arXiv:2602.04479v1 [math.OC] 4 Feb 2026

复杂度　混合仿射约束下的去中心化优化

Demyan Yarmoshik [1] Nhat Trung Nguyen [1] Alexander Rogozin [1] Alexander Gasnikov [1 2 3]

摘要

本文研究带有混合仿射等式约束的凸函数去中心化优化问题，这些约束涉及局部变量和全局变量。网络中不同节点的全局变量约束可能各不相同，而局部变量则受耦合约束和节点特定约束的限制。这类问题建模常见于机器学习应用，包括联邦学习和多任务学习，也常见于资源分配和分布式控制。我们在光滑与非光滑假设下分析该问题，并考虑强凸和一般凸目标函数。我们的主要贡献是针对光滑、强凸情形提出一种最优算法，其收敛率达到已有的复杂度下界。对于其余情形，我们还提供了近乎最优的方法。

1. 引言

我们研究分布式优化问题，其中目标函数分解到多个计算节点上，决策变量同时受到局部和全局仿射约束。具体而言，我们考虑如下形式的问题

$$\min_{x_1,\dots,x_n,\tilde{x}} \sum_{i=1}^{n} f_i(x_i,\tilde{x}) \qquad \text{(P)}$$
$$\text{s.t.} \sum_{i=1}^{n}(A_i x_i - b_i) = 0,\ C_i x_i = c_i,\ \widetilde{C}_i\tilde{x} = \tilde{c}_i.$$

有两组变量： $x_i \in \mathbb{R}^{d_i}$ 对每个节点而言都是独立的，而 $\tilde{x} \in \mathbb{R}^d$ 则是所有~节点共有的。每个节点在本地拥有 x_i、 A_i、 b_i、 C_i、 c_i、 C_i、 $\tilde{c}_i$ 和 f_i，其中 $A_i \in \mathbb{R}^{m\times d_i}$、 $b_i \in \mathbb{R}^m$、 $C_i \in \mathbb{R}^{p_i\times d_i}$、 $c_i \in \mathbb{R}^{p_i}$、 $\widetilde{C}_i \in \mathbb{R}^{\tilde{p}_i\times \tilde{d}}$、 $\tilde{c}_i \in \mathbb{R}^{\tilde{p}_i}$ 和 f_i 是凸函数。每个节点可以利用其矩阵执行矩阵-向量乘法，并计算 f_i的函数值和梯度。

节点（智能体）组成一个计算网络 $\mathcal{G} = (\mathcal{V}, \mathcal{E})$，其中 $\mathcal{V} = \{1,\dots,n\}$ 是顶点集合， $\mathcal{E}$ 是边集合。每个智能体都可以与通过一条边与其相连的节点交换信息。（P）中的约束可分为三类:

耦合约束。 这些约束将多个节点上的变量联系起来，需要节点间协调才能满足。主要的耦合约束为

$$\sum_{i=1}^{n}(A_i x_i - b_i) = 0, \qquad (1)$$

它们共同约束变量 x_i。

局部约束。 这些约束仅取决于每个节点 i处的局部变量和数据。具体而言，

$$C_i x_i = c_i, \qquad (2)$$

这些约束可由节点 i 独立执行，无需与其他节点协调。它们通常表示特定于各节点的局部限制或运行限制。

共享变量约束。 此类约束施加于共享变量 $\tilde{x}$。这些约束的形式为

$$\widetilde{C}_i\tilde{x} = \tilde{c}_i, \qquad (3)$$

变量 $\tilde{x}$ 在所有节点处都相同，但约束~中的矩阵 C_i 各不相同。

如果问题包含耦合约束、局部约束和共享变量约束的组合，我们就称之为具有混合约束的问题。我们的研究主要受到关于分布式但集中式算法的经典工作的启发（Boyd 等人，2011年），该工作展示了如何通过调整仿射约束的重新表述，实现各类问题的分解与分布式优化。我们的主要目标是将这种灵活性引入去中心化优化，并开发理论工具，以比较不同重新表述的复杂度。

