x})\rangle \ \leq\ \delta\left[\max_{\mathbf{z}\in P} F(\mathbf{z}) \ +\ \frac{LD^2}{2}\right] \qquad \textit{for all } \mathbf{y}\in P. \tag{125}$$

将一致性证书（125）与弱 DR 不等式（引理 2.2）以及局部到格点比较（引理 3.1）结合，即可得到我们的主要界限。

定理 3.2. 设 $F\colon [0,1]^n \to \mathbb{R}_{\geq 0}$ 非负且为 $L$-光滑，并假设对于某个 $\gamma \in (0,1]$， $F$ 是 $\gamma$ 弱 DR 次模。设 $P \subseteq [0,1]^n$ 为直径为 $D$ 的可求解凸体，且令 $\delta \in (0,1)$。那么存在一个多项式时间算法，输出 $\mathbf{x} \in P$，使得对于每个 $\mathbf{y} \in P$，

$$F(\mathbf{x}) \ \geq\ \frac{\gamma^2 F(\mathbf{x}\vee\mathbf{y}) + F(\mathbf{x}\wedge\mathbf{y})}{1+\gamma^2} \ -\ \frac{\delta\gamma}{1+\gamma^2}\left[\max_{\mathbf{z}\in P} F(\mathbf{z}) \ +\ \frac{LD^2}{2}\right]. \tag{126}$$

证明。 令 $\mathbf{x} \in P$ 为引理 B.1 返回的结果；则对于所有 125，式（$\mathbf{y} \in P$）成立。与引理 3.1 的证明类似，引理 2.2 意味着对于每个 $\mathbf{y} \in [0,1]^n$，

$$\langle \mathbf{y}-\mathbf{x}, \nabla F(\mathbf{x})\rangle \ \geq\ \gamma\, F(\mathbf{x}\vee\mathbf{y}) \ +\ \frac{1}{\gamma} F(\mathbf{x}\wedge\mathbf{y}) \ -\ \frac{1+\gamma^2}{\gamma} F(\mathbf{x}). \tag{127}$$

将 (125) 和 (127) 对任意 $\mathbf{y} \in P$ 结合可得

$$\delta\left[\max_{\mathbf{z}\in P} F(\mathbf{z}) \ +\ \frac{LD^2}{2}\right] \ \geq\ \gamma\, F(\mathbf{x}\vee\mathbf{y}) \ +\ \frac{1}{\gamma} F(\mathbf{x}\wedge\mathbf{y}) \ -\ \frac{1+\gamma^2}{\gamma} F(\mathbf{x}). \tag{128}$$

将 (128) 中的 $F(\mathbf{x})$ 项移至等式左边，整理得

$$\frac{1+\gamma^2}{\gamma} F(\mathbf{x}) \ \geq\ \gamma\, F(\mathbf{x}\vee\mathbf{y}) \ +\ \frac{1}{\gamma} F(\mathbf{x}\wedge\mathbf{y}) \ -\ \delta\left[\max_{\mathbf{z}\in P} F(\mathbf{z}) \ +\ \frac{LD^2}{2}\right]. \tag{129}$$

将 (129) 两边同时乘以 $\frac{\gamma}{1+\gamma^2}$ 可得

$$F(\mathbf{x}) \ \geq\ \frac{\gamma^2}{1+\gamma^2} F(\mathbf{x}\vee\mathbf{y}) \ +\ \frac{1}{1+\gamma^2} F(\mathbf{x}\wedge\mathbf{y}) \ -\ \frac{\delta\gamma}{1+\gamma^2}\left[\max_{\mathbf{z}\in P} F(\mathbf{z}) \ +\ \frac{LD^2}{2}\right]. \tag{130}$$

这正是 (126)，证明完毕。 □

## C 双重贪心算法及定理 3.3 的证明

本附录将构建并分析一种 $\gamma$-感知型双重贪心算法，其保证见定理 3.3。我们的分析采用算法 4 中的网格离散化变体。为具体起见，我们假设 $\varepsilon^{-1}$ 是偶整数；若不是，则将 $\varepsilon$ 替换为 $\varepsilon' = 1/(2\lceil\varepsilon^{-1}\rceil) \in (0,1]$,，这样除了所述的 $\varepsilon$-依赖性外，界限保持不变。全文中，我们用 $\mathbf{o} \in [0,1]^n$ 表示任意比较对象；当 $\mathbf{o}$ 取为最大化点时，有 $F(\mathbf{o}) = \max_{\mathbf{u}\in[0,1]^n} F(\mathbf{u})$。

定理 3.3。存在一种多项式时间算法，对于给定的非负 $\gamma$-弱 DR 次模函数 $F\colon [0,1]^n \to \mathbb{R}_{\geq 0}$和参数 $\varepsilon \in (0,1)$，输出 $\mathbf{x} \in [0,1]^n$，使得对于每个固定的 $\mathbf{o} \in [0,1]^n$，

$$F(\mathbf{x}) \ \geq\ \max_{r\geq 0} \frac{\left(2\gamma^{3/2} - 4\varepsilon\,\gamma^{9/2}\right) r\, F(\mathbf{o}) \ +\ F(\mathbf{0}) \ +\ r^2\, F(\mathbf{1})}{r^2 \ +\ 2\gamma^{3/2} r \ +\ 1}. \tag{131}$$

算法 4 维护两个向量 $\mathbf{x}$，$\mathbf{y}\ \in [0,1]^n$，它们分别从 $\mathbf{0}$ 和 $\mathbf{1}$开始，并且单调收敛 到同一个向量，每次迭代使一个坐标相同。赋给所选坐标的新值（在 $\mathbf{x}$ 和 $\mathbf{y}$中）取自一个均匀网格；我们将该网格记为

$$V \;:=\; \Big\{\, j\,\frac{\varepsilon}{n} \;:\; j\in\mathbb{Z},\; 0\le j\le n\,\varepsilon^{-1}\Big\} \;\subseteq\; [0,1].$$

正如以下引理所示，使用 $V$所导致的离散化损失得到了明确控制，所得下界随 $\gamma$ 连续变化，并且在 $\gamma=1$时退化为经典 DR 保证。

---

算法 **4** $\gamma$-双重贪心

---

1：输入：对 $F$的预言机访问：$[0,1]^n\to\mathbb{R}_{\ge 0}$，基集 $N$，网格参数 $\varepsilon\in(0,1)$，弱 DR 参数 $\gamma\in(0,1]$。2：令 $V\leftarrow\left\{\frac{j\varepsilon}{n}:\quad j\in\mathbb{Z},\ 0\le j\le n\varepsilon^{-1}\right\}\subseteq[0,1]$。3：按任意顺序将 $N$中的元素记为 $u_1,\cdots,u_n$。4：令 $\mathbf{x}\leftarrow\mathbf{0}$ 和 $\mathbf{y}\ \leftarrow\mathbf{1}$。5：对于 $i=1$ 到 $n$，执行 6：$a_i\in\arg\max_{v\in V}F(\mathbf{x}+v\,\mathbf{1}_{u_i})$；$\Delta_{a,i}\leftarrow F(\mathbf{x}+a_i\,\mathbf{1}_{u_i})\ -F(\mathbf{x})$。7：$b_i\in\arg\max_{v\in V}F(\mathbf{y}\ -v\mathbf{1}_{u_i})$；$\Delta_{b,i}\leftarrow F(\mathbf{y}\ -b_i\,\mathbf{1}_{u_i})\ -F(\mathbf{y})$。8：如果 $\Delta_{a,i}+\gamma\,\Delta_{b,i}>0$，则 9：$w_i\leftarrow{}^{\Delta_a}{}_{,i}\,a_i+\gamma\,\Delta_{b,i}(1-b_i)\,\Delta_{a,i}+\gamma\,\Delta_{b,i}$。10：否则$\Delta_{a,i}=0$ 且$\Delta_{b,i}=0$11：$w_i\leftarrow 1-b_i$12：结束 if 13：设置 $x_{u_i}\leftarrow w_i$ 和 $y_{u_i}\leftarrow w_i$。14：结束 for 15：返回 $\mathbf{x}$

---

现在，我们量化算法 4 返回的向量所达到的值。令 $\mathbf{x}(i)$ 和$\mathbf{y}(i)$ 分别表示主循环进行 $\mathbf{x}$ 和 $\mathbf{y}$ 次迭代后 $i$ 的取值。对于一个固定坐标 $u_i$，令

$$v^*\ \in\ \arg\max_{v\in[0,1]}\ F\big(\mathbf{x}(i-1)+v\,\mathbf{e}_{u_i}\big) \tag{132}$$

表示在第 $u_i$ 次迭代时，沿第 $i$ 个坐标方向取得的连续（不受网格限制）最大化点。下一个引理界定了算法 4 所用网格选取 $a_i$ 的离散化损失。对于每个 $\gamma\in(0,1]$ 均成立；当 $\gamma=1$时，该界与 [5] 的精确界相同。

引理 C.1. 对任意整数 $i\in\{1,\dots,n\}$，以下结论成立：

$$\textit{If } v^*\ge\tfrac12,\quad F\big(\mathbf{x}(i-1)+a_i\,\mathbf{e}_{u_i}\big)\ \ge\ \max_{v\in[0,1]}F\big(\mathbf{x}(i-1)+v\,\mathbf{e}_{u_i}\big)\ -\ \frac{2\varepsilon}{\gamma^2 n}\,F(\mathbf{o}). \tag{133}$$

$$\textit{If } v^*<\tfrac12,\quad F\big(\mathbf{x}(i-1)+a_i\,\mathbf{e}_{u_i}\big)\ \ge\ \max_{v\in[0,1]}F\big(\mathbf{x}(i-1)+v\,\mathbf{e}_{u_i}\big)\ -\ \frac{2\varepsilon}{n}\,\gamma^2\,F(\mathbf{o}). \tag{134}$$

证明。令 $v^*\in[0,1]$ 为使函数取得最大值的点

$$v\ \longmapsto\ F\big(\mathbf{x}(i-1)+v\,\mathbf{e}_{u_i}\big), \tag{135}$$

使得（132）成立。根据 $v^*$ 的大小，我们分两种情况讨论。

情况 1：$v^*\ge\frac12$。令 $v\in V$ 为满足 $v\le v^*$ 的最大网格点。根据 $a_i$ 是网格上的最大化点的定义，我们有

$$F\big(\mathbf{x}(i-1)+v^*\mathbf{e}_{u_i}\big)-F\big(\mathbf{x}(i-1)+a_i\mathbf{e}_{u_i}\big)\ \le\ F\big(\mathbf{x}(i-1)+v^*\mathbf{e}_{u_i}\big)-F\big(\mathbf{x}(i-1)+v\mathbf{e}_{u_i}\big), \tag{136}$$

因为 $F(\mathbf{x}(i-1) + a_i \mathbf{e}_{u_i})$ 至少不小于任何其他网格点处的值，尤其是 $v$ 处的值。定义单变量函数

$$\phi(t) \;:=\; F\big(\mathbf{x}(i-1) + t\,\mathbf{e}_{u_i}\big), \qquad t \in [0,1]. \tag{137}$$

根据可微性和链式法则，

$$\phi'(t) \;=\; \big\langle \nabla F\big(\mathbf{x}(i-1) + t\,\mathbf{e}_{u_i}\big),\, \mathbf{e}_{u_i} \big\rangle. \tag{138}$$

沿此坐标方向，$\gamma$-弱 DR-次模性意味着对于任意 $0 \le s \le t \le 1$，

$$\phi'(s) \;\ge\; \gamma\,\phi'(t). \tag{139}$$

应用（139），令 $s = v$ 且 $t \in [v, v^*]$ 可得

$$\phi'(t) \;\le\; \frac{1}{\gamma}\,\phi'(v), \qquad t \in [v, v^*]. \tag{140}$$

对（140）关于 $t \in [v, v^*]$ 积分，得到

$$\phi(v^*) - \phi(v) \;=\; \int_v^{v^*} \phi'(t)\,dt \;\le\; \int_v^{v^*} \frac{1}{\gamma}\,\phi'(v)\,dt \;=\; \frac{v^* - v}{\gamma}\,\phi'(v). \tag{141}$$

类似地，应用（139），令 $0 \le s \le v$ 且 $t = v$ 可得

$$\phi'(s) \;\ge\; \gamma\,\phi'(v), \qquad s \in [0, v]. \tag{142}$$

对（142）关于 $s \in [0, v]$ 积分，得到

$$\phi(v) - \phi(0) \;=\; \int_0^{v} \phi'(s)\,ds \;\ge\; \int_0^{v} \gamma\,\phi'(v)\,ds \;=\; v\,\gamma\,\phi'(v). \tag{143}$$

重新整理（143）得

$$\phi'(v) \;\le\; \frac{\phi(v) - \phi(0)}{\gamma\, v}. \tag{144}$$

结合（141）和（144），得到

$$\phi(v^*) - \phi(v) \;\le\; \frac{v^* - v}{\gamma} \cdot \frac{\phi(v) - \phi(0)}{\gamma\, v} \;=\; \frac{v^* - v}{\gamma^2 v}\,\big[\phi(v) - \phi(0)\big]. \tag{145}$$

现在我们对因子 $\phi(v) - \phi(0)$ 进行界定。由于 $\phi(v) = F(\mathbf{x}(i-1) + v\,\mathbf{e}_{u_i})$ 且 $\phi(0) = F(\mathbf{x}(i-1))$，由 $F$ 的非负性和 o 的最优性可得

$$\phi(v) - \phi(0) \;\le\; F(\mathbf{o}), \tag{146}$$

因为 $\phi(v) \le F(\mathbf{o})$ 且 $\phi(0) \ge 0$。将（146）代入（145）得

$$\phi(v^*) - \phi(v) \;\le\; \frac{v^* - v}{\gamma^2 v}\,F(\mathbf{o}). \tag{147}$$

根据网格 $V$ 的构造，我们有

$$v^* - v \;\le\; \frac{\varepsilon}{n}. \tag{148}$$

此外，由于 $v^* \ge \frac{1}{2}$ 和 $\frac{1}{2} \in V$，选择不超过 $v^*$的最大网格点 $v$ 意味着

$$v \;\ge\; \tfrac{1}{2}. \tag{149}$$

结合式（147）、（148）和（149），可得

$$\phi(v^*) - \phi(v) \;\le\; \frac{\varepsilon/n}{\gamma^2 \cdot (1/2)}\, F(\mathbf{o}) \;=\; \frac{2\varepsilon}{\gamma^2 n}\, F(\mathbf{o}). \tag{150}$$

利用（136）和（137），（150）可得

$$F\big(\mathbf{x}(i-1) + v^* \mathbf{e}_{u_i}\big) - F\big(\mathbf{x}(i-1) + a_i \mathbf{e}_{u_i}\big) \;\le\; \frac{2\varepsilon}{\gamma^2 n}\, F(\mathbf{o}). \tag{151}$$

由于 $v^*$ 使 $v \mapsto F\big(\mathbf{x}(i-1) + v\mathbf{e}_{u_i}\big)$ 在 $[0,1]$，（151）上的最大化等价于（133），故第一项结论得证。

情形 2：　$v^* < \frac{1}{2}$。令 $v \in V$ 为满足 $v \ge v^*$的最小网格点。则

$$v - v^* \;\le\; \frac{\varepsilon}{n}, \tag{152}$$

并且，由于 $\varepsilon^{-1}$ 为偶数且$\frac{1}{2} \in V$，我们有

$$v \;\le\; \tfrac{1}{2}. \tag{153}$$

根据 $a_i$ 作为网格上的最大化点的选取方式，

$$F\big(\mathbf{x}(i-1) + a_i\, \mathbf{e}_{u_i}\big) - F\big(\mathbf{x}(i-1) + v^* \mathbf{e}_{u_i}\big) \;\ge\; F\big(\mathbf{x}(i-1) + v\, \mathbf{e}_{u_i}\big) - F\big(\mathbf{x}(i-1) + v^* \mathbf{e}_{u_i}\big). \tag{154}$$

使用与（ $\phi$ ）中相同的函数137），对于 $t \in [v^*, v]$ 我们有 $t \le v$，因此由（139），

$$\phi'(t) \;\ge\; \gamma\, \phi'(v), \qquad t \in [v^*, v]. \tag{155}$$

对（155）在 $t \in [v^*, v]$ 上积分，得到

$$\phi(v) - \phi(v^*) \;=\; \int_{v^*}^{v} \phi'(t)\, dt \;\ge\; \int_{v^*}^{v} \gamma\, \phi'(v)\, dt \;=\; (v - v^*)\, \gamma\, \phi'(v). \tag{156}$$

对于 $s \in [v, 1]$，有 $v \le s$，因此（139）意味着

$$\phi'(v) \;\ge\; \gamma\, \phi'(s), \qquad s \in [v, 1]. \tag{157}$$

对（157）在 $s \in [v, 1]$ 上积分，得到

$$\phi(1) - \phi(v) \;=\; \int_{v}^{1} \phi'(s)\, ds \;\le\; \int_{v}^{1} \frac{1}{\gamma}\, \phi'(v)\, ds \;=\; \frac{1 - v}{\gamma}\, \phi'(v). \tag{158}$$

整理（158），得到

$$\phi'(v) \;\ge\; \frac{\gamma}{1 - v}\, \big[\phi(1) - \phi(v)\big]. \tag{159}$$

结合（156）和（159），可得

$$\phi(v) - \phi(v^*) \;\ge\; (v - v^*)\, \gamma \cdot \frac{\gamma}{1 - v}\, \big[\phi(1) - \phi(v)\big] \;=\; \frac{(v - v^*)\, \gamma^2}{1 - v}\, \big[\phi(1) - \phi(v)\big]. \tag{160}$$

利用（137），注意到

$$\phi(1) - \phi(v) \;=\; F\big(\mathbf{x}(i-1) + \mathbf{e}_{u_i}\big) - F\big(\mathbf{x}(i-1) + v\, \mathbf{e}_{u_i}\big) \;\le\; F(\mathbf{o}), \tag{161}$$

因为 $F$ 是非负的，并在$\mathbf{o}$处取最大值。将（161）以及界（152）–（153）代入（160），得

$$\begin{aligned}
\phi(v) - \phi(v^*) \;&\ge\; \frac{(v - v^*)\, \gamma^2}{1 - v}\, \big[\phi(1) - \phi(v)\big] \\
&\ge\; \frac{(v - v^*)\, \gamma^2}{1 - v}\, (-F(\mathbf{o})) \qquad (\text{since } \phi(1) - \phi(v) \ge -F(\mathbf{o})) \\
&\ge\; -\frac{\varepsilon/n}{1/2}\, \gamma^2 F(\mathbf{o}) \;=\; -\frac{2\varepsilon}{n}\, \gamma^2 F(\mathbf{o}),
\end{aligned} \tag{162}$$

其中我们用到了 $v - v^* \le \varepsilon/n$ 和 $v \le \frac{1}{2}$（因此 $1 - v \ge \frac{1}{2}$）。

利用 (154)、(137) 和 (162)，我们得到

$$F\big(\mathbf{x}(i-1) + a_i\,\mathbf{e}_{u_i}\big) - F\big(\mathbf{x}(i-1) + v^*\mathbf{e}_{u_i}\big) \;\ge\; -\frac{2\varepsilon}{n}\,\gamma^2 F(\mathbf{o}). \tag{163}$$

由于 $v^*$ 使 $v \mapsto F(\mathbf{x}(i-1) + v\,\mathbf{e}^{u_i})$ 在 $[0,1]$, (163) 上取最大值等价于 (134)，故第二个断言得证。

情形 (133) 和 (134) 合在一起即可证明该引理。 □

类似地，我们可以得到关于 $\mathbf{y}$ 的类似结果。证明从略，因为其论证与引理 C.1 相同。

引理 C.2. 对于任意整数 $i \in \{1, \dots, n\}$，令

$$v^* \;\in\; \arg\max_{v\in[0,1]} F\big(\mathbf{y}(i-1) - v\,\mathbf{e}_{u_i}\big). \tag{164}$$

若 $v^* \ge \frac{1}{2}$，则

$$F\big(\mathbf{y}(i-1) - b_i\,\mathbf{e}_{u_i}\big) \;\ge\; \max_{v\in[0,1]} F\big(\mathbf{y}(i-1) - v\,\mathbf{e}_{u_i}\big) \;-\; \frac{2\varepsilon}{n}\,\gamma^2\,F(\mathbf{o}), \tag{165}$$

并且若 $v^* < \frac{1}{2}$，则

$$F\big(\mathbf{y}(i-1) - b_i\,\mathbf{e}_{u_i}\big) \;\ge\; \max_{v\in[0,1]} F\big(\mathbf{y}(i-1) - v\,\mathbf{e}_{u_i}\big) \;-\; \frac{2\varepsilon}{\gamma^2 n}\,F(\mathbf{o}). \tag{166}$$

在每次迭代中，$\gamma$感知混合步骤（算法 4 的第 8 行）保证目标函数有可量化的提升。下面的引理为两条轨迹的逐坐标进展给出下界，表明增益是一个随$\gamma$缩放的凸组合型二次项。

引理 C.3. 对每个整数 $1 \le i \le n$,

$$F\big(\mathbf{x}(i)\big) - F\big(\mathbf{x}(i-1)\big) \;\ge\; \frac{\Delta_{a,i}^2}{\Delta_{a,i} + \gamma^3\Delta_{b,i}} \tag{167}$$

*and*

$$F\big(\mathbf{y}(i)\big) - F\big(\mathbf{y}(i-1)\big) \;\ge\; \frac{\gamma^3\Delta_{b,i}^2}{\Delta_{a,i} + \gamma^3\Delta_{b,i}}. \tag{168}$$

证明. 沿 $\mathbf{x}$ 的 $F$ 的增加。根据 $\mathbf{x}(i)$ 的定义以及

$$w_i \;=\; \frac{\Delta_{a,i}\,a_i + \gamma\,\Delta_{b,i}\,(1-b_i)}{\Delta_{a,i} + \gamma\,\Delta_{b,i}}, \tag{169}$$

we c可以将新点 $\mathbf{x}(i)$ 写成两个一维更新的凸组合 es:

$$\mathbf{x}(i) \;=\; \mathbf{x}(i-1) \;+\; \frac{\Delta_{a,i}}{\Delta_{a,i} + \gamma\Delta_{b,i}}\,a_i\,\mathbf{e}_{u_i} \;+\; \frac{\gamma\Delta_{b,i}}{\Delta_{a,i} + \gamma\Delta_{b,i}}\,(1-b_i)\,\mathbf{e}_{u_i}. \tag{170}$$

等价地，$\mathbf{x}(i)$ 是以下凸组合

$$\mathbf{x}(i) \;=\; \frac{\Delta_{a,i}}{\Delta_{a,i} + \gamma\Delta_{b,i}}\big(\mathbf{x}(i-1) + a_i\,\mathbf{e}_{u_i}\big) \;+\; \frac{\gamma\Delta_{b,i}}{\Delta_{a,i} + \gamma\Delta_{b,i}}\big(\mathbf{x}(i-1) + (1-b_i)\,\mathbf{e}_{u_i}\big). \tag{171}$$

因此

$$F\big(\mathbf{x}(i)\big) - F\big(\mathbf{x}(i-1)\big) = F\Big(\frac{\Delta_{a,i}}{\Delta_{a,i}+\gamma\Delta_{b,i}}\big(\mathbf{x}(i-1)+a_i\,\mathbf{e}_{u_i}\big) + \frac{\gamma\Delta_{b,i}}{\Delta_{a,i}+\gamma\Delta_{b,i}}\big(\mathbf{x}(i-1)+(1-b_i)\,\mathbf{e}_{u_i}\big)\Big) - F\big(\mathbf{x}(i-1)\big). \tag{172}$$

现在我们将引理2.1 （一维 $\gamma$–弱 DR 凸性型界）应用于这一对

$$\mathbf{z}^{(1)} = \mathbf{x}(i-1) + a_i\,\mathbf{e}_{u_i}, \qquad \mathbf{z}^{(2)} = \mathbf{x}(i-1) + (1-b_i)\,\mathbf{e}_{u_i},$$

采用混合权重

$$\lambda \;=\; \frac{\gamma\Delta_{b,i}}{\Delta_{a,i}+\gamma\Delta_{b,i}}, \qquad 1-\lambda \;=\; \frac{\Delta_{a,i}}{\Delta_{a,i}+\gamma\Delta_{b,i}}. \tag{173}$$

引理2.1 指出，对于这样的凸组合，我们有

$$F\big((1-\lambda)\mathbf{z}^{(1)} + \lambda\mathbf{z}^{(2)}\big) \;\geq\; \frac{(1-\lambda)\,F(\mathbf{z}^{(1)}) + \gamma^2\lambda\,F(\mathbf{z}^{(2)})}{(1-\lambda)+\gamma^2\lambda}. \tag{174}$$

将（173）代入（174），并注意到

$$(1-\lambda)+\gamma^2\lambda \;=\; \frac{\Delta_{a,i}+\gamma^3\Delta_{b,i}}{\Delta_{a,i}+\gamma\Delta_{b,i}},$$

我们得到

$$F\big(\mathbf{x}(i)\big) \geq \frac{\Delta_{a,i}\,F\big(\mathbf{x}(i-1)+a_i\,\mathbf{e}_{u_i}\big) + \gamma^3\Delta_{b,i}\,F\big(\mathbf{x}(i-1)+(1-b_i)\,\mathbf{e}_{u_i}\big)}{\Delta_{a,i}+\gamma^3\Delta_{b,i}}. \tag{175}$$

从（ $F\big(\mathbf{x}(i-1)\big)$ 的两边减去175），并合并各项，得到

$$F\big(\mathbf{x}(i)\big) - F\big(\mathbf{x}(i-1)\big) \geq \frac{\Delta_{a,i}\big[F\big(\mathbf{x}(i-1)+a_i\,\mathbf{e}_{u_i}\big) - F\big(\mathbf{x}(i-1)\big)\big] + \gamma^3\Delta_{b,i}\big[F\big(\mathbf{x}(i-1)+(1-b_i)\,\mathbf{e}_{u_i}\big) - F\big(\mathbf{x}(i-1)\big)\big]}{\Delta_{a,i}+\gamma^3\Delta_{b,i}}. \tag{176}$$

这里我们只是在分子中减去 $F(\mathbf{x}(i-1))$，以分离方向增益。

根据$\Delta_{a,i}$的定义，

$$\Delta_{a,i} \;=\; F\big(\mathbf{x}(i-1)+a_i\,\mathbf{e}_{u_i}\big) - F\big(\mathbf{x}(i-1)\big), \tag{177}$$

因此，(176) 分子中的第一项恰好是 $\Delta_{a,i}^2$。对于第二项，我们利用$\gamma$–弱 DR 性质，比较在 $\mathbf{x}(i-1)$ 处的增益与在 $\mathbf{y}(i-1)$ 处的对应增益。沿着第 $u_i$ 个坐标，弱 DR 性质意味着：在 $\mathbf{y}(i-1)$ 处，从 $\mathbf{1}$ 降至 $b_i$ 时的边际减量，至少是 $\mathbf{x}(i-1)$ 处对应边际减量的 $\gamma^2$ 倍。由此得到

$$F\big(\mathbf{x}(i-1)+(1-b_i)\,\mathbf{e}_{u_i}\big) - F\big(\mathbf{x}(i-1)\big) \;\geq\; \gamma\Big[F\big(\mathbf{y}(i-1)-b_i\,\mathbf{e}_{u_i}\big) - F\big(\mathbf{y}(i-1)-\mathbf{e}_{u_i}\big)\Big]. \tag{178}$$

将 (178) 乘以 $\gamma^3\Delta_{b,i}$，分子中便得到一个 $\gamma^4$ 倍因子。将 (177) 和 (178) 代入 (176，可得

$$F\big(\mathbf{x}(i)\big) - F\big(\mathbf{x}(i-1)\big) \geq \frac{\Delta_{a,i}^2 + \gamma^4\Delta_{b,i}\Big[F\big(\mathbf{y}(i-1)-b_i\,\mathbf{e}_{u_i}\big) - F\big(\mathbf{y}(i-1)-\mathbf{e}_{u_i}\big)\Big]}{\Delta_{a,i}+\gamma^3\Delta_{b,i}}. \tag{179}$$

根据 $b_i$ 在算法 4中的定义，表达式

$$F\big(\mathbf{y}(i-1)-b_i\,\mathbf{e}_{u_i}\big) - F\big(\mathbf{y}(i-1)-\mathbf{e}_{u_i}\big)$$

为非负，或者在最坏情况下，不超过网格上的相应候选值。特别是，$\gamma^4\Delta_{b,\,i}$ 在（179）中相乘的项为非负，因此我们可以将其舍去，从而得到更简单的界

$$F\big(\mathbf{x}(i)\big) - F\big(\mathbf{x}(i-1)\big) \;\geq\; \frac{\Delta_{a,i}^2}{\Delta_{a,i} + \gamma^3 \Delta_{b,i}}, \tag{180}$$

这正是（167）。

沿 y 增加 $F$。关于 y 的论证是对称的。根据 y($i$) 的更新规则以及相同的权重 $w_i$，可得

$$\mathbf{y}(i) \;=\; \frac{\Delta_{a,i}}{\Delta_{a,i} + \gamma\Delta_{b,i}}\big(\mathbf{y}(i-1) + (a_i - 1)\,\mathbf{e}_{u_i}\big) \;+\; \frac{\gamma\Delta_{b,i}}{\Delta_{a,i} + \gamma\Delta_{b,i}}\big(\mathbf{y}(i-1) - b_i\,\mathbf{e}_{u_i}\big). \tag{181}$$

因此

$$F\big(\mathbf{y}(i)\big) - F\big(\mathbf{y}(i-1)\big) = F\left(\frac{\Delta_{a,i}}{\Delta_{a,i} + \gamma\Delta_{b,i}}\big(\mathbf{y}(i-1) + (a_i - 1)\,\mathbf{e}_{u_i}\big) + \frac{\gamma\Delta_{b,i}}{\Delta_{a,i} + \gamma\Delta_{b,i}}\big(\mathbf{y}(i-1) - b_i\,\mathbf{e}_{u_i}\big)\right) - F\big(\mathbf{y}(i-1)\big). \tag{182}$$

将引理 2.1 应用于这一对

$$\tilde{\mathbf{z}}^{(1)} = \mathbf{y}(i-1) + (a_i - 1)\,\mathbf{e}_{u_i}, \qquad \tilde{\mathbf{z}}^{(2)} = \mathbf{y}(i-1) - b_i\,\mathbf{e}_{u_i},$$

使用与（173）中相同的权重，可得

$$F\big(\mathbf{y}(i)\big) \;\geq\; \frac{\Delta_{a,i}\,F\big(\mathbf{y}(i-1) + (a_i - 1)\,\mathbf{e}_{u_i}\big) + \gamma^3\Delta_{b,i}\,F\big(\mathbf{y}(i-1) - b_i\,\mathbf{e}_{u_i}\big)}{\Delta_{a,i} + \gamma^3\Delta_{b,i}}. \tag{183}$$

减去 $F$(y($i-1$)) 并重新分组，得到

$$F\big(\mathbf{y}(i)\big) - F\big(\mathbf{y}(i-1)\big) \geq \frac{\Delta_{a,i}\big[F\big(\mathbf{y}(i-1) + (a_i - 1)\,\mathbf{e}_{u_i}\big) - F\big(\mathbf{y}(i-1)\big)\big] + \gamma^3\Delta_{b,i}\big[F\big(\mathbf{y}(i-1) - b_i\,\mathbf{e}_{u_i}\big) - F\big(\mathbf{y}(i-1)\big)\big]}{\Delta_{a,i} + \gamma^3\Delta_{b,i}}. \tag{184}$$

利用 $\gamma$–弱 DR 性质，沿第 $u_i$ 个坐标比较 y($i-1$) 处的截面与 x($i-1$) 处的截面，可得

$$F\big(\mathbf{y}(i-1) + (a_i - 1)\,\mathbf{e}_{u_i}\big) - F\big(\mathbf{y}(i-1)\big) \;\geq\; \frac{1}{\gamma}\Big[F\big(\mathbf{x}(i-1) + a_i\,\mathbf{e}_{u_i}\big) - F\big(\mathbf{x}(i-1) + \mathbf{e}_{u_i}\big)\Big]. \tag{185}$$

根据 $\Delta_{b,i}$ 的定义，

$$\Delta_{b,i} \;=\; F\big(\mathbf{y}(i-1) - b_i\,\mathbf{e}_{u_i}\big) - F\big(\mathbf{y}(i-1)\big), \tag{186}$$

因此，分子中 184 式的第二项为 $\gamma^3\Delta_{b,i}^2$。将 185 式和 186 式代入 184 式，得到

$$F\big(\mathbf{y}(i)\big) - F\big(\mathbf{y}(i-1)\big) \geq \frac{\frac{1}{\gamma}\Delta_{a,i}\Big[F\big(\mathbf{x}(i-1) + a_i\,\mathbf{e}_{u_i}\big) - F\big(\mathbf{x}(i-1) + \mathbf{e}_{u_i}\big)\Big] + \gamma^3\Delta_{b,i}^2}{\Delta_{a,i} + \gamma^3\Delta_{b,i}}. \tag{187}$$

根据 $a_i$ 在算法 4 中的选择，项

$$F\big(\mathbf{x}(i-1) + a_i\,\mathbf{e}_{u_i}\big) - F\big(\mathbf{x}(i-1) + \mathbf{e}_{u_i}\big)$$

为非正（因为 $a_i$ 与单位步长进行最优权衡），因此 187 式分子中的第一项为非正。去掉这一非正项，可得更简单的下界

$$F\big(\mathbf{y}(i)\big) - F\big(\mathbf{y}(i-1)\big) \;\geq\; \frac{\gamma^3\Delta_{b,i}^2}{\Delta_{a,i} + \gamma^3\Delta_{b,i}}, \tag{188}$$

这恰好是（168）。结合（180）和（188），即可完成证明。 □

参考路径及其收缩。为了将两条轨迹 $\mathbf{x}(i)$ 和 $\mathbf{y}(i)$ 与任意初始向量 $\mathbf{o}$ 联系起来，我们引入标准的格耦合参考序列

$$\mathbf{o}^{(i)} \;:=\; \big(\,\mathbf{o} \vee \mathbf{x}(i)\,\big) \;\wedge\; \mathbf{y}(i) \qquad \text{for } i = 0, 1, \ldots, n. \tag{189}$$

注意到 $\mathbf{o}^{(0)} = \mathbf{o}$ 和 $\mathbf{o}^{(n)} = \mathbf{x}(n) = \mathbf{y}(n)$，因为算法 4 在结束时使两条轨迹相等。下一个引理给出了 $F(\mathbf{o}^{(i)})$ 的单步下降量上界；将此上界沿 $i$ 累加，即可得到最终保证中使用的全局比较。

引理 C.4. 对每个整数 $1 \le i \le n$,

$$F\big(\mathbf{o}^{(i-1)}\big) - F\big(\mathbf{o}^{(i)}\big) \;\le\; \frac{\Delta_{a,i}\,\Delta_{b,i}}{\Delta_{a,i} + \gamma^2 \Delta_{b,i}} \;+\; \frac{2\varepsilon}{n}\,\gamma^3\,F(\mathbf{o}). \tag{190}$$

证明。我们先考虑以下情形

$$\mathbf{o}_{u_i}^{(i-1)} \;\le\; \mathbf{o}_{u_i}^{(i)}. \tag{191}$$

定义单坐标设置器

$$\mathrm{set}_u(\mathbf{z}, t) \;:=\; \mathbf{z} - (\mathbf{z}_u - t)\,\mathbf{e}_u, \tag{192}$$