形如（P）的问题在若干机器学习问题中均有应用，包括横向与纵向联邦学习、分布式多任务学习以及分布式能源系统的控制。

[1]MIRIAI，俄罗斯莫斯科 [2]Innopolis University，俄罗斯Innopolis [3]HDI Lab，HSE University，俄罗斯莫斯科。通讯作者：Demyan Yarmoshik <yarmoshik.d@miriai.org>。

预印本。*2026*年*2*月*5*日。

横向联邦学习（横向联邦学习，或共识优化）。设各节点局部持有的变量通过共识约束相连，即 $x_1 = \ldots = x_n$。我们便得到共识优化问题。

$$\begin{aligned} \min_{x_1,\ldots,x_n} \ & \sum_{i=1}^n f_i(x_i) \\ \text{s.t.} \ & x_1 = \ldots = x_n. \end{aligned} \tag{4}$$

这是一个标准的去中心化优化（或横向联邦学习）问题表述（Boyd 等人，2011年；Yang 等人，2019；Kairouz 等人，2021；Nedi´c 和 Ozdaglar，2009；Scaman 等人，2017年；Kovalev 等人，2021b；a；Li 和 Lin，2021）。训练数据按样本分布在各个计算实体之间，即每个节点持有数据集的一部分。在优化过程中，各智能体旨在保持模型权重相等 x_i。

耦合约束能够表达共识约束：例如，循环等式 $x_1 - x_2 = 0$，$x_2 - x_3 = 0$，……，$x_n - x_1 = 0$ 可以写成 $\sum_{i=1}^n A_i x_i = 0$，其中 $x_i \in \mathbb{R}^d$，$A_1 = \begin{pmatrix} I_d & 0_d & \cdots & 0_d & -I_d \end{pmatrix}^\top$ 依此类推。不过，有趣的是，所有将共识约束黑盒归约为耦合约束的方法，经证明对一阶去中心化算法均无效，因为任意选择矩阵 A_i 都会使通信复杂度至少增加 $\sqrt{n}$ 倍，相较于最优的专用共识优化算法（Yarmoshik 等人，2024b，附录 A）。这一事实促使我们在（P）中明确区分共识变量与耦合约束变量。

纵向联邦学习（VFL）与数据按样本分布在各节点之间的共识优化不同，在纵向联邦学习中，数据按特征分布（Liu 等，2024年；Chen 等，2020）。每个节点对应一个持有其局部权重子集 X_i和特征子矩阵 F_i的参与方。在深度纵向联邦学习中，每个参与方将其局部特征转换为中间表示，再由共享的“顶层”模型处理。一个简单的实例是将参与方侧映射设为线性，$H_i := F_i X_i \in \mathbb{R}^{N \times m}$其中 $X_i \in \mathbb{R}^{d_i \times m}$，聚合 $Z := \sum_{i=1}^n H_i$，并进行预测 $\widehat{y} = g(Z; X)$，其中顶层模型参数为 X。由此得到混合约束表述

$$\begin{aligned} \min_{Z,\widetilde{X},X_1,\ldots,X_n} \ & \ell(g(Z,\widetilde{X})) + \sum_{i=1}^n r_i(X_i) + r(\widetilde{X}) \\ \text{s.t.} \ & \sum_{i=1}^n F_i X_i = Z, \end{aligned}$$

其中 r_i，r 是正则化项。注意，仿射约束 $\sum_{i=1}^n F_i X_i - IZ = 0$ 是耦合约束的一种特殊情况。论文（Vepakomma 等人，2018年；Xie 等人，2024年）研究了类似的设置，但对中间表示使用了非线性映射。

完全去中心化实现可以通过复制顶层模型参数并划分样本维度以实现计算并行，从而避免设置专门的“顶层”节点。将每个参与方特征矩阵划分为垂直分块 $F_i =$列$(F_{i,1}, \ldots\ldots, F_{i,n})$，这会相应地将 $H_i =$划分为列$(F_{i,1}X_i, \ldots\ldots, F_{i,p}X_i)$。引入局部聚合表示 Z_i 以及顶层模型参数的局部副本$\widetilde{\ }$。X_i那么，纵向联邦学习可以写成一个具有混合耦合约束和共识约束的问题，可直接纳入（P）：