其中 将 $\mathbf{z}$ 的第 $u$ 个坐标替换为 $t$，并保持所有其他坐标不变。

根据 $\gamma$–弱 DR 次模性，对于任意 $\mathbf{p} \le \mathbf{q}$ 和任意标量 $\alpha \le \beta$，有如下单维比较

$$F\big(\mathrm{set}_u(\mathbf{p}, \beta)\big) - F\big(\mathrm{set}_u(\mathbf{p}, \alpha)\big) \;\ge\; \gamma\Big[F\big(\mathrm{set}_u(\mathbf{q}, \beta)\big) - F\big(\mathrm{set}_u(\mathbf{q}, \alpha)\big)\Big]. \tag{193}$$

这里，左侧表示在较低点 $\mathbf{p}$ 处将坐标 $u$ 从 $\alpha$ 改为 $\beta$ 所获得的增益；右侧则将其与较高点 $\mathbf{q}$ 处对应的增益进行比较，并按 $\gamma$ 缩放。

根据 $\mathbf{o}^{(i-1)}$ 和 $\mathbf{y}(i-1)$ 在 (189) 中的构造可知,

$$\mathbf{o}^{(i-1)} \;\le\; \mathbf{y}(i-1). \tag{194}$$

在当前情形下，我们还有 (191)。令

$$\mathbf{p} = \mathbf{o}^{(i-1)}, \quad \mathbf{q} = \mathbf{y}(i-1), \quad u = u_i, \quad \alpha = \mathbf{o}_{u_i}^{(i-1)}, \quad \beta = \mathbf{o}_{u_i}^{(i)}, \tag{195}$$

它们满足 (193) 的要求。于是，由 (193) 可得

$$\begin{aligned} F\big(\mathrm{set}_{u_i}(\mathbf{o}^{(i-1)}, \mathbf{o}_{u_i}^{(i)})\big) - F\big(\mathrm{set}_{u_i}(\mathbf{o}^{(i-1)}, \mathbf{o}_{u_i}^{(i-1)})\big) \\ \ge\; \gamma\Big[F\big(\mathrm{set}_{u_i}(\mathbf{y}^{(i-1)}, \mathbf{o}_{u_i}^{(i)})\big) - F\big(\mathrm{set}_{u_i}(\mathbf{y}^{(i-1)}, \mathbf{o}_{u_i}^{(i-1)})\big)\Big]. \end{aligned} \tag{196}$$

根据参考路径的定义 (189) 以及 setter 的定义 (192),

$$\mathrm{set}_{u_i}(\mathbf{o}^{(i-1)}, \mathbf{o}_{u_i}^{(i)}) = \mathbf{o}^{(i)}, \qquad \mathrm{set}_{u_i}(\mathbf{o}^{(i-1)}, \mathbf{o}_{u_i}^{(i-1)}) = \mathbf{o}^{(i-1)}. \tag{197}$$

利用（197）代入（196），左侧变为 $F(\mathbf{o}^{(i)}) - F(\mathbf{o}^{(i-1)})$，因此得到

$$F\big(\mathbf{o}^{(i)}\big) - F\big(\mathbf{o}^{(i-1)}\big) \;\ge\; \gamma\Big[F\big(\mathrm{set}_{u_i}(\mathbf{y}^{(i-1)}, \mathbf{o}_{u_i}^{(i)})\big) - F\big(\mathrm{set}_{u_i}(\mathbf{y}^{(i-1)}, \mathbf{o}_{u_i}^{(i-1)})\big)\Big]. \tag{198}$$

利用显式形式（192），对于任意 $\mathbf{z}$ $t$ 我们有集合$u_i(\mathbf{z}\ t) = \mathbf{z} - (\mathbf{z}_{u_i} - t)\mathbf{e}^{u_i}$。因此

$$\mathrm{set}_{u_i}(\mathbf{y}^{(i-1)}, \mathbf{o}_{u_i}^{(i)}) = \mathbf{y}^{(i-1)} - (\mathbf{y}_{u_i}^{(i-1)} - \mathbf{o}_{u_i}^{(i)})\,\mathbf{e}_{u_i} \tag{199}$$

and

$$\mathrm{set}_{u_i}(\mathbf{y}^{(i-1)}, \mathbf{o}_{u_i}^{(i-1)}) = \mathbf{y}^{(i-1)} - (\mathbf{y}_{u_i}^{(i-1)} - \mathbf{o}_{u_i}^{(i-1)})\,\mathbf{e}_{u_i}. \tag{200}$$

由于 $\mathbf{o}^{(i)}$ 和 $\mathbf{o}^{(i-1)}$ 除可能的 $u_i$ 外共享所有坐标，且 $\mathbf{y}_{u_i}^{(i-1)} \in [0,1]$，量 $\mathbf{y}_{u_i}^{(i-1)} - \mathbf{o}_{u_i}^{(i)}$和 $\mathbf{y}_{u_i}^{(i-1)} - \mathbf{o}_{u_i}^{(i-1)}$属于 $[0,1]$。为方便记号，记

$$\begin{aligned} 1 - \mathbf{o}_{u_i}^{(i)} &= \mathbf{y}_{u_i}^{(i-1)} - \mathbf{o}_{u_i}^{(i)}, \\ 1 - \mathbf{o}_{u_i}^{(i-1)} &= \mathbf{y}_{u_i}^{(i-1)} - \mathbf{o}_{u_i}^{(i-1)}, \end{aligned} \tag{201}$$

因此我们可以将 (199)–(200) 改写为

$$\mathrm{set}_{u_i}(\mathbf{y}^{(i-1)}, \mathbf{o}_{u_i}^{(i)}) = \mathbf{y}^{(i-1)} - (1 - \mathbf{o}_{u_i}^{(i)})\,\mathbf{e}_{u_i}, \qquad \mathrm{set}_{u_i}(\mathbf{y}^{(i-1)}, \mathbf{o}_{u_i}^{(i-1)}) = \mathbf{y}^{(i-1)} - (1 - \mathbf{o}_{u_i}^{(i-1)})\,\mathbf{e}_{u_i}. \tag{202}$$

将 (202代入 (198，得到

$$F\big(\mathbf{o}^{(i)}\big) - F\big(\mathbf{o}^{(i-1)}\big) \;\geq\; \gamma\Big[F\big(\mathbf{y}(i-1) - (1 - \mathbf{o}_{u_i}^{(i)})\,\mathbf{e}_{u_i}\big) - F\big(\mathbf{y}(i-1) - (1 - \mathbf{o}_{u_i}^{(i-1)})\,\mathbf{e}_{u_i}\big)\Big]. \tag{203}$$

现在我们对右侧的每一项进行界定。

首先，根据引理C.2，对于沿着 $u_i$-th 个坐标的序列，

$$\max_{v \in [0,1]} F\big(\mathbf{y}(i-1) - v\,\mathbf{e}_{u_i}\big) \;\leq\; F\big(\mathbf{y}(i-1) - b_i\,\mathbf{e}_{u_i}\big) \;+\; \frac{2\varepsilon}{n}\,\gamma^2\,F(\mathbf{o}). \tag{204}$$

由于 $1 - \mathbf{o}_{u_i}^{(i-1)} \in [0,1]$，我们有

$$\begin{aligned} F\big(\mathbf{y}(i-1) - (1 - \mathbf{o}_{u_i}^{(i-1)})\,\mathbf{e}_{u_i}\big) \;&\leq\; \max_{v \in [0,1]} F\big(\mathbf{y}(i-1) - v\,\mathbf{e}_{u_i}\big) \\ &\leq\; F\big(\mathbf{y}(i-1) - b_i\,\mathbf{e}_{u_i}\big) \;+\; \frac{2\varepsilon}{n}\,\gamma^2\,F(\mathbf{o}). \end{aligned} \tag{205}$$

接下来，回忆一下参考点的第 $u_i$ 个坐标满足

$$\mathbf{o}_{u_i}^{(i)} = \frac{\Delta_{a,i}}{\Delta_{a,i} + \Delta_{b,i}}\,a_i + \frac{\Delta_{b,i}}{\Delta_{a,i} + \Delta_{b,i}}\,(1 - b_i), \tag{206}$$

根据算法 4中坐标更新的定义。沿着$u_i$-th 坐标，在点 $\mathbf{y}(i-1)$ 处应用引理

$$(1 - a_i) \quad \text{and} \quad b_i,$$

并采用与 $\Delta_{a,i}$ 和 $\Delta_{b,i}$ 成比例的混合权重，可得

$$\begin{aligned} &F\big(\mathbf{y}(i-1) - (1 - \mathbf{o}_{u_i}^{(i)})\,\mathbf{e}_{u_i}\big) \\ &\quad\geq \frac{\Delta_{a,i}}{\Delta_{a,i} + \gamma^2\Delta_{b,i}}\,F\big(\mathbf{y}(i-1) - (1 - a_i)\,\mathbf{e}_{u_i}\big) \;+\; \frac{\gamma^2\Delta_{b,i}}{\Delta_{a,i} + \gamma^2\Delta_{b,i}}\,F\big(\mathbf{y}(i-1) - b_i\,\mathbf{e}_{u_i}\big). \end{aligned} \tag{207}$$

结合式（203）、（205）和（207），可得

$$\begin{aligned} &F\big(\mathbf{o}^{(i)}\big) - F\big(\mathbf{o}^{(i-1)}\big) \\ &\quad\geq \gamma\left[\frac{\Delta_{a,i}}{\Delta_{a,i} + \gamma^2\Delta_{b,i}}\Big(F\big(\mathbf{y}(i-1) - (1 - a_i)\,\mathbf{e}_{u_i}\big) - F\big(\mathbf{y}(i-1) - b_i\,\mathbf{e}_{u_i}\big)\Big) - \frac{2\varepsilon}{n}\,\gamma^2\,F(\mathbf{o})\right]. \end{aligned} \tag{208}$$

方括号内的第一项是坐标为 $1 - a_i$ 和 $b_i$ 的两点之间的“真实”方向增益；第二项来自引理 C.2 中的离散化损失。

通过加上并减去 $F(\mathbf{y}(i-1))$，并利用定义，重写式（208）中方括号内的第一项

$$\Delta_{b,i} = F\big(\mathbf{y}(i-1) - b_i\,\mathbf{e}_{u_i}\big) - F\big(\mathbf{y}(i-1)\big), \tag{209}$$

以获得

$$\begin{aligned} &F\big(\mathbf{y}(i-1) - (1 - a_i)\,\mathbf{e}_{u_i}\big) - F\big(\mathbf{y}(i-1) - b_i\,\mathbf{e}_{u_i}\big) \\ &\qquad= \Big(F\big(\mathbf{y}(i-1) - (1 - a_i)\,\mathbf{e}_{u_i}\big) - F\big(\mathbf{y}(i-1)\big)\Big) - \Delta_{b,i}. \end{aligned} \tag{210}$$

将（210）代入（208，得到

$$\begin{aligned} &F\big(\mathbf{o}^{(i)}\big) - F\big(\mathbf{o}^{(i-1)}\big) \\ &\quad \geq \frac{\gamma\,\Delta_{a,i}}{\Delta_{a,i} + \gamma^2 \Delta_{b,i}} \Big( F\big(\mathbf{y}(i-1) - (1-a_i)\,\mathbf{e}_{u_i}\big) - F\big(\mathbf{y}(i-1)\big) - \Delta_{b,i} \Big) - \frac{2\varepsilon}{n}\,\gamma^3\,F(\mathbf{o}). \end{aligned} \tag{211}$$

现在，我们利用 $\gamma$–弱 DR 性质，将这个表达式从 $\mathbf{y}(i-1)$ 转换到 $\mathbf{x}(i-1)$。由于

$$\mathbf{x}(i-1) + a_i\,\mathbf{e}_{u_i} \;\leq\; \mathbf{y}(i-1) - (1-a_i)\,\mathbf{e}_{u_i}, \tag{212}$$

弱 DR 性质意味着，当坐标 $u_i$从 1 向下移动至 $1-a_i$ 时，$\mathbf{y}(i-1)$ 处的边际增益至多为当坐标 $u_i$ 从 0 向上移动至 $a_i$ 时，$\mathbf{x}(i-1)$ 处边际增益的 $1/\gamma$ 倍。形式化地，

$$F\big(\mathbf{y}(i-1) - (1-a_i)\,\mathbf{e}_{u_i}\big) - F\big(\mathbf{y}(i-1)\big) \;\geq\; \frac{1}{\gamma}\Big[F\big(\mathbf{x}(i-1) + a_i\,\mathbf{e}_{u_i}\big) - F\big(\mathbf{x}(i-1) + \mathbf{e}_{u_i}\big)\Big]. \tag{213}$$

将（213）代入（211，可得

$$\begin{aligned} &F\big(\mathbf{o}^{(i)}\big) - F\big(\mathbf{o}^{(i-1)}\big) \\ &\quad \geq \frac{\Delta_{a,i}}{\Delta_{a,i} + \gamma^2 \Delta_{b,i}} \Big( F\big(\mathbf{x}(i-1) + a_i\,\mathbf{e}_{u_i}\big) - F\big(\mathbf{x}(i-1) + \mathbf{e}_{u_i}\big) - \Delta_{b,i} \Big) - \frac{2\varepsilon}{n}\,\gamma^3\,F(\mathbf{o}). \end{aligned} \tag{214}$$

根据算法 $a_i$ 中的定义 4 （并且注意到 $1 \in V$），沿网格选择 $a_i$ 可确保

$$F\big(\mathbf{x}(i-1) + a_i\,\mathbf{e}_{u_i}\big) - F\big(\mathbf{x}(i-1) + \mathbf{e}_{u_i}\big) \;\leq\; 0. \tag{215}$$

因此

$$F\big(\mathbf{x}(i-1) + a_i\,\mathbf{e}_{u_i}\big) - F\big(\mathbf{x}(i-1) + \mathbf{e}_{u_i}\big) - \Delta_{b,i} \;\leq\; -\Delta_{b,i}. \tag{216}$$

将（216）代入（214，得到

$$F\big(\mathbf{o}^{(i)}\big) - F\big(\mathbf{o}^{(i-1)}\big) \;\geq\; -\frac{\Delta_{a,i}\,\Delta_{b,i}}{\Delta_{a,i} + \gamma^2 \Delta_{b,i}} - \frac{2\varepsilon}{n}\,\gamma^3\,F(\mathbf{o}). \tag{217}$$

重新整理（217）得

$$F\big(\mathbf{o}^{(i-1)}\big) - F\big(\mathbf{o}^{(i)}\big) \;\leq\; \frac{\Delta_{a,i}\,\Delta_{b,i}}{\Delta_{a,i} + \gamma^2 \Delta_{b,i}} + \frac{2\varepsilon}{n}\,\gamma^3\,F(\mathbf{o}), \tag{218}$$

这恰好就是情形（190）下的（191）。

剩余情形 $\mathbf{o}_{u_i}^{(i-1)} > \mathbf{o}_{u_i}^{(i)}$ 类似：我们交换 $u_i$-th 坐标上“左”端点和“右”端点的角色，并进行相同的论证，从而得到相同的界（190）。证明至此完成。

□

结合轨迹 $\mathbf{x}(i)$ 和 $\mathbf{y}(i)$ 的进展以及参考路径 $\mathbf{o}^{(i)}$ 的收缩，可得任意权衡参数 $r \geq 0$下的不等式。该不等式将在 $i$ 上逐项相消，从而得出最终保证。

推论 C.5. 对于每个 $r \geq 0$ 和整数 $1 \leq i \leq n$,

$$\begin{aligned} &\frac{1}{r}\left[F(\mathbf{x}(i)) - F(\mathbf{x}(i-1))\right] \;+\; r\left[F(\mathbf{y}(i)) - F(\mathbf{y}(i-1))\right] \\ &\qquad\qquad \geq\; 2\gamma^{3/2}\left(F\big(\mathbf{o}^{(i-1)}\big) - F\big(\mathbf{o}^{(i)}\big) - \frac{2\varepsilon}{n}\,\gamma^3\,F(\mathbf{o})\right). \end{aligned} \tag{219}$$

证明。 根据引理 C.3，对于每个 $i$ 我们有

$$F\big(\mathbf{x}(i)\big) - F\big(\mathbf{x}(i-1)\big) \;\geq\; \frac{\Delta_{a,i}^2}{\Delta_{a,i} + \gamma^3 \Delta_{b,i}}, \qquad F\big(\mathbf{y}(i)\big) - F\big(\mathbf{y}(i-1)\big) \;\geq\; \frac{\gamma^3 \Delta_{b,i}^2}{\Delta_{a,i} + \gamma^3 \Delta_{b,i}}. \tag{220}$$

将第一个不等式 (220) 乘以 $1/r$ ，第二个不等式乘以 $r$，然后将两者相加，得到

$$\frac{1}{r}\big[F(\mathbf{x}(i)) - F(\mathbf{x}(i-1))\big] + r\big[F(\mathbf{y}(i)) - F(\mathbf{y}(i-1))\big] \geq \frac{(1/r)\,\Delta_{a,i}^2}{\Delta_{a,i} + \gamma^3 \Delta_{b,i}} + \frac{r\,\gamma^3\,\Delta_{b,i}^2}{\Delta_{a,i} + \gamma^3 \Delta_{b,i}}. \tag{221}$$

式 (221中的分子可以改写为完全平方与混合项之和 :

$$\frac{(1/r)\,\Delta_{a,i}^2 + r\gamma^3 \Delta_{b,i}^2}{\Delta_{a,i} + \gamma^3 \Delta_{b,i}} = \frac{\left(\frac{\Delta_{a,i}}{\sqrt{r}} - \Delta_{b,i}\,\gamma^{3/2}\sqrt{r}\right)^2 + 2\gamma^{3/2}\,\Delta_{a,i}\Delta_{b,i}}{\Delta_{a,i} + \gamma^3 \Delta_{b,i}}. \tag{222}$$

Sub将式 (222) 代入式 (221)，并利用平方项非负这一性质，可得

$$\frac{1}{r}\big[F(\mathbf{x}(i)) - F(\mathbf{x}(i-1))\big] + r\big[F(\mathbf{y}(i)) - F(\mathbf{y}(i-1))\big] \geq \frac{2\gamma^{3/2}\,\Delta_{a,i}\Delta_{b,i}}{\Delta_{a,i} + \gamma^3 \Delta_{b,i}}. \tag{223}$$

因此，除去因子 $\Delta_{a,\ i}\Delta_{b,\ i}$，每步进展是一个依赖于 $\gamma$ 的凸组合型量。

Ne接下来，我们将 $\Delta_{a、\ i}\Delta_{b、\ i}$ 与参考路径的收缩联系起来。引理 C.4指出 t

$$F\big(\mathbf{o}^{(i-1)}\big) - F\big(\mathbf{o}^{(i)}\big) \;\leq\; \frac{\Delta_{a,i}\,\Delta_{b,i}}{\Delta_{a,i} + \gamma^2 \Delta_{b,i}} \;+\; \frac{2\varepsilon}{n}\,\gamma^3\,F(\mathbf{o}). \tag{224}$$

重新整理式 (224，得到

$$\Delta_{a,i}\Delta_{b,i} \;\geq\; \left(F\big(\mathbf{o}^{(i-1)}\big) - F\big(\mathbf{o}^{(i)}\big) - \frac{2\varepsilon}{n}\,\gamma^3\,F(\mathbf{o})\right)\big(\Delta_{a,i} + \gamma^2 \Delta_{b,i}\big). \tag{225}$$

将（225）代入（223）得到

$$\begin{aligned}\frac{1}{r}\big[F(\mathbf{x}(i)) - F(\mathbf{x}(i-1))\big] + r\big[F(\mathbf{y}(i)) - F(\mathbf{y}(i-1))\big] \\ \geq \frac{2\gamma^{3/2}\big(F(\mathbf{o}^{(i-1)}) - F(\mathbf{o}^{(i)}) - \frac{2\varepsilon}{n}\gamma^3 F(\mathbf{o})\big)\big(\Delta_{a,i} + \gamma^2 \Delta_{b,i}\big)}{\Delta_{a,i} + \gamma^3 \Delta_{b,i}}.\end{aligned} \tag{226}$$

由于 $\gamma \in (0, 1]$, ，我们有 $\gamma^2 \geq \gamma^3$，因此

$$\Delta_{a,i} + \gamma^2 \Delta_{b,i} \;\geq\; \Delta_{a,i} + \gamma^3 \Delta_{b,i}, \tag{227}$$

因此

$$\frac{\Delta_{a,i} + \gamma^2 \Delta_{b,i}}{\Delta_{a,i} + \gamma^3 \Delta_{b,i}} \;\geq\; 1. \tag{228}$$

将（228）应用于（226）可得

$$\begin{aligned}\frac{1}{r}\big[F(\mathbf{x}(i)) - F(\mathbf{x}(i-1))\big] + r\big[F(\mathbf{y}(i)) - F(\mathbf{y}(i-1))\big] \\ \geq 2\gamma^{3/2}\left(F\big(\mathbf{o}^{(i-1)}\big) - F\big(\mathbf{o}^{(i)}\big) - \frac{2\varepsilon}{n}\,\gamma^3\,F(\mathbf{o})\right).\end{aligned} \tag{229}$$

这正是（219），证明完毕。 □

现在我们完成对 $\gamma$–感知型双重贪心过程的分析。下述定理由逐次迭代耦合界与格耦合参考路径的收缩性通过望远镜求和得到。该保证随 $\gamma$ 连续变化，并在 $\gamma = 1$时退化为经典 DR 界。

定理 C.6。存在一种多项式时间算法，给定非负的 $\gamma$-弱 DR 次模函数 $F\colon [0,1]^n \to \mathbb{R}_{\geq 0}$ 以及参数 $\varepsilon \in (0,1)$，输出 $\mathbf{x} \in [0,1]^n$，使得对于每个固定的 $\mathbf{o} \in [0,1]^n$，

$$F(\mathbf{x}) \;\geq\; \max_{r \geq 0} \frac{\left(2\gamma^{3/2} - 4\varepsilon\,\gamma^{9/2}\right) r\,F(\mathbf{o}) \;+\; F(\mathbf{0}) \;+\; r^2\,F(\mathbf{1})}{r^2 \;+\; 2\gamma^{3/2} r \;+\; 1}. \tag{230}$$

证明。 任取 $r > 0$。将推论 C.5 中的逐次迭代不等式对 $i = 1, \ldots, n$求和可得

$$\begin{aligned}
&\frac{1}{r}\big[F(\mathbf{x}(n)) - F(\mathbf{x}(0))\big] \;+\; r\big[F(\mathbf{y}(n)) - F(\mathbf{y}(0))\big] \\
&\qquad = \sum_{i=1}^{n} \left( \frac{1}{r}\big[F(\mathbf{x}(i)) - F(\mathbf{x}(i-1))\big] \;+\; r\big[F(\mathbf{y}(i)) - F(\mathbf{y}(i-1))\big] \right) \\
&\qquad \geq \sum_{i=1}^{n} \left( 2\gamma^{3/2}\left[F(\mathbf{o}^{(i-1)}) - F(\mathbf{o}^{(i)})\right] \;-\; \frac{2\varepsilon}{n} \cdot 2\gamma^{9/2} F(\mathbf{o}) \right) \\
&\qquad = 2\gamma^{3/2}\left[F(\mathbf{o}^{(0)}) - F(\mathbf{o}^{(n)})\right] \;-\; 4\varepsilon\,\gamma^{9/2} F(\mathbf{o}).
\end{aligned} \tag{231}$$

这里，第一个等式只是将 $F(\mathbf{x}^{(i)})$ 和 $F(\mathbf{y}^{(i)})$的增量望远镜求和，而不等式是在每次迭代 $i$中应用推论 C.5 得到的。

根据参考路径的构造，我们有 $\mathbf{o}^{(0)} = \mathbf{o}$，并且

$$\mathbf{x}(n) = \mathbf{y}(n) = \mathbf{o}^{(n)}. \tag{232}$$

此外，算法 4从

$$\mathbf{x}(0) = \mathbf{0}, \qquad \mathbf{y}(0) = \mathbf{1}. \tag{233}$$

在（232）–（233）中代入（231），可得

$$\frac{1}{r}\big[F(\mathbf{x}(n)) - F(\mathbf{0})\big] \;+\; r\big[F(\mathbf{x}(n)) - F(\mathbf{1})\big] \;\geq\; 2\gamma^{3/2}\left[F(\mathbf{o}) - F(\mathbf{x}(n))\right] \;-\; 4\varepsilon\,\gamma^{9/2} F(\mathbf{o}). \tag{234}$$

我们不 我们将所有涉及 $F(\mathbf{x}(n))$ 的项移到等式（234）的左边。整理后得 es

$$F(\mathbf{x}(n)) \left( \frac{1}{r} + r + 2\gamma^{3/2} \right) \;\geq\; \left(2\gamma^{3/2} - 4\varepsilon\,\gamma^{9/2}\right) F(\mathbf{o}) \;+\; \frac{1}{r} F(\mathbf{0}) \;+\; r\,F(\mathbf{1}), \tag{235}$$

其中，右边汇总了 $F(\mathbf{o})$、 $F(\mathbf{0})$ 和 $F(\mathbf{1})$ 的贡献。

将等式（235）两边同时除以 $\frac{1}{r} + r + 2\gamma^{3/2}$，得

$$F(\mathbf{x}(n)) \;\geq\; \frac{\left(2\gamma^{3/2} - 4\varepsilon\,\gamma^{9/2}\right) F(\mathbf{o}) \;+\; \frac{1}{r} F(\mathbf{0}) \;+\; r\,F(\mathbf{1})}{\frac{1}{r} + r + 2\gamma^{3/2}}. \tag{236}$$

将等式（236）右边的分子和分母同时乘以 $r$，得

$$F(\mathbf{x}(n)) \;\geq\; \frac{\left(2\gamma^{3/2} - 4\varepsilon\,\gamma^{9/2}\right) r\,F(\mathbf{o}) \;+\; F(\mathbf{0}) \;+\; r^2\,F(\mathbf{1})}{r^2 + 2\gamma^{3/2} r + 1}. \tag{237}$$

由于$\mathbf{x} = \mathbf{x}^{(n)}$ 是算法的输出，不等式（237）对 $r > 0$的任意取值都成立。利用右端关于 $r$ 的连续性，将其扩展到 $r = 0$，并对 $r \geq 0$取最大值，即得到（230），证明完毕。

□

## D 引理的证明 4.1 和定理 4.2

本节首先证明引理4.1，然后证明定理 4.2

**引理 4.1.** 设 $F:\ [0,1]^n \to \mathbb{R}_{\geq 0}$ 是非负且 $\gamma$-弱 DR 次模的某个 $0 < \gamma \leq 1$，并设 $P \subseteq [0,1]^n$ 向下封闭。存在一个常数规模的三元组集合 $\mathcal{G} \subseteq \mathbb{R}^3_{\geq 0}$（其规模仅取决于 $\varepsilon$ 和 $\gamma$），使得 $\mathcal{G}$ 中包含一个三元组 $(g, g_\odot, g_\oplus)$，满足

$$(1-\varepsilon)\,F(\mathbf{o}) \leq g \leq F(\mathbf{o}), \tag{238a}$$

$$F(\mathbf{z}\odot\mathbf{o}) - \varepsilon\, g \leq g_\odot \leq F(\mathbf{z}\odot\mathbf{o}), \tag{238b}$$

$$F(\mathbf{z}\oplus\mathbf{o}) - \varepsilon\, g \leq g_\oplus \leq F(\mathbf{z}\oplus\mathbf{o}). \tag{238c}$$

证明。假设我们有一个常数因子估计 $v$，满足

$$c\,F(\mathbf{o}) \ \leq\ v \ \leq\ F(\mathbf{o}) \tag{239}$$

对于s 某个绝对常数 $c \in (0,1]$）。我们将使用 $v$ 构造一个常数规模的猜测集合 .
定义一维猜测集合

$$G_o \ :=\ \Big\{(1-\varepsilon)^i \cdot \frac{v}{c} \ :\ i = 0,1,\ldots,\lceil \log_{1-\varepsilon} c \rceil\Big\}. \tag{240}$$

其大小为 $G_o$，即 $|G_o| = \mathcal{O}_\varepsilon(1)$，因为（240）中的指数范围仅取决于 $\varepsilon$ 和 $c$。根据构造，$G_o$ 中的值构成一个几何网格，该网格以 $\varepsilon$-覆盖区间 $[F(\mathbf{o})\ F(\mathbf{o})/c]$，因此存在

$$g \in G_o \quad \text{with} \quad (1-\varepsilon)\,F(\mathbf{o}) \ \leq\ g \ \leq\ F(\mathbf{o}), \tag{241}$$

这证明了（238a）；关于这一标准论证，参见例如 [5] 。

接下来，我们根据 $F(\mathbf{o})$ 给出 $F(\mathbf{z}\odot\mathbf{o})$ 和 $F(\mathbf{z}\oplus\mathbf{o})$ 可能取值范围的上界。由于 $\mathbf{z}\odot\mathbf{o} \leq \mathbf{o}$ 且 $F$ 非负，我们总有

$$0 \ \leq\ F(\mathbf{z}\odot\mathbf{o}) \ \leq\ F(\mathbf{o}). \tag{242}$$

如果 $F$ 是单调的，那么（242）显然成立；否则，我们只使用平凡的非负性上界。

对于 $\oplus$ 操作，利用 $\gamma$-弱 DR 次模性和非负性，我们可以得到上界：

$$\begin{aligned}
F(\mathbf{z}\oplus\mathbf{o}) &= F\big(\mathbf{o} + (\mathbf{1}-\mathbf{o})\odot\mathbf{z}\big) \\
&\leq F(\mathbf{o}) \ +\ \frac{1}{\gamma}\Big[F\big((\mathbf{1}-\mathbf{o})\odot\mathbf{z}\big) - F(\mathbf{0})\Big] && \text{(by } \gamma\text{–weakly DR along } (\mathbf{1}-\mathbf{o})\odot\mathbf{z}) \\
&\leq F(\mathbf{o}) \ +\ \frac{1}{\gamma}F(\mathbf{o}) && \text{(since } F((\mathbf{1}-\mathbf{o})\odot\mathbf{z}) \leq F(\mathbf{o}) \text{ and } F(\mathbf{0}) \geq 0) \\
&= \Big(1+\tfrac{1}{\gamma}\Big)\,F(\mathbf{o}).
\end{aligned} \tag{243}$$

因此，$F(\mathbf{z}\odot\mathbf{o})$ 和 $F(\mathbf{z}\oplus\mathbf{o})$ 的取值范围均受 $F(\mathbf{o})$ 线性有界 .
对于任意满足条件（ $g \in G_o$ 241）的，现在我们为这些范围构造 $\varepsilon g$-网。对于$\odot$-情形，定义

$$G_\odot(g) \ :=\ \Big\{\varepsilon\, i \cdot g \ :\ i = 0,1,\ldots,\ \Big\lceil \tfrac{1}{\varepsilon(1-\varepsilon)} \Big\rceil\Big\}. \tag{244}$$

由于（242）和（241）推出

$$0 \ \leq\ F(\mathbf{z}\odot\mathbf{o}) \ \leq\ F(\mathbf{o}) \ \leq\ \frac{g}{1-\varepsilon}, \tag{245}$$

(244) 中的网格 $\varepsilon g$-覆盖区间 $[0, F(\mathbf{z}\odot\mathbf{o})]$: 对于任意值 $x \in [0, F(\mathbf{z}\odot\mathbf{o})],$，都存在某个 $g_\odot \in G_\odot(g)$ 使得

$$x - \varepsilon g \ \leq\ g_\odot \ \leq\ x. \tag{246}$$

特别地，我们可以选择 $g_\odot$ 使得（246）在 $x = F(\mathbf{z} \odot \mathbf{o})$ 下成立，这恰好就是（238b）。类似地，对于 $\oplus$-情形，定义

$$G_\oplus(g) \;:=\; \Big\{\varepsilon\, i \cdot g \;:\; i = 0, 1, \ldots, \Big\lceil \tfrac{1+1/\gamma}{\varepsilon(1-\varepsilon)} \Big\rceil \Big\}. \tag{247}$$

利用 (243) 和 (241)，可得

$$0 \;\le\; F(\mathbf{z} \oplus \mathbf{o}) \;\le\; \Big(1 + \frac{1}{\gamma}\Big) F(\mathbf{o}) \;\le\; \frac{1 + 1/\gamma}{1 - \varepsilon}\, g. \tag{248}$$

因此，(247) 中的网格 $\varepsilon g$覆盖区间 $[0, F(\mathbf{z} \oplus \mathbf{o})],$，并且存在 $g_\oplus \in G_\oplus(g)$，使得

$$F(\mathbf{z} \oplus \mathbf{o}) - \varepsilon g \;\le\; g_\oplus \;\le\; F(\mathbf{z} \oplus \mathbf{o}), \tag{249}$$

即 (238c)。

最后，令

$$\mathcal{G} \;:=\; \bigcup_{g \in G_o} \{g\} \times G_\odot(g) \times G_\oplus(g). \tag{250}$$

由 (241)、(246) 和 (249)，存在三元组 $(g,\ g_\odot,\ g_\oplus) \in \mathcal{G}$，满足 (238) 中的全部三个界。此外，$|\mathcal{G}|$ 仅通过 $\varepsilon$ 和 $\gamma$ 的基数依赖于 $G_o$、$G_\odot(g)$ 和 $G_\oplus(g)$，因此$\mathcal{G}$ 具有常数大小。

□

为了证明定理 4.2，我们现在回顾迭代点 $\{\mathbf{y}(i)\}_{i=0}^{\delta^{-1}}$ 的闭式 [5],以及终端迭代的可行性。这些公式将用于将$F(\mathbf{y}(\delta^{-1}))$ 与基准值 $F(\mathbf{o})$ 联系起来。

引理 D.1（$\mathbf{y}(i)$ [5] 的闭式）。对于每个整数 $0 \le i \le \delta^{-1}$,

$$\mathbf{y}(i) = \begin{cases} (\mathbf{1} - \mathbf{z}) \;\odot\; \displaystyle\bigoplus_{j=1}^{i} (\delta\, \mathbf{x}(j)), & i \le i_s, \\[2ex] (\mathbf{1} - \mathbf{z}) \;\odot\; \displaystyle\bigoplus_{j=1}^{i} (\delta\, \mathbf{x}(j)) \;+\; \mathbf{z} \;\odot\; \bigoplus_{j=i_s+1}^{i} (\delta\, \mathbf{x}(j)), & i \ge i_s. \end{cases} \tag{251}$$

按照约定，对于任意索引 $a$,

$$\bigoplus_{j=a}^{a-1} (\delta\, \mathbf{x}(j)) \;:=\; \mathbf{0}, \tag{252}$$

因此，(251) 中的两个表达式在其边界索引处均仍然成立。

观察 3 （可行性 [5]）。 终端迭代满足

$$\mathbf{y}(\delta^{-1}) \in P. \tag{253}$$

现在，我们先证明定理 4.2 在以下情形下的结论

$$Q(i) \neq \varnothing \qquad \text{for all } i \in \{1, \ldots, \delta^{-1}\}, \tag{254}$$

其中，两个引理（引理 D.2 和 D.3）给出了定理 4.2 (1) 中所述的界，概括为推论 D.4。另一种情形，即某个 $Q(i) = \varnothing$ 的情形，将在之后单独处理。

观察 4. 若 $Q(i) \neq \varnothing$ 对于某个 $i \in [\delta^{-1}]$，则

$$F\big(\mathbf{y}(i)\big) - F\big(\mathbf{y}(i-1)\big) \;\ge\; \delta\gamma \Big[V(i-1) - F\big(\mathbf{y}(i-1)\big)\Big] \;-\; \frac{\delta^2 L D^2}{2}. \tag{255}$$