$$\begin{aligned} \min_{\substack{Z_1,\ldots,Z_n \\ \widetilde{X}_1,\ldots,\widetilde{X}_n \\ X_1,\ldots,X_n}} \ & \sum_{i=1}^n \ell_i\left(g(Z_i,\widetilde{X}_i)\right) + \sum_{i=1}^n r_i(X_i) + \frac{1}{n}\sum_{i=1}^n r(\widetilde{X}_i) \\ \text{s.t.} \ & \sum_{i=1}^n F_{i,j} X_i = Z_j, \quad j = 1,\ldots,n, \\ & \widetilde{X}_1 = \ldots = \widetilde{X}_n, \end{aligned}$$

其中，ℓ_i 汇总了第 i个样本块上的损失，而共识约束则确保顶层模型共享。

其他可归结为（P）的问题形式包括全局与局部模型的混合（Zhang 等人，2015；Hanzely&Richt´arik，2020；Hanzely 等人，2020）、分布式多任务学习（Wang 等人，2016）以及联邦自监督学习（Makhija 等人，2022）。我们在附录 B中讨论这些问题。

相关工作。问题（P）最常见的特例是共识约束和耦合约束。这两种情形在文献中都已得到广泛研究。

共识约束的形式为 $x_1 = \ldots = x_n$，当仅存在共享变量 $\tilde{x}$ 时，它是（P）的一个特例。去中心化共识优化的研究源于相关工作（Nedi´c & Ozdaglar, 2009; Boyd 等人, 2011），随后发展到建立复杂度下界和最优一阶算法（Scaman 等，2017 年; Kovalev 等人，2020），并推广到时变图（Nedic et al., 2017; Kovalev 等人, 2021b;a; Li & Lin, 2021）。随机变化网络上的推广也见于相关研究（Koloskova et al., 2020; 2021）。针对非光滑问题的算法见于（Scaman 等人，2018; Dvinskikh 和 Gasnikov，2021a; Gorbunov 等, 2019; Kovalev 等人，2024）。

共识优化源于大规模模型训练，即联邦学习（Kairouz 等，2019年；McMahan 等，2017年；Lian 等，2017年）。

耦合约束问题最初出现在分布式电力系统的控制中。不同变体

及相应工程问题表述的方法已在（Necoara et al., 2011年; Necoara & Nedelcu, 2014; 2015）中得到研究。耦合约束的一个特例是资源分配问题（Doan & Olshevsky, 2017年; Li et al., 2018; Nedi´c et al., 2018）。除了耦合等式约束之外，研究者还探讨了非线性不等式约束（Liang et al., 2019; Gong & Zhang, 2023; Wu et al., 2022）、局部约束（Nedic et al., 2010; Zhu & Martinez, 2011）以及局部函数的受限域（Wang & Hu, 2022; Liang et al., 2019; Nedi´c et al., 2018; Gong & Zhang, 2023; Zhang et al., 2021; Wu et al., 2022）。关于时变网络的推广见（Zhang et al., 2021; Nedi´c et al., 2018）。

耦合约束优化的两种主要方法包括近端ADMM型算法和梯度方法。近端方法已在（Boyd 等人，2011年; Chang，2016; Falsone 等人，2020; Wu 等人，2022）中得到研究，而（Gong & Zhang，2023）研究了一种采用不精确近端算子的算法。当目标函数结构适合近端处理时，计算近端映射在计算上是可行的。因此，近端方法主要应用于电力系统控制领域，因为该领域的损失函数足够简单。机器学习领域的情况则不同，需要使用一阶方法（Doan & Olshevsky，2017年; Nedi´c 等人，2018; Yarmoshik 等人，2024b）。复杂度下界和一阶最优方法见（Yarmoshik 等人，2024b）。

耦合约束对于机器学习研究的价值主要体现在其对纵向联邦学习的应用（Chen 等人，2020; Liu 等人，2024年; Stanko 等人，2026）。

混合约束。据我们所知，文献对混合约束的研究很少。开创性工作（Boyd 等人，2011年）提出了一般形式的共识问题，其中每个局部变量 x_i 都是全局变量 x 的子向量。然而，如果局部函数不依赖于 x，这种约束就可以写成耦合约束（而非混合约束）的形式。论文（Du 与 Meng，2023; 2025）提出了聚合优化，其中每个局部函数都依赖于全局决策向量 x 以及全局向量部分 x_i，并针对耦合不等式约束进行优化。该方法采用一阶方法，并实现了线性收敛速率。其问题可以写成形式（P），但其中的局部约束将局部个体变量与共享变量联系起来。本文研究的是一种略有不同的设定。