证明. 考虑线段

$$\mathbf{u}(s) \;:=\; \mathbf{y}(i-1) \;+\; s\,\Big(\big(\mathbf{1}-\mathbf{y}(i-1)-\mathbf{z}(i-1)\big)\odot\mathbf{x}(i)\Big), \qquad s\in[0,\delta]. \tag{256}$$

根据构造，$\mathbf{u}(0)=\mathbf{y}(i-1)$，且 $\mathbf{u}(\delta)=\mathbf{y}(i)$。

沿着 (256) 应用微积分基本定理，

$$\begin{aligned}
&F\big(\mathbf{y}(i)\big)-F\big(\mathbf{y}(i-1)\big) &(257)\\
&\quad=\int_0^{\delta}\Big\langle\big(\mathbf{1}-\mathbf{y}(i-1)-\mathbf{z}(i-1)\big)\odot\mathbf{x}(i),\;\nabla F\big(\mathbf{u}(s)\big)\Big\rangle\,ds &(258)\\
&\quad=\int_0^{\delta}\Big\langle\big(\mathbf{1}-\mathbf{y}(i-1)-\mathbf{z}(i-1)\big)\odot\mathbf{x}(i),\;\nabla F\big(\mathbf{y}(i-1)\big)\Big\rangle\,ds\\
&\qquad+\int_0^{\delta}\Big\langle\big(\mathbf{1}-\mathbf{y}(i-1)-\mathbf{z}(i-1)\big)\odot\mathbf{x}(i),\;\nabla F\big(\mathbf{u}(s)\big)-\nabla F\big(\mathbf{y}(i-1)\big)\Big\rangle\,ds\\
&\quad=\delta\,\big\langle\mathbf{w}(i),\,\mathbf{x}(i)\big\rangle+\int_0^{\delta}\Big\langle\big(\mathbf{1}-\mathbf{y}(i-1)-\mathbf{z}(i-1)\big)\odot\mathbf{x}(i),\;\nabla F\big(\mathbf{u}(s)\big)-\nabla F\big(\mathbf{y}(i-1)\big)\Big\rangle ds.
\end{aligned}$$

这里我们用到了

$$\mathbf{w}(i):=\big(\mathbf{1}-\mathbf{y}(i-1)-\mathbf{z}(i-1)\big)\odot\nabla F\big(\mathbf{y}(i-1)\big),$$

因此 $\langle\mathbf{w}(i),\;\mathbf{x}(i)\rangle=\big\langle(\mathbf{1}-\mathbf{y}(i-1)\;-\mathbf{z}(i-1))\;\odot\mathbf{x}(i)\;\;\nabla F(\mathbf{y}(i-1))\big\rangle$。

对第二个积分应用柯西–施瓦茨不等式可得

$$\begin{aligned}
F\big(\mathbf{y}(i)\big)-F\big(\mathbf{y}(i-1)\big)\;\ge\;\delta\,\big\langle\mathbf{w}(i),\,\mathbf{x}(i)\big\rangle\\
-\int_0^{\delta}\Big\|\big(\mathbf{1}-\mathbf{y}(i-1)-\mathbf{z}(i-1)\big)\odot\mathbf{x}(i)\Big\|_2\,\Big\|\nabla F\big(\mathbf{u}(s)\big)-\nabla F\big(\mathbf{y}(i-1)\big)\Big\|_2\,ds.
\end{aligned}\tag{259}$$

此步骤用到了 $\langle a,b\rangle\ge-\|a\|_2\|b\|_2$。

由于 $F$ 是 $L$-光滑的，其梯度是 $L$-Lipschitz连续的，因此

$$\big\|\nabla F(\mathbf{u}(s))-\nabla F(\mathbf{y}(i-1))\big\|_2\;\le\;L\,\|\mathbf{u}(s)-\mathbf{y}(i-1)\|_2\;=\;L\,s\,\Big\|\big(\mathbf{1}-\mathbf{y}(i-1)-\mathbf{z}(i-1)\big)\odot\mathbf{x}(i)\Big\|_2.\tag{260}$$

将（260）代入（259）可得

$$\begin{aligned}
F\big(\mathbf{y}(i)\big)-F\big(\mathbf{y}(i-1)\big)\;&\ge\;\delta\,\big\langle\mathbf{w}(i),\,\mathbf{x}(i)\big\rangle-\int_0^{\delta}s\,L\,\Big\|\big(\mathbf{1}-\mathbf{y}(i-1)-\mathbf{z}(i-1)\big)\odot\mathbf{x}(i)\Big\|_2^2\,ds\\
&=\;\delta\,\big\langle\mathbf{w}(i),\,\mathbf{x}(i)\big\rangle-\frac{L}{2}\,\delta^2\,\Big\|\big(\mathbf{1}-\mathbf{y}(i-1)-\mathbf{z}(i-1)\big)\odot\mathbf{x}(i)\Big\|_2^2,
\end{aligned}\tag{261}$$

其中用到了 $\int_0^{\delta}s\,ds=\delta^2/2$。

Let

$$\mathbf{d}(i):=\big(\mathbf{1}-\mathbf{y}(i-1)-\mathbf{z}(i-1)\big)\odot\mathbf{x}(i).$$

根据可行性和逐坐标界 $0\le\mathbf{y}(i-1)$、$\mathbf{z}(i-1)$、$\mathbf{x}(i)\le\mathbf{1}$，$\mathbf{d}(i)$的每个分量都属于 $[0,1]$，，因此

$$\|\mathbf{d}(i)\|_2\;\le\;D,$$

其中 $D$ 是直径。将其代入（261）可得

$$F\big(\mathbf{y}(i)\big)-F\big(\mathbf{y}(i-1)\big)\;\ge\;\delta\,\big\langle\mathbf{w}(i),\,\mathbf{x}(i)\big\rangle\;-\;\frac{\delta^2LD^2}{2}.\tag{262}$$

最后，由于 $Q(i)\neq\varnothing$，选择 $\mathbf{x}(i)\in Q(i)$ 可保证

$$\big\langle\mathbf{w}(i),\,\mathbf{x}(i)\big\rangle\;\ge\;\gamma\Big[V(i-1)-F\big(\mathbf{y}(i-1)\big)\Big],\tag{263}$$

根据 $Q$（$i$）的定义以及弱 DR 结构。将（263）代入（262）可得

$$F\big(\mathbf{y}(i)\big) - F\big(\mathbf{y}(i-1)\big) \;\geq\; \delta\,\gamma\Big[V(i-1) - F\big(\mathbf{y}(i-1)\big)\Big] \;-\; \frac{\delta^2 L D^2}{2},$$

这正是（255）。 □

界（255）为 $F(\mathbf{y}(i))$ 给出了一个递归下界。接下来（直到推论 D.4），我们将展开这一递推关系并推导出一个闭式。

**Le引理 D.2**。假设对于每个 $i \in [i_s]$, $Q(i) \neq \varnothing$。固定 $0 < \delta \leq 1$ 和 $0 < \gamma \leq 1$，并令 $t$

$$\alpha := \delta\gamma, \qquad \beta := \frac{\gamma^2\delta}{1-\delta+\gamma^2\delta}.$$

对于 $i \geq 0$，定义几何简写

$$\Delta_i \;:=\; 1-(1-\alpha)^i, \qquad \Theta_i \;:=\; (1-\beta)^i - (1-\alpha)^i.$$

于是，对于每个整数 $0 \leq i \leq i_s$,

$$F\big(\mathbf{y}(i)\big) \;\geq\; \left[\frac{(1-2\varepsilon)\,\Delta_i}{\gamma} + \frac{\delta(1-\gamma)}{\beta-\alpha}\,\Theta_i\right] g \;-\; \frac{\Delta_i}{\gamma}\,g_\odot \;-\; \left[\frac{\Delta_i}{\gamma} + \frac{\delta}{\beta-\alpha}\,\Theta_i\right] g_\oplus \;-\; i\,\delta^2 D^2 L. \tag{264}$$

证明。对于 $i=0$，有 $\Delta_0 = 0$ 且 $\Theta_0 = 0$，因此（264）的右侧等于 0。由于$F(\mathbf{y}(0)) \geq 0$根据 $F$的非负性，结论对 $i = 0$成立。因此，在证明余下部分时，我们固定一个整数$1 \leq i \leq i_s$。

根据观察 4 （以及假设 $Q(i-1) \neq \varnothing$ 对于 $i \leq i_s$），

$$F\big(\mathbf{y}(i)\big) - F\big(\mathbf{y}(i-1)\big) \;\geq\; \delta\,\gamma\Big(V(i-1) - F\big(\mathbf{y}(i-1)\big)\Big) \;-\; \frac{\delta^2 D^2 L}{2}. \tag{265}$$

重排（265）得

$$F\big(\mathbf{y}(i)\big) \;\geq\; (1-\alpha)\,F\big(\mathbf{y}(i-1)\big) \;+\; \delta\gamma\,V(i-1) \;-\; \frac{\delta^2 D^2 L}{2}, \tag{266}$$

其中我们使用了 $\alpha = \delta\gamma$。

量 $V(i-1)$ 的显式形式如（23）所示

$$V(i-1) = \left[(1-\beta)^{i-1} + \frac{1-(1-\beta)^{i-1}-2\varepsilon}{\gamma}\right] g \;-\; \frac{1}{\gamma}\,g_\odot \;-\; \frac{1-(1-\beta)^{i-1}}{\gamma}\,g_\oplus. \tag{267}$$

现在，将（267）乘以 $\delta\gamma$ 并代入（266）。首先，对于 $g$项，

$$\begin{aligned}
\delta\gamma\left[(1-\beta)^{i-1} + \frac{1-(1-\beta)^{i-1}-2\varepsilon}{\gamma}\right] &= \delta\gamma(1-\beta)^{i-1} + \delta\Big(1-(1-\beta)^{i-1}-2\varepsilon\Big) \\
&= \delta\Big[\gamma(1-\beta)^{i-1} + 1 - (1-\beta)^{i-1} - 2\varepsilon\Big] \\
&= \delta\Big[1-(1-\gamma)(1-\beta)^{i-1} - 2\varepsilon\Big].
\end{aligned} \tag{268}$$

对于 $g_\odot$项，我们得到

$$\delta\gamma\cdot\Big(-\frac{1}{\gamma}\,g_\odot\Big) = -\delta\,g_\odot. \tag{269}$$

对于 $g_\oplus$项，

$$\delta\gamma\cdot\Big(-\frac{1-(1-\beta)^{i-1}}{\gamma}\,g_\oplus\Big) = -\delta\Big(1-(1-\beta)^{i-1}\Big)g_\oplus. \tag{270}$$

将（268）–（270）代入（266）可得

$$F\big(\mathbf{y}(i)\big) \;\geq (1-\alpha)\,F\big(\mathbf{y}(i-1)\big) + \;\delta\Big(1-(1-\gamma)(1-\beta)^{i-1}-2\varepsilon\Big)\,g \\ -\;\delta\,g_{\odot}\;-\;\delta\Big(1-(1-\beta)^{i-1}\Big)\,g_{\oplus}\;-\;\frac{\delta^2 D^2 L}{2}. \tag{271}$$

现在，我们从271开始，将 $k=1$ 到 $k=i$的递推关系展开。为简便起见，记 $F_k := F(\mathbf{y}(k))$，则(271)变为

$$F_k \;\geq\; (1-\alpha)\,F_{k-1} + \delta\,A_{k-1}\,g - \delta\,g_{\odot} - \delta\,B_{k-1}\,g_{\oplus} - \frac{\delta^2 D^2 L}{2}, \tag{272}$$

其中

$$A_{k-1} := 1-(1-\gamma)(1-\beta)^{k-1}-2\varepsilon, \qquad B_{k-1} := 1-(1-\beta)^{k-1}. \tag{273}$$

从 $k=1$ 迭代（272）至 $k=i$ 可得

$$F_i \geq (1-\alpha)^i F_0 + \delta\sum_{k=1}^{i}(1-\alpha)^{i-k}A_{k-1}\,g \\ -\delta\sum_{k=1}^{i}(1-\alpha)^{i-k}g_{\odot} - \delta\sum_{k=1}^{i}(1-\alpha)^{i-k}B_{k-1}\,g_{\oplus} - \frac{\delta^2 D^2 L}{2}\sum_{k=1}^{i}(1-\alpha)^{i-k}. \tag{274}$$

由于 $F_0 = F\left(\mathbf{y}\left(\mathbf{0}\right)\right) \geq 0$ 根据 $F$ 的非负性，第一项 $(1-\alpha)^{i}F_0$ 为非负项，因此在求下界时可以舍去。接下来，我们计算（274）中出现的几何级数和。

首先，

$$\sum_{k=1}^{i}(1-\alpha)^{i-k} = \sum_{t=0}^{i-1}(1-\alpha)^{t} = \frac{1-(1-\alpha)^i}{\alpha} = \frac{\Delta_i}{\alpha}. \tag{275}$$

接下来，对于混合和，使用变量代换 $m=k-1$,

$$\sum_{k=1}^{i}(1-\alpha)^{i-k}(1-\beta)^{k-1} = \sum_{m=0}^{i-1}(1-\alpha)^{i-1-m}(1-\beta)^{m}. \tag{276}$$

这是关于 $m$ 的几何级数。提出公因子 $(1-\alpha)^{i-1}$ 后，我们得到

$$\begin{aligned}\sum_{m=0}^{i-1}(1-\alpha)^{i-1-m}(1-\beta)^{m} &= (1-\alpha)^{i-1}\sum_{m=0}^{i-1}\left(\frac{1-\beta}{1-\alpha}\right)^{m} \\ &= (1-\alpha)^{i-1}\cdot\frac{1-\left(\frac{1-\beta}{1-\alpha}\right)^{i}}{1-\frac{1-\beta}{1-\alpha}} \\ &= (1-\alpha)^{i-1}\cdot\frac{1-\frac{(1-\beta)^i}{(1-\alpha)^i}}{\frac{\beta-\alpha}{1-\alpha}} \\ &= \frac{(1-\alpha)^i-(1-\beta)^i}{\beta-\alpha}.\end{aligned} \tag{277}$$

回想起 $\Theta_i = (1-\beta)^i - (1-\alpha)^i$，我们还可以写成

$$\sum_{k=1}^{i}(1-\alpha)^{i-k}(1-\beta)^{k-1} = \frac{(1-\alpha)^i-(1-\beta)^i}{\beta-\alpha} = -\frac{\Theta_i}{\beta-\alpha}. \tag{278}$$

现在我们将其代入（274）中的各个系数。

的系数。 利用（273），我们将其拆分为

$$\sum_{k=1}^{i}(1-\alpha)^{i-k}A_{k-1} = \sum_{k=1}^{i}(1-\alpha)^{i-k}\Big(1-(1-\gamma)(1-\beta)^{k-1}-2\varepsilon\Big)$$
$$= (1-2\varepsilon)\sum_{k=1}^{i}(1-\alpha)^{i-k} - (1-\gamma)\sum_{k=1}^{i}(1-\alpha)^{i-k}(1-\beta)^{k-1}. \tag{279}$$

利用（275）和（278），我们得到

$$\sum_{k=1}^{i}(1-\alpha)^{i-k}A_{k-1} = (1-2\varepsilon)\cdot\frac{\Delta_i}{\alpha} - (1-\gamma)\cdot\Big(-\frac{\Theta_i}{\beta-\alpha}\Big)$$
$$= \frac{(1-2\varepsilon)\Delta_i}{\alpha} + \frac{(1-\gamma)\Theta_i}{\beta-\alpha}. \tag{280}$$

乘以 $\delta$ 并利用 $\alpha=\delta\gamma$（因此 $\delta/\alpha = 1/\gamma$），得到

$$\delta\sum_{k=1}^{i}(1-\alpha)^{i-k}A_{k-1} = \frac{(1-2\varepsilon)\Delta_i}{\gamma} + \frac{\delta(1-\gamma)\Theta_i}{\beta-\alpha}. \tag{281}$$

的系数。由（274）和（275），

$$-\delta\sum_{k=1}^{i}(1-\alpha)^{i-k}g_{\odot} = -\delta\cdot\frac{\Delta_i}{\alpha}\,g_{\odot} = -\frac{\Delta_i}{\gamma}\,g_{\odot}, \tag{282}$$

再次利用 $\alpha=\delta\gamma$。

的系数。利用（273），

$$\sum_{k=1}^{i}(1-\alpha)^{i-k}B_{k-1} = \sum_{k=1}^{i}(1-\alpha)^{i-k}\Big(1-(1-\beta)^{k-1}\Big)$$
$$= \sum_{k=1}^{i}(1-\alpha)^{i-k} - \sum_{k=1}^{i}(1-\alpha)^{i-k}(1-\beta)^{k-1}. \tag{283}$$

将（275）和（278）代入（283），

$$\sum_{k=1}^{i}(1-\alpha)^{i-k}B_{k-1} = \frac{\Delta_i}{\alpha} - \Big(-\frac{\Theta_i}{\beta-\alpha}\Big)$$
$$= \frac{\Delta_i}{\alpha} + \frac{\Theta_i}{\beta-\alpha}. \tag{284}$$

Thus

$$-\delta\sum_{k=1}^{i}(1-\alpha)^{i-k}B_{k-1}\,g_{\oplus} = -\delta\left(\frac{\Delta_i}{\alpha}+\frac{\Theta_i}{\beta-\alpha}\right)g_{\oplus} = -\frac{\Delta_i}{\gamma}\,g_{\oplus} - \frac{\delta}{\beta-\alpha}\,\Theta_i\,g_{\oplus}, \tag{285}$$

其中我们再次使用了 $\delta/\alpha = 1/\gamma$。

平滑性惩罚。 (274) 中的最后一项是

$$-\frac{\delta^2 D^2 L}{2}\sum_{k=1}^{i}(1-\alpha)^{i-k}. \tag{286}$$

利用对所有 $k$均有 $(1-\alpha)^{i-k}\le 1$，我们得到

$$\sum_{k=1}^{i}(1-\alpha)^{i-k} \;\le\; i, \tag{287}$$

因此

$$-\frac{\delta^2 D^2 L}{2}\sum_{k=1}^{i}(1-\alpha)^{i-k} \;\ge\; -\frac{\delta^2 D^2 L}{2}\,i \;\ge\; -\,i\,\delta^2 D^2 L, \tag{288}$$

其中，最后一个不等式只是放宽了因子 $1/2$，从而得到一个略弱但更简单的界。

将（274）与界（281）、（282）、（285）和（288）结合起来，并回顾 $\mathbf{y}$（$i$）对应于 $F_i$，我们得到

$$F\big(\mathbf{y}(i)\big) \;\ge\; \left[\frac{(1-2\varepsilon)\,\Delta_i}{\gamma} + \frac{\delta(1-\gamma)}{\beta-\alpha}\,\Theta_i\right] g \;-\; \frac{\Delta_i}{\gamma}\,g_{\odot} \;-\; \left[\frac{\Delta_i}{\gamma} + \frac{\delta}{\beta-\alpha}\,\Theta_i\right] g_{\oplus} \;-\; i\,\delta^2 D^2 L,$$

这正是（264）。 □

引理 D.3。假设 $0<\delta\le 1$ 和 $0<\gamma\le 1$，并令

$$\alpha := \delta\gamma, \qquad \beta := \frac{\gamma^2\delta}{1-\delta+\gamma^2\delta}.$$

设 $i_s < i \le \delta^{-1}$ 并假设对每个整数 $i_s < i \le \delta^{-1}$，$Q(i) \neq \varnothing$。定义常数

$$A \;:=\; \frac{(1-\beta)^{-i_s}}{\gamma} - \Big(1+\frac{3}{\gamma}\Big)\varepsilon + 1 - \frac{1}{\gamma}, \qquad C_\gamma \;:=\; \frac{(1-\beta)^{-i_s}-1}{\gamma}. \tag{289}$$

对每个整数 $i_s \le i \le \delta^{-1}$，采用以下简写

$$S_1(i) \;:=\; \sum_{k=i_s+1}^{i}(1-\alpha)^{\,i-k}(1-\beta)^{\,k}, \qquad S_2(i) \;:=\; \sum_{k=i_s+1}^{i}(1-\alpha)^{\,i-k}(1-\beta)^{\,k}\Big(C_\gamma - \beta\,(k-i_s)\Big), \tag{290}$$

有如下界：

$$F\big(\mathbf{y}(i)\big) \;\ge\; (1-\alpha)^{\,i-i_s}\,F\big(\mathbf{y}(i_s)\big) \;+\; \alpha\,A\,S_1(i)\,g \;-\; \alpha\,S_2(i)\,g_{\oplus} \;-\; (i-i_s)\,\delta^2 D^2 L. \tag{291}$$

此外，令 $n := i - i_s$ 和 $q := \frac{1-\beta}{1-\alpha}$，则可得以下闭式表达式

$$\begin{aligned} S_1(i) &= (1-\alpha)^n(1-\beta)^{i_s}\cdot\frac{q(1-q^n)}{1-q} \;=\; \frac{(1-\alpha)^n(1-\beta)^{\,i_s+1}-(1-\beta)^{\,i+1}}{\beta-\alpha}, \\ S_2(i) &= C_\gamma\,S_1(i) \;-\; \beta\,(1-\alpha)^n(1-\beta)^{i_s}\cdot\frac{q\big(1-(n+1)q^n+nq^{n+1}\big)}{(1-q)^2}. \end{aligned} \tag{292}$$

证明。令 $\alpha := \delta\gamma$ 并固定一个索引 $i$ 满足 $i_s < i \le \delta^{-1}$。由观察结果 4，对于每个 $k\in\{i_s+1,\dots,i\}$，我们都有如下单步递推关系

$$F\big(\mathbf{y}(k)\big) \;\ge\; (1-\alpha)\,F\big(\mathbf{y}(k-1)\big) \;+\; \alpha\,V(k-1) \;-\; \frac{\delta^2 D^2 L}{2}. \tag{293}$$

这里我们使用 $\alpha = \delta\gamma$ 将项 $\delta\gamma\, V(k-1)$ 改写为 $\alpha V(k-1)$。

I在切换后的阶段（$k > i_s$），替代项具有式（24）给出的显式形式

$$V(k-1) \;=\; (1-\beta)^{k}\Big[A\,g \;-\; \Big(C_\gamma - \beta\,(k-i_s)\Big)\,g_{\oplus}\Big], \tag{294}$$

其中 $A$ 和 $C_\gamma$ 如式（289）所示。

从 $k = i_s + 1$ 到 $k = i$ 反复应用式（293），由（关于 $k$的标准归纳法）可得

$$\begin{aligned} F\big(\mathbf{y}(i)\big) \;\geq\; & (1-\alpha)^{\,i-i_s}\,F\big(\mathbf{y}(i_s)\big) \\ & +\;\alpha\sum_{k=i_s+1}^{i}(1-\alpha)^{\,i-k}\,V(k-1) \;-\; \frac{\delta^2 D^2 L}{2}\sum_{k=i_s+1}^{i}(1-\alpha)^{\,i-k}. \end{aligned} \tag{295}$$

因子 $(1-\alpha)^{\,i-i_s}$ 来自在递推关系的齐次部分中反复乘以 $(1-\alpha)$。

将式（294）代入式（295）得

$$\begin{aligned} F\big(\mathbf{y}(i)\big) \geq\; & (1-\alpha)^{\,i-i_s}\,F\big(\mathbf{y}(i_s)\big) \;+\; \alpha\sum_{k=i_s+1}^{i}(1-\alpha)^{\,i-k}(1-\beta)^{k}\,A\,g \\ & -\;\alpha\sum_{k=i_s+1}^{i}(1-\alpha)^{\,i-k}(1-\beta)^{k}\Big(C_\gamma - \beta\,(k-i_s)\Big)\,g_{\oplus} \\ & -\;\frac{\delta^2 D^2 L}{2}\sum_{k=i_s+1}^{i}(1-\alpha)^{\,i-k}. \end{aligned} \tag{296}$$

第一个求和收集所有 $g$ 项，第二个求和收集所有 $g_{\oplus}$ 项，最后一个求和是累积的平滑性惩罚。

现在定义这些求和式

$$S_1(i) \;:=\; \sum_{k=i_s+1}^{i}(1-\alpha)^{\,i-k}(1-\beta)^{k}, \quad S_2(i) \;:=\; \sum_{k=i_s+1}^{i}(1-\alpha)^{\,i-k}(1-\beta)^{k}\Big(C_\gamma - \beta\,(k-i_s)\Big). \tag{297}$$

由（297），不等式（296）变为

$$F\big(\mathbf{y}(i)\big) \;\geq\; (1-\alpha)^{\,i-i_s}\,F\big(\mathbf{y}(i_s)\big) \;+\; \alpha\,A\,S_1(i)\,g \;-\; \alpha\,S_2(i)\,g_{\oplus} \;-\; \frac{\delta^2 D^2 L}{2}\sum_{k=i_s+1}^{i}(1-\alpha)^{\,i-k}. \tag{298}$$

因此，剩下的任务只有对光滑性项进行界定，并计算$S_1(i)$和 $S_2(i)$的闭式表达式。

为得到闭式表达式，令 $n := i - i_s$ 和 $q := \frac{1-\beta}{1-\alpha}$，并设 $t := k - i_s$，使得 $t = 1, \ldots, n$。则

$$S_1(i) = \sum_{k=i_s+1}^{i}(1-\alpha)^{\,i-k}(1-\beta)^{k} = (1-\alpha)^{n}(1-\beta)^{i_s}\sum_{t=1}^{n} q^{t}, \tag{299}$$

因为 $(1-\alpha)^{i-k} = (1-\alpha)^{n-t}$ 且 $(1-\beta)^{k} = (1-\beta)^{i_s+t}$，所以每一项都是

$$(1-\alpha)^{n-t}(1-\beta)^{i_s+t} = (1-\alpha)^{n}(1-\beta)^{i_s}\left(\frac{1-\beta}{1-\alpha}\right)^{t} = (1-\alpha)^{n}(1-\beta)^{i_s}q^{t}.$$

利用等比求和公式

$$\sum_{t=1}^{n} q^{t} = \frac{q(1-q^{n})}{1-q},$$

我们得到

$$S_1(i) \;=\; (1-\alpha)^{\,n}(1-\beta)^{\,i_s}\cdot\frac{q(1-q^{\,n})}{1-q}. \tag{300}$$

记 $1-q=\frac{\beta-\alpha}{1-\alpha}$ 和 $q^n=\left(\frac{1-\beta}{1-\alpha}\right)^n$，经过简单的代数变形可得等价形式

$$S_1(i) \;=\; \frac{(1-\alpha)^{\,n}(1-\beta)^{\,i_s+1}-(1-\beta)^{\,i+1}}{\beta-\alpha}, \tag{301}$$

这就是（292）的第一行。

对于 $S_2(i)$，定义

$$T(i) \;:=\; \sum_{k=i_s+1}^{i}(1-\alpha)^{\,i-k}(1-\beta)^{\,k}\,(k-i_s) \;=\; (1-\alpha)^{\,n}(1-\beta)^{\,i_s}\sum_{t=1}^{n}t\,q^{\,t}.$$

标准恒等式

$$\sum_{t=1}^{n}t\,q^{\,t} \;=\; \frac{q\left(1-(n+1)q^{\,n}+nq^{\,n+1}\right)}{(1-q)^2}$$

于是得到

$$T(i) \;=\; (1-\alpha)^{\,n}(1-\beta)^{\,i_s}\cdot\frac{q\left(1-(n+1)q^{\,n}+nq^{\,n+1}\right)}{(1-q)^2}. \tag{302}$$

根据 $S_2(i)$ 的定义，

$$S_2(i)=C_\gamma\,S_1(i)-\beta\,T(i),$$

结合（300）和（302可得（292）的第二行。

最后，我们界定（298）中的平滑性惩罚。由于 $0\le\;(1-\alpha)^{\,i-k}\le 1$,

$$\sum_{k=i_s+1}^{i}(1-\alpha)^{\,i-k}=\sum_{t=0}^{n-1}(1-\alpha)^t \;\le\; n. \tag{303}$$

因此

$$-\frac{\delta^2D^2L}{2}\sum_{k=i_s+1}^{i}(1-\alpha)^{\,i-k} \;\ge\; -\frac{\delta^2D^2L}{2}\,n \;\ge\; -(i-i_s)\,\delta^2D^2L, \tag{304}$$

其中，我们放宽了因子[1 2]，以得到稍微简单一些的界。

将（304）代入（298）恰好得到（291），从而完成证明。 □

结合引理 D.2 和 D.3，我们得到以下推论；当 $Q(i)$ 对所有 $i\in[\delta^{-1}]$都非空时，该推论完成了定理 4.2 的证明。

推论 D.4。假设对于每个 $i\in[\delta^{-1}]$，$Q(i)\neq\varnothing$。固定 $0<\delta\le 1$、$0<\gamma\le 1$，并令集合

$$\alpha:=\delta\gamma,\qquad \beta:=\frac{\gamma^2\delta}{1-\delta+\gamma^2\delta}. \tag{305}$$

于是，令 $i_s\in\{0,1,\dots,\delta^{-1}\}$ 和 $\mathbf{y}(\cdot)$ 如引理 D.1所示，我们有

$$\begin{aligned}
F\big(\mathbf{y}(\delta^{-1})\big) \;&\ge\; (1-\alpha)^{\,\delta^{-1}-i_s}\,F\big(\mathbf{y}(i_s)\big) \;+\; \alpha\,A\,S_1(\delta^{-1})\,g \;-\; \alpha\,S_2(\delta^{-1})\,g_{\oplus} \;-\; (\delta^{-1}-i_s)\,\delta^2D^2L\\
&\ge\; (1-\alpha)^{\,\delta^{-1}-i_s}\Bigg\{\Big[\tfrac{1-(1-\alpha)^{i_s}}{\gamma}\,(1-2\varepsilon)+\tfrac{\delta(1-\gamma)}{\beta-\alpha}\big((1-\beta)^{i_s}-(1-\alpha)^{i_s}\big)\Big]g\\
&\qquad -\;\tfrac{1-(1-\alpha)^{i_s}}{\gamma}\,g_{\odot}-\Big[\tfrac{1-(1-\alpha)^{i_s}}{\gamma}+\tfrac{\delta}{\beta-\alpha}\big((1-\beta)^{i_s}-(1-\alpha)^{i_s}\big)\Big]g_{\oplus}-\;i_s\,\delta^2D^2L\Bigg\}\\
&\quad +\;\alpha\,A\,S_1(\delta^{-1})\,g \;-\; \alpha\,S_2(\delta^{-1})\,g_{\oplus} \;-\; (\delta^{-1}-i_s)\,\delta^2D^2L,
\end{aligned} \tag{306}$$

其中

$$A := \frac{(1-\beta)^{-i_s}}{\gamma} - \Big(1+\frac{3}{\gamma}\Big)\varepsilon + 1 - \frac{1}{\gamma}, \qquad C_\gamma := \frac{(1-\beta)^{-i_s}-1}{\gamma}, \tag{307}$$

以及引理 D.3中的求和可写成闭式表达式

$$\begin{aligned} S_1(\delta^{-1}) &= \frac{(1-\alpha)^{\delta^{-1}-i_s}(1-\beta)^{i_s+1} - (1-\beta)^{\delta^{-1}+1}}{\beta-\alpha}, \\ S_2(\delta^{-1}) &= C_\gamma\, S_1(\delta^{-1}) \;-\; \beta\,(1-\alpha)^{\delta^{-1}-i_s}(1-\beta)^{i_s} \\ &\qquad \frac{\frac{1-\beta}{1-\alpha}\Big(1-(\delta^{-1}-i_s+1)\big(\frac{1-\beta}{1-\alpha}\big)^{\delta^{-1}-i_s} + (\delta^{-1}-i_s)\big(\frac{1-\beta}{1-\alpha}\big)^{\delta^{-1}-i_s+1}\Big)}{\Big(1-\frac{1-\beta}{1-\alpha}\Big)^2}. \end{aligned} \tag{308}$$

证明。 我们使用 ( $\alpha$ 和 $\beta$ ) 中的简写参数 305。根据引理D.2，在$i = i_s$处取值可得

$$\begin{aligned} F\big(\mathbf{y}(i_s)\big) \;\geq\; & \left[\frac{1-(1-\alpha)^{i_s}}{\gamma}(1-2\varepsilon) + \frac{\delta(1-\gamma)}{\beta-\alpha}\Big((1-\beta)^{i_s}-(1-\alpha)^{i_s}\Big)\right] g \\ & - \frac{1-(1-\alpha)^{i_s}}{\gamma}\, g_\odot - \left[\frac{1-(1-\alpha)^{i_s}}{\gamma} + \frac{\delta}{\beta-\alpha}\Big((1-\beta)^{i_s}-(1-\alpha)^{i_s}\Big)\right] g_\oplus - \; i_s\,\delta^2 D^2 L. \end{aligned} \tag{309}$$

这就是引理 D.2 在 $i = i_s$处的情形，以及定义 $\Delta_{i_s} = 1-(1-\alpha)^{i_s}$和 $\Theta_{i_s} = (1-\beta)^{i_s} - (1-\alpha)^{i_s}$。

接下来，根据引理 D.3，在 $i = \delta^{-1}$处取值可得

$$F\big(\mathbf{y}(\delta^{-1})\big) \;\geq\; (1-\alpha)^{\delta^{-1}-i_s}\, F\big(\mathbf{y}(i_s)\big) + \alpha\, A\, S_1(\delta^{-1})\, g - \alpha\, S_2(\delta^{-1})\, g_\oplus - (\delta^{-1}-i_s)\,\delta^2 D^2 L \tag{310}$$

其中 $A$ 和 $C_\gamma$ 与 (307) 中的一致， $S_1(\delta^{-1})$ 中的 $S_2(\delta^{-1})$ 由 (308) 给出。

将下界（309）代入 $F(\mathbf{y}(i_s))$，再代入（310），随后收集 $g$、$g_\odot$和 $g_\oplus$的系数，并合并平滑性惩罚项 $i_s\delta^2 D^2 L$ 和 $(\delta^{-1}-i_s)\,\delta^2 D^2 L$，恰好得到离散不等式（306）。这只是简单的代数变形。

为了从离散时间界（306）推导出连续时间保证，定义切换时刻

$$t_s := \delta\, i_s \in [0,1],$$

并令 $\delta \to 0^+$，同时保持 $t_s$ 不变。利用

$$(1-\alpha)^{\delta^{-1}-i_s} = \big(1-\delta\gamma\big)^{(1-t_s)/\delta} \;\longrightarrow\; e^{-\gamma(1-t_s)}, \qquad (1-\alpha)^{i_s} = \big(1-\delta\gamma\big)^{t_s/\delta} \;\longrightarrow\; e^{-\gamma t_s},$$

以及闭式表达式 (308)，每个离散求和都收敛到连续时间分析中相应的时间积分。具体而言， $S_1(\delta^{-1})$ 和 $S_2(\delta^{-1})$ 可视为步长 $\delta$下的黎曼和；对 $(1-\alpha)$ 和 $(1-\beta)$ 应用标准的一阶展开（或者等价地，对相关极限应用洛必达法则），即可得到连续时间系数 $A_\gamma(t_s)$、 $B_\gamma(t_s)$ 和$C_\gamma(t_s)$:

$$F\big(\mathbf{y}^*\big) \;\geq\; A_\gamma(t_s)\, g \;+\; B_\gamma(t_s)\, g_\odot \;+\; C_\gamma(t_s)\, g_\oplus \;-\; O(\varepsilon)\,\big(g + g_\odot + g_\oplus\big) \;-\; \delta\, L D^2, \tag{311}$$

with

$$A_\gamma(t_s) := -\frac{e^{\gamma t_s-\gamma}}{1-\gamma} + \frac{e^{-\gamma^2}}{\gamma(1-\gamma)}\Big(e^{\gamma^2 t_s} - (1-\gamma)\Big), \tag{312}$$

$$B_\gamma(t_s) := \frac{e^{-\gamma} - e^{\gamma t_s-\gamma}}{\gamma}, \tag{313}$$

$$C_\gamma(t_s) := \frac{e^{\gamma^2 t_s}-1}{\gamma(1-\gamma)}\Big(e^{-\gamma(1-t_s)-\gamma^2 t_s} - e^{-\gamma^2}\Big) + \frac{e^{-\gamma(1-t_s)}}{\gamma}\left[-\big(1-e^{-\gamma t_s}\big) + \frac{e^{-\gamma^2 t_s} - e^{-\gamma t_s}}{1-\gamma}\right]$$
$$+\ e^{-\gamma(1-t_s)-\gamma^2 t_s}\left[\frac{\gamma^2}{1-\gamma}(1-t_s)\,e^{\gamma(1-\gamma)(1-t_s)} + \frac{\gamma}{(1-\gamma)^2}\Big(1-e^{\gamma(1-\gamma)(1-t_s)}\Big)\right]. \tag{314}$$

最后，将 $(g, g_\odot, g_\oplus)$ 与 $\big(F(\mathbf{o})\ \ F(\mathbf{z}\odot\mathbf{o})\ \ F(\mathbf{z}\oplus\mathbf{o})\big)$ 对应起来可知，(311) 正是定理 4.2(1) 中使用的连续时间保证。 □

此时，我们处理存在某个 $i \in [\delta^{-1}]$ 使得 $Q(i) = \varnothing$ 的情形。注意，$Q(i) = \varnothing$ 尤其意味着，$\mathbf{o} \notin Q(i)$。因此，定义

$$i_o := \min\big\{\, i \in [\delta^{-1}] \;:\; \mathbf{o} \notin Q(i) \,\big\}.$$

观察 5. 对于每个 $i \in [i_o - 1]$,

$$F\big(\mathbf{x}(i)\big) \;\geq\; \frac{\gamma^2 F\big(\mathbf{x}(i)\vee\mathbf{o}\big) + F\big(\mathbf{x}(i)\wedge\mathbf{o}\big)}{1+\gamma^2} \;-\; \varepsilon F(\mathbf{o}) \;-\; \delta LD^2. \tag{315}$$

证明。固定 $i \in [i_o - 1]$。由于 $i < i_o$，根据 $\mathbf{o} \in Q(i)$ 的定义，我们有 $i_o$。我们将定理 3.2 应用于向下闭体 $P := Q(i)$，并取比较点 $\mathbf{y} := \mathbf{o} \in Q(i)$，然后以精度参数运行局部例程

$$\eta \;:=\; \min\{\varepsilon,\, 2\delta\}.$$

根据定理 3.2，输出 $\mathbf{x}(i) \in Q(i)$ 满足

$$F\big(\mathbf{x}(i)\big) \geq \frac{\gamma^2 F\big(\mathbf{x}(i)\vee\mathbf{o}\big) + F\big(\mathbf{x}(i)\wedge\mathbf{o}\big)}{1+\gamma^2} - \frac{\eta\gamma}{1+\gamma^2}\left(\max_{\mathbf{y}'\in Q(i)} F(\mathbf{y}') + \tfrac{1}{2}LD^2\right). \tag{316}$$

这里第一项恰好是定理 3.2中的基于格的比较，第二项则是参数为 $\eta$ 的一致一阶误差。

由于 $\mathbf{o} \in Q(i)$，我们有

$$\max_{\mathbf{y}'\in Q(i)} F(\mathbf{y}') \;\geq\; F(\mathbf{o}).$$

在（316）中，$\max_{\mathbf{y}'\in Q(i)} F(\mathbf{y}')$ 前面的系数为负，即 $-\frac{\eta\gamma}{1+\gamma^2}$。因此，将 $\max_{\mathbf{y}'\in Q(i)} F(\mathbf{y}')$ 替换为更小的值 $F(\mathbf{o})$，可得到关于 $F(\mathbf{x}(i))$ 的更强下界：

$$F\big(\mathbf{x}(i)\big) \;\geq\; \frac{\gamma^2 F\big(\mathbf{x}(i)\vee\mathbf{o}\big) + F\big(\mathbf{x}(i)\wedge\mathbf{o}\big)}{1+\gamma^2} - \frac{\eta\gamma}{1+\gamma^2}F(\mathbf{o}) - \frac{\eta\gamma}{2(1+\gamma^2)}LD^2. \tag{317}$$

现在我们来简化这两个误差项。由于 $\gamma \in (0, 1]$,

$$\frac{\gamma}{1+\gamma^2} \;\leq\; 1, \qquad \frac{\gamma}{2(1+\gamma^2)} \;\leq\; \frac{1}{2}.$$

利用 $\eta \leq \varepsilon$ 和 $\eta \leq 2\delta$（根据定义 $\eta = \min\{\varepsilon, 2\delta\}$），可得

$$\frac{\eta\gamma}{1+\gamma^2}F(\mathbf{o}) \;\leq\; \varepsilon F(\mathbf{o}), \qquad \frac{\eta\gamma}{2(1+\gamma^2)}LD^2 \;\leq\; \delta LD^2. \tag{318}$$

将（318）中的界代入（317），得到

$$F\big(\mathbf{x}(i)\big) \;\geq\; \frac{\gamma^2 F\big(\mathbf{x}(i)\vee\mathbf{o}\big) + F\big(\mathbf{x}(i)\wedge\mathbf{o}\big)}{1+\gamma^2} \;-\; \varepsilon F(\mathbf{o}) \;-\; \delta LD^2,$$

这正是（315）。 □

观察5 给出了定理4.2 只要我们能够找到某个 $i \in [i_o - 1]$ 满足

$$F\big(\mathbf{x}(i) \oplus \mathbf{o}\big) \;\le\; F\big(\mathbf{z} \oplus \mathbf{o}\big) \;-\; \varepsilon\, F(\mathbf{o}). \tag{319}$$

因此，我们的任务归结为证明间隙条件（319）对至少一个索引成立$i \in [i_o - 1]$。引理 D.6 和引理 D.7 恰好证明了这一点，从而完成了定理 4.2的证明。在陈述这些引理之前，我们先记录一个将在证明中用到的辅助引理。

引理 D.5。必有

$$F\big(\mathbf{y}(i_o - 1) \oplus \mathbf{o} \;-\; \mathbf{z}(i_o - 1) \odot \mathbf{o}\big) \;\le\; V_\gamma(i_o - 1).$$

证明。根据 $i_o$的定义，我们有 $\mathbf{o} \notin Q(i_o)$。在迭代 $i_o$ 时，弱 DR 成员资格失败条件表明

$$\big\langle \mathbf{w}(i_o),\, \mathbf{o} \big\rangle \;\le\; \gamma\big(V_\gamma(i_o - 1) - F(\mathbf{y}(i_o - 1))\big). \tag{320}$$

另一方面，对于任意 $i \in [\delta^{-1}]$，弱 DR 梯度界给出

$$\big\langle \mathbf{w}(i),\, \mathbf{o} \big\rangle \;\ge\; \gamma\Big(F\big(\mathbf{y}(i-1) \oplus \mathbf{o} \;-\; \mathbf{z}(i-1) \odot \mathbf{o}\big) - F(\mathbf{y}(i-1))\Big). \tag{321}$$

这对 $i = i_o$ 尤其适用，因此

$$\big\langle \mathbf{w}(i_o),\, \mathbf{o} \big\rangle \;\ge\; \gamma\Big(F\big(\mathbf{y}(i_o - 1) \oplus \mathbf{o} \;-\; \mathbf{z}(i_o - 1) \odot \mathbf{o}\big) - F(\mathbf{y}(i_o - 1))\Big). \tag{322}$$

结合（320）和（322）可得

$$\gamma\Big(F\big(\mathbf{y}(i_o - 1) \oplus \mathbf{o} \;-\; \mathbf{z}(i_o - 1) \odot \mathbf{o}\big) - F(\mathbf{y}(i_o - 1))\Big) \;\le\; \gamma\big(V_\gamma(i_o - 1) - F(\mathbf{y}(i_o - 1))\big). \tag{323}$$

由于 $\gamma > 0$，我们可以将（323）两边除以 $\gamma$并在两边加上 $F(\mathbf{y}(i_o - 1))$，得到

$$F\big(\mathbf{y}(i_o - 1) \oplus \mathbf{o} \;-\; \mathbf{z}(i_o - 1) \odot \mathbf{o}\big) \;\le\; V_\gamma(i_o - 1),$$

这就是所需的不等式。 □

引理 D.6. 设 $\beta := \beta_\gamma(\delta) = \dfrac{\gamma^2\delta}{1 - \delta + \gamma^2\delta}$. 则必有 $i_o > i_{s\circ}$

证明. 反设 $i_o \le i_{s\circ}$ 由于 $i_o - 1 < i_s$，在时刻 D.5 的引理 $i_o - 1$ (参见$v_1(\cdot)$及三元组 $(g, g_\odot, g_\oplus)$)可推出

$$\begin{aligned} &F\big(\mathbf{y}(i_o - 1) \oplus \mathbf{o} \;-\; \mathbf{z} \odot \mathbf{o}\big) \\ &\quad\le\; \Big[(1-\beta)^{i_o - 1} + \frac{1 - (1-\beta)^{i_o - 1} - 2\varepsilon}{\gamma}\Big]\, g \;-\; \frac{1}{\gamma}\, g_\odot \;-\; \frac{1 - (1-\beta)^{i_o - 1}}{\gamma}\, g_\oplus \qquad (324) \\ &\quad\le\; \Big[(1-\beta)^{i_o - 1} + \frac{1 - (1-\beta)^{i_o - 1}}{\gamma}\Big]\, F(\mathbf{o}) \;-\; \frac{1}{\gamma}\, F(\mathbf{z} \odot \mathbf{o}) \;-\; \frac{1 - (1-\beta)^{i_o - 1}}{\gamma}\, F(\mathbf{z} \oplus \mathbf{o}). \qquad (325) \end{aligned}$$

这里，（324）正是在 $i = i_o - 1$处取值的基准不等式，而（325）使用了引理 4.1：

$$(1 - \varepsilon)F(\mathbf{o}) \le g \le F(\mathbf{o}), \quad F(\mathbf{z} \odot \mathbf{o}) - \varepsilon g \le g_\odot \le F(\mathbf{z} \odot \mathbf{o}), \quad F(\mathbf{z} \oplus \mathbf{o}) - \varepsilon g \le g_\oplus \le F(\mathbf{z} \oplus \mathbf{o}),$$

并且，由于将 $g$替换为 $F(\mathbf{o})$ 和 $g_\odot, g_\oplus$替换为 $F(\mathbf{z} \odot \mathbf{o})$ $F(\mathbf{z} \oplus \mathbf{o})$只会增大（324）的右侧（因为它们在 $+1$ 前的系数为 $g$，在 $-1/\gamma$ 前的系数为 $g_\odot$，$g_\oplus$）。

接下来，根据 $\mathbf{y}(\cdot)$ 的闭式（引理 D.1，取 $i = i_o - 1$），

$$\mathbf{y}(i_o - 1) \;=\; (\mathbf{1} - \mathbf{z}) \;\odot\; \bigoplus_{j=1}^{i_o - 1} \big(\delta\, \mathbf{x}(j)\big). \tag{326}$$

应用推论 E.2，取 $h=1,\ \ r=i_o-1,\ \ p_j=\delta$ 对所有 $j$，并取外层掩码（$\mathbf{1}-\mathbf{z}$），可得混合下界

$$F\big(\mathbf{y}(i_o-1)\oplus\mathbf{o}\ -\ \mathbf{z}\odot\mathbf{o}\big)=F\Big((\mathbf{1}-\mathbf{z})\odot\bigoplus_{j=1}^{i_o-1}(\delta\,\mathbf{x}(j))\ \oplus\ \mathbf{o}\ -\ \mathbf{z}\odot\mathbf{o}\Big)$$
$$\geq\sum_{S\subseteq[i_o-1]}\beta^{|S|}(1-\beta)^{i_o-1-|S|}\,F\Big((\mathbf{1}-\mathbf{z})\odot\bigoplus_{j\in S}\mathbf{x}(j)\ \oplus\ \mathbf{o}\Big).\quad(327)$$

现在我们分别界定 $S=\varnothing$ 和 $S\neq\varnothing$ 的贡献。

该 $S=\varnothing$ 项。当 $S=\varnothing$时，内部向量化简为 $(\mathbf{1}-\mathbf{z})\odot\mathbf{o}$。利用引理 2.2 形式的弱 DR 不等式以及非负性，可得

$$F\big((\mathbf{1}-\mathbf{z})\odot\mathbf{o}\big)\ \geq\ F(\mathbf{o})\ -\ \frac{1}{\gamma}F(\mathbf{z}\odot\mathbf{o}),\quad(328)$$

这表明，沿着$\mathbf{z}$ 方向将$\mathbf{z}\odot\mathbf{o}$ 减少 $F$时，其减少幅度不会超过 $1/\gamma$倍。因此，式 ( $S=\varnothing$ 中的贡献 327满足

$$(1-\beta)^{i_o-1}F\big((\mathbf{1}-\mathbf{z})\odot\mathbf{o}\big)\ \geq\ (1-\beta)^{i_o-1}\Big[F(\mathbf{o})-\frac{1}{\gamma}F(\mathbf{z}\odot\mathbf{o})\Big].\quad(329)$$

$S\neq\varnothing$ 项。固定任意非空的 $S\subseteq[i_o-1]$。我们首先写出

$$(\mathbf{1}-\mathbf{z})\odot\bigoplus_{j\in S}\mathbf{x}(j)\ \oplus\ \mathbf{o}\ =\ \Big((\mathbf{1}-\mathbf{z})\odot\bigoplus_{j\in S}\mathbf{x}(j)\ \oplus\ \mathbf{o}\ -\ \mathbf{z}\odot\mathbf{o}\Big)\ +\ \mathbf{z}\odot\mathbf{o},$$

然后以与式 (328) 相同的方式应用弱 DR 不等式和非负性。这给出

$$F\Big((\mathbf{1}-\mathbf{z})\odot\bigoplus_{j\in S}\mathbf{x}(j)\ \oplus\ \mathbf{o}\Big)\geq F\Big((\mathbf{1}-\mathbf{z})\odot\bigoplus_{j\in S}\mathbf{x}(j)\ \oplus\ \mathbf{o}\ -\ \mathbf{z}\odot\mathbf{o}\Big)\ -\ \frac{1}{\gamma}F(\mathbf{z}\odot\mathbf{o})$$
$$\geq\frac{1}{\gamma}\Big[F(\mathbf{o})-F(\mathbf{z}\oplus\mathbf{o})-F(\mathbf{z}\odot\mathbf{o})\Big],\quad(330)$$

其中，在第二个不等式中，我们利用弱 DR 差值界，将$\mathbf{o}$ 处的值与$\mathbf{z}\oplus\mathbf{o}$ 和$\mathbf{z}\odot\mathbf{o}$处的值进行比较。将式 (330) 乘以 $\beta^{|S|}(1-\beta)^{i_o-1-|S|}$，并对所有非空的 $S$ 求和可得

$$\sum_{\varnothing\neq S\subseteq[i_o-1]}\beta^{|S|}(1-\beta)^{i_o-1-|S|}F\Big((\mathbf{1}-\mathbf{z})\odot\bigoplus_{j\in S}\mathbf{x}(j)\ \oplus\ \mathbf{o}\Big)$$
$$\geq\Big[1-(1-\beta)^{i_o-1}\Big]\frac{1}{\gamma}\Big[F(\mathbf{o})-F(\mathbf{z}\oplus\mathbf{o})-F(\mathbf{z}\odot\mathbf{o})\Big],\quad(331)$$

因为 $\sum_{\varnothing\neq S\subseteq[i_o-1]}\beta^{|S|}(1-\beta)^{i_o-1-|S|}=1-(1-\beta)^{i_o-1}$。

综合以上结果。 将 $S=\varnothing$ 的贡献（329）与 $S\neq\varnothing$的贡献（331）结合混合表示（327），我们得到

$$F\big(\mathbf{y}(i_o-1)\oplus\mathbf{o}\ -\ \mathbf{z}\odot\mathbf{o}\big)\geq\Big[(1-\beta)^{i_o-1}+\frac{1-(1-\beta)^{i_o-1}}{\gamma}\Big]F(\mathbf{o})$$
$$-\ \frac{1-(1-\beta)^{i_o-1}}{\gamma}F(\mathbf{z}\oplus\mathbf{o})\ -\ \frac{1}{\gamma}F(\mathbf{z}\odot\mathbf{o}).\quad(332)$$

（332）的右侧严格大于（325）中的上界，因为（325）缺少（324）中存在的 $-2\varepsilon$ 松弛项，而三元组（$g$、$g_{\odot}$、$g_{\oplus}$）对 $(F(\mathbf{o})\ F(\mathbf{z}\odot\mathbf{o})\ F(\mathbf{z}\oplus\mathbf{o}))$的近似误差至多为加性 $O(\varepsilon)$ 项。这与（325）矛盾。因此，我们的假设 $i_o\leq i_s$ 不成立，故可得 $i_o>i_s$。

□

引理 D.7. 若 $i_o > i_s$，则存在某个 $i \in [i_o - 1]$ 使得

$$F\big(\mathbf{x}(i) \oplus \mathbf{o}\big) \;\le\; F\big(\mathbf{z} \oplus \mathbf{o}\big) \;-\; \varepsilon\, F(\mathbf{o})\,.$$

证明。令 $\beta := \beta_\gamma(\delta) = \dfrac{\gamma^2\delta}{1-\delta+\gamma^2\delta}$，并注意到 $0 < \beta \le \varepsilon \le \frac{1}{2}$。由于 $i_o > i_s$，我们有 $i_o - 1 \ge i_s$。

上界：根据在 $i = i_o - 1$ 处取值的切换后代理函数（即 $v_2$ 界）以及对猜测的三元组 $(g, g_\odot, g_\oplus)$ 的成功继承者假设（参见（24）和（238）），我们得到

$$\begin{aligned} F\big(\mathbf{y}(i_o-1)\oplus\mathbf{o}\big) &\overset{\text{(a)}}{\le} (1-\beta)^{i_o-1}\Bigg[\left(\frac{(1-\beta)^{-i_s}}{\gamma} - \Big(1+\frac{3}{\gamma}\Big)\varepsilon + 1 - \frac{1}{\gamma}\right) g \\ &\qquad - \left(\frac{(1-\beta)^{-i_s}}{\gamma} - \frac{1}{\gamma} - \beta\,(i_o - 1 - i_s)\right) g_\oplus\Bigg] \qquad (333) \\ &\overset{\text{(b)}}{\le} (1-\beta)^{i_o-1}\Bigg[\left(\frac{(1-\beta)^{-i_s}}{\gamma} - \varepsilon + 1 - \frac{1}{\gamma}\right) F(\mathbf{o}) \\ &\qquad - \left(\frac{(1-\beta)^{-i_s}}{\gamma} - \frac{1}{\gamma} - \beta\,(i_o - 1 - i_s)\right) F(\mathbf{z}\oplus\mathbf{o})\Bigg]. \qquad (334) \end{aligned}$$

这里：(a) 只是将常数 $A$ 和 $C_\gamma$ 展开后，$v_2(i_o-1)$ 的显式公式。对于 (b)，我们使用三元组保证（238）：

$$g \le F(\mathbf{o}), \qquad g_\oplus \;\ge\; F(\mathbf{z}\oplus\mathbf{o}) - \varepsilon\, g.$$

记 $B := \frac{(1-\beta)^{-i_s} - 1}{\gamma} - \beta\,(i_o - 1 - i_s) \ge 0$，则有

$$Ag - Bg_\oplus \;\le\; Ag - B\big(F(\mathbf{z}\oplus\mathbf{o}) - \varepsilon g\big) \;=\; (A + B\varepsilon)\, g \;-\; B\, F(\mathbf{z}\oplus\mathbf{o}),$$

然后，$g \le F(\mathbf{o})$ 给出 $(A+B\varepsilon)\, g \le (A+B\varepsilon)F(\mathbf{o})$。一个粗略的界 $(1-\beta)^{-i_s} \le (1-\beta)^{-1/\beta} \le 4$（使用 $\beta \le 1/2$）意味着

$$B = \frac{(1-\beta)^{-i_s} - 1}{\gamma} - \beta(i_o - 1 - i_s) \;\le\; \frac{(1-\beta)^{-i_s} - 1}{\gamma} \;\le\; \frac{3}{\gamma}.$$

Thus

$$A + B\varepsilon \;\le\; \left(\frac{(1-\beta)^{-i_s}}{\gamma} - \Big(1+\frac{3}{\gamma}\Big)\varepsilon + 1 - \frac{1}{\gamma}\right) + \frac{3}{\gamma}\varepsilon = \frac{(1-\beta)^{-i_s}}{\gamma} - \varepsilon + 1 - \frac{1}{\gamma},$$

这恰好是 $F(\mathbf{o})$ 在 (334) 中的系数。

下界：由 $\mathbf{y}(\cdot)$ 的闭式（引理 D.1 及 $i = i_o - 1$），

$$\mathbf{y}(i_o - 1) \;=\; (\mathbf{1} - \mathbf{z}) \odot \bigoplus_{j=1}^{i_o-1} \big(\delta\, \mathbf{x}(j)\big) \;+\; \mathbf{z} \odot \bigoplus_{j=i_s+1}^{i_o-1} \big(\delta\, \mathbf{x}(j)\big). \qquad (335)$$

将带掩蔽的弱 DR 混合不等式（推论 E.2）应用于 (335)。我们只保留三组非负项：- 空集项 $S = \varnothing$；- 所有子集 $S \subseteq [i_s]$；- 所有单元素集合 $S = \{j\}$，其中 $j \in \{i_s + 1, \dots, i_o - 1\}$。

利用引理 2.2（弱 DR 梯度界）以及 $F$的非负性，这三组项给出

$$\begin{aligned} F\big(\mathbf{y}(i_o-1)\oplus\mathbf{o}\big) &\overset{\text{(c)}}{\ge} (1-\beta)^{i_o-1} F(\mathbf{o}) \;+\; (1-\beta)^{i_o-1-i_s}\big(1-(1-\beta)^{i_s}\big)\,\frac{1}{\gamma}\big[F(\mathbf{o}) - F(\mathbf{z}\oplus\mathbf{o})\big] \\ &\qquad +\; \beta(1-\beta)^{i_o-2} \sum_{j=i_s+1}^{i_o-1} F\big(\mathbf{x}(j)\oplus\mathbf{o}\big), \qquad (336) \end{aligned}$$

其中 (c) 的来源如下：- $S = \varnothing$：给出项 $(1-\beta)^{i_o-1}F(\mathbf{o})$；- $S \subseteq [i_s], S \ne \varnothing$：合并后下界为 $\frac{1}{\gamma}\big[F(\mathbf{o}) - F(\mathbf{z}\oplus\mathbf{o})\big]$，总权重为 $(1-\beta)^{i_o-1-i_s}\big(1-(1-\beta)^{i_s}\big)$；- 满足 $j > i_s$ 的单元素集合 $S = \{j\}$：每个权重为 $\beta(1-\beta)^{i_o-2}$，并贡献 $F(\mathbf{x}(j)\oplus\mathbf{o})$。

比较这些界。结合（334）和（336），并将两边除以公因子 $(1-\beta)^{\,i_o-1}>0$ 得到

$$F(\mathbf{o})+(1-\beta)^{-i_s}\bigl(1-(1-\beta)^{\,i_s}\bigr)\,\frac{1}{\gamma}\bigl[F(\mathbf{o})-F(\mathbf{z}\oplus\mathbf{o})\bigr]+\frac{\beta}{1-\beta}\sum_{j=i_s+1}^{i_o-1}F\bigl(\mathbf{x}(j)\oplus\mathbf{o}\bigr)$$

$$\le\ \left(\frac{(1-\beta)^{-i_s}}{\gamma}-\varepsilon+1-\frac{1}{\gamma}\right)F(\mathbf{o})-\left(\frac{(1-\beta)^{-i_s}}{\gamma}-\frac{1}{\gamma}-\beta\,(i_o-1-i_s)\right)F(\mathbf{z}\oplus\mathbf{o}).$$

重新排列并化简 $F(\mathbf{o})$ 和 $F(\mathbf{z}\oplus\mathbf{o})$ 的系数，得到

$$\beta\sum_{j=i_s+1}^{i_o-1}F\bigl(\mathbf{x}(j)\oplus\mathbf{o}\bigr)\ \le\ \bigl(1-(1-\beta)^{\,i_o-1-i_s}\bigr)\Bigl[F(\mathbf{z}\oplus\mathbf{o})-\varepsilon\,F(\mathbf{o})\Bigr]. \tag{337}$$

此外，

$$1-(1-\beta)^{\,i_o-1-i_s}\ \le\ \beta\,(i_o-1-i_s),$$

因此由（337）可得

$$\frac{1}{i_o-1-i_s}\sum_{j=i_s+1}^{i_o-1}F\bigl(\mathbf{x}(j)\oplus\mathbf{o}\bigr)\ \le\ F(\mathbf{z}\oplus\mathbf{o})-\varepsilon\,F(\mathbf{o}).$$

通过取平均，必定存在某个 $j\in\{i_s+1,\dots,i_o-1\}$ 使得

$$F\bigl(\mathbf{x}(j)\oplus\mathbf{o}\bigr)\ \le\ F(\mathbf{z}\oplus\mathbf{o})-\varepsilon\,F(\mathbf{o}),$$

这就证明了该引理。 □

# E 支持性结果

在本节中，我们证明两个辅助结果，它们将在引理 D.6 和引理 D.7的证明中用到。全文中，我们采用如下约定

$$\bigoplus_{i\in\varnothing}\mathbf{x}(i)\ :=\ \mathbf{0}.$$

**引理 E.1.** 设 $F\colon[0,1]^n\to\mathbb{R}_{\ge0}$ 是可微且 $\gamma$-弱 DR 次模的，并且 $0<\gamma\le1$。固定一个整数 $r\ge1$、向量 $\mathbf{x(1)}\dots,\mathbf{x}(r)\in[0,1]^n$，以及标量 $p_1,\dots,p_r\in[0,1]$。定义

$$\beta_\gamma(p)\ :=\ \frac{\gamma^2\,p}{1-p+\gamma^2p}\qquad(p\in[0,1]).$$

*Then*

$$F\left(\bigoplus_{i=1}^{r}p_i\,\mathbf{x}(i)\right)\ \ge\ \sum_{S\subseteq[r]}\left(\prod_{i\in S}\beta_\gamma(p_i)\prod_{i\notin S}\bigl(1-\beta_\gamma(p_i)\bigr)\right)F\left(\bigoplus_{i\in S}\mathbf{x}(i)\right).$$

证明。我们通过对 $r$ 进行归纳法来证明该命题。

基例 $r=1$。应用引理 2.1(2) with $\mathbf{x}=\mathbf{0}$、$\mathbf{y}=\mathbf{x}(1)$ and $\lambda=p_1$:

$$F\bigl(p_1\,\mathbf{x}(1)\bigr)-F(\mathbf{0})\ \ge\ \frac{\gamma^2p_1}{1-p_1+\gamma^2p_1}\bigl(F(\mathbf{x}(1))-F(\mathbf{0})\bigr).$$

整理后可得

$$F\bigl(p_1\,\mathbf{x}(1)\bigr)\ \ge\ \beta_\gamma(p_1)\,F(\mathbf{x}(1))+\bigl(1-\beta_\gamma(p_1)\bigr)\,F(\mathbf{0}),$$

这恰好就是当 $r=1$ 和 $S\in\{\varnothing,\{1\}\}$时所声称的公式。

归纳步骤。假设该命题对某个 $r-1 \geq 1$成立，并考虑 $r$。定义

$$G_1(\mathbf{x}) \;:=\; F\!\left(\mathbf{x} \;\oplus\; \bigoplus_{i=1}^{r-1} p_i\,\mathbf{x}(i)\right), \qquad G_2(\mathbf{x}) \;:=\; F\big(\mathbf{x} \oplus \mathbf{x}(r)\big).$$

根据引理 2.3，两者 $G_1$ 和 $G_2$ 均为非负的，且 $\gamma$-弱 DR 次模。

沿着射线 2.1中的引理 (2)，从基点 $G_1$ 0 出发，步长为 $\mathbf{x}r$: $\lambda = p_r$

$$\begin{aligned} F\!\left(\bigoplus_{i=1}^{r} p_i\,\mathbf{x}(i)\right) &= G_1\big(p_r\,\mathbf{x}(r)\big) \\ &\geq \beta_\gamma(p_r)\,G_1\big(\mathbf{x}(r)\big) \;+\; \big(1-\beta_\gamma(p_r)\big)\,G_1(\mathbf{0}). \end{aligned} \tag{338}$$

根据 $G_1$ 和 $G_2$的定义,

$$G_1(\mathbf{x}(r)) = G_2\!\left(\bigoplus_{i=1}^{r-1} p_i\,\mathbf{x}(i)\right), \qquad G_1(\mathbf{0}) = F\!\left(\bigoplus_{i=1}^{r-1} p_i\,\mathbf{x}(i)\right).$$

代入式 (338) 得

$$F\!\left(\bigoplus_{i=1}^{r} p_i\,\mathbf{x}(i)\right) \;\geq\; \beta_\gamma(p_r)\,G_2\!\left(\bigoplus_{i=1}^{r-1} p_i\,\mathbf{x}(i)\right) \;+\; \big(1-\beta_\gamma(p_r)\big)\,F\!\left(\bigoplus_{i=1}^{r-1} p_i\,\mathbf{x}(i)\right).$$

现在将归纳假设应用于 $G_2$（其中有 $r-1$ 个向量 $\mathbf{x(1)}\ldots,\mathbf{x}(r-1)$，权重为 $p_1,\ldots,p_{r-1}$），并应用于 $F$:

$$G_2\!\left(\bigoplus_{i=1}^{r-1} p_i\,\mathbf{x}(i)\right) \geq \sum_{S\subseteq[r-1]} \left(\prod_{i\in S}\beta_\gamma(p_i)\prod_{i\notin S}\big(1-\beta_\gamma(p_i)\big)\right) G_2\!\left(\bigoplus_{i\in S}\mathbf{x}(i)\right),$$

$$F\!\left(\bigoplus_{i=1}^{r-1} p_i\,\mathbf{x}(i)\right) \geq \sum_{S\subseteq[r-1]} \left(\prod_{i\in S}\beta_\gamma(p_i)\prod_{i\notin S}\big(1-\beta_\gamma(p_i)\big)\right) F\!\left(\bigoplus_{i\in S}\mathbf{x}(i)\right).$$

注意

$$G_2\!\left(\bigoplus_{i\in S}\mathbf{x}(i)\right) = F\!\left(\bigoplus_{i\in S\cup\{r\}}\mathbf{x}(i)\right).$$

代入 将这些展开式代入（338）的右侧，我们得到对所有 $S\subseteq[r-1]$ o f:

- 项，其因子为 $\beta_\gamma(p_r)$，值为 $F\big(\bigoplus_{i\in S\cup\{r\}}\mathbf{x}(i)\big)$;
- 项，其因子为 $(1-\beta_\gamma(p_r))$，值为 $F\big(\bigoplus_{i\in S}\mathbf{x}(i)\big)$。

将第一组按 $S' = S\cup\{r\}$ 重新索引（因此 $S'\subseteq[r]$ 满足 $r\in S'$），将第二组按$S' = S$ 重新索引（因此 $S'\subseteq[r]$ 满足 $r\notin S'$）。于是每个 $F\big(\bigoplus_{i\in S'}\mathbf{x}(i)\big)$ 前的系数为

$$\prod_{i\in S'}\beta_\gamma(p_i)\prod_{i\notin S'}\big(1-\beta_\gamma(p_i)\big),$$

这恰好给出了对所有 $S'\subseteq[r]$的所需混合不等式。归纳证明至此完成。

□

推论 E.2. 设 $F\colon [0,1]^N\to\mathbb{R}_{\geq 0}$ 非负且为 $\gamma$-弱 DR 次模，并满足 $0<\gamma\leq 1$。固定整数 $r,h\geq 1$。对每个 $i\in[h]$，令 $\mathbf{x}^{(i)}(1)\ldots,\mathbf{x}^{(i)}(r)\in[0,1]^N$和 $\mathbf{b}(i)\in[0,1]^N$满足 $\sum_{i=1}^{h}\mathbf{b}(i)=\mathbf{1}$（逐坐标），并令 $p_1,\ldots,p_r\in[0,1]$。定义

$$\beta_\gamma(p) \;:=\; \frac{\gamma^2 p}{1-p+\gamma^2 p} \qquad (p\in[0,1]).$$

*Then*

$$F\!\left(\sum_{i=1}^{h}\mathbf{b}(i)\;\odot\;\bigoplus_{j=1}^{r}\big(p_j\,\mathbf{x}^{(i)}(j)\big)\right) \;\geq\; \sum_{S\subseteq[r]}\left(\prod_{j\in S}\beta_\gamma(p_j)\prod_{j\notin S}\big(1-\beta_\gamma(p_j)\big)\right) F\!\left(\sum_{i=1}^{h}\mathbf{b}(i)\;\odot\;\bigoplus_{j\in S}\mathbf{x}^{(i)}(j)\right).$$

证明. 在 $h$ 个块上定义 $G\colon [0,1]^{Nh} \to \mathbb{R}_{\geq 0}$ 如下

$$G\big(\mathbf{c}(1),\dots,\mathbf{c}(h)\big) \;:=\; F\left(\sum_{i=1}^{h} \mathbf{b}(i)\odot \mathbf{c}(i)\right).$$

由于 $F \geq 0$ 且映射（$\mathbf{c}(1)$..., $\mathbf{c}(h)$ $\mapsto \sum_{i=1}^{h} \mathbf{b}(i) \odot \mathbf{c}(i)$ 在坐标上非负且为线性映射，其中 $\sum_i \mathbf{b}(i) = \mathbf{1}$，由引理 2.3 （逐块应用）可知， $G$ 也非负，且在 $\gamma$上是 $[0,1]^{Nh}$-弱 DR 次模函数。

对每个 $j \in [r]$，定义块向量

$$\mathbf{x}(j) \;:=\; \big(\mathbf{x}^{(1)}(j),\, \mathbf{x}^{(2)}(j),\, \dots,\, \mathbf{x}^{(h)}(j)\big) \;\in\; [0,1]^{Nh}.$$

Then

$$G\left(\bigoplus_{j=1}^{r} p_j\,\mathbf{x}(j)\right) \;=\; F\left(\sum_{i=1}^{h} \mathbf{b}(i)\odot \bigoplus_{j=1}^{r} p_j\,\mathbf{x}^{(i)}(j)\right),$$

这恰好是所需不等式的左侧。

将引理 E.1 应用于 $G$，输入为$\mathbf{x}(1)$..., $\mathbf{x}(r)$，系数为 $p_1$，..., $p_r$：

$$G\left(\bigoplus_{j=1}^{r} p_j\,\mathbf{x}(j)\right) \;\geq\; \sum_{S\subseteq[r]} \left(\prod_{j\in S}\beta_\gamma(p_j)\prod_{j\notin S}\big(1-\beta_\gamma(p_j)\big)\right) G\left(\bigoplus_{j\in S}\mathbf{x}(j)\right).$$

最后，注意对每个 $S \subseteq [r]$,

$$G\left(\bigoplus_{j\in S}\mathbf{x}(j)\right) \;=\; F\left(\sum_{i=1}^{h} \mathbf{b}(i)\odot \bigoplus_{j\in S}\mathbf{x}^{(i)}(j)\right),$$

因此代入 将此恒等式代入上式右端，恰好得到所声称的不等式 y.

□