我们的贡献。

- 提出一类具有混合仿射约束的去中心化问题，它推广了大多数常见的共识约束设置和耦合约束设置，并使它们能够高效结合。
- 一阶算法的紧确复杂度分析（并给出相应的新下界，适用于光滑且强凸情形），以及针对非光滑且非强凸情形据称近乎最优的上界。
- 对分布式仿射约束与通信网络的联合结构如何影响优化性能进行量化。

概率	预言机	互补
共识约束 (4)	梯度	$\sqrt{\kappa_f}$
	通信	$\sqrt{\kappa_f}\sqrt{\kappa_W}$
	论文	（SCAMAN 等人。，2017年）
耦合约束 (1)	梯度	$\sqrt{\kappa_f}$
	矩阵 A	$\sqrt{\kappa_f}\sqrt{\widehat{\kappa}_A}$
	通信	$\sqrt{\kappa_f}\sqrt{\widehat{\kappa}_A}\sqrt{\kappa_W}$
	论文	(YARMOSHIK 等人., 2024年B)
共享 VAR. 约束 (S)	梯度	$\sqrt{\kappa_f}$
	矩阵 $\widetilde{C}$	$\sqrt{\kappa_f}\sqrt{\widehat{\kappa}_{\widetilde{C}^\top}}$
	通信	$\sqrt{\kappa_f}\sqrt{\widehat{\kappa}_{\widetilde{C}^\top}}\sqrt{\kappa_W}$
	论文	本论文，定理。 4.1
局部 VAR. 约束 (C&L)	梯度	$\sqrt{\kappa_f}$
	矩阵 A	$\sqrt{\kappa_f}\sqrt{\widetilde{\kappa}_{AC}}$
	矩阵 C	$\sqrt{\kappa_f}\sqrt{\widetilde{\kappa}_{AC}}\sqrt{\kappa_C}$
	通信	$\sqrt{\kappa_f}\sqrt{\widetilde{\kappa}_{AC}}\sqrt{\kappa_W}$
	论文	本文 , TH. 4.5
混合约束 (P)	梯度	$\sqrt{\kappa_f}$
	矩阵 A	$\sqrt{\kappa_f}\sqrt{\widetilde{\kappa}_{AC}}$
	矩阵 C	$\sqrt{\kappa_f}\sqrt{\widetilde{\kappa}_{AC}}\sqrt{\kappa_C}$
	矩阵 $\widetilde{C}$	$\sqrt{\kappa_f}\sqrt{\widehat{\kappa}_{\widetilde{C}^\top}}$
	通信	$\sqrt{\kappa_f}\left(\sqrt{\widetilde{\kappa}_{AC}}+\sqrt{\widehat{\kappa}_{\widetilde{C}^\top}}\right)\sqrt{\kappa_W}$
	论文	本文 , TH. 4.6

表 *1*。 具有仿射约束的去中心化光滑强凸优化的收敛率。混合条件数记为 $\widehat{\kappa}_A$， $\widehat{\kappa}_{C^\top}$（见定义 3.4）。条件数 $\widetilde{\kappa}_{AC}$ 定义见 4.2。省略项 $\log\left(\frac{1}{\varepsilon}\right)$。

符号说明。对于向量 $x_1,\ldots,x_n$，我们将按列堆叠的向量记为 $\mathbf{x}=\mathrm{col}\,(x_1,\ldots,x_n)$。我们用 $\|\cdot\|$ 表示欧几里得范数，用 $\langle\cdot,\cdot\rangle$ 表示标准内积。符号 $\otimes$ 表示克罗内克积。对于由不同智能体堆叠而成的矩阵，我们使用粗体字母表示。这些矩阵包括分块矩阵以及通过克罗内克积得到的矩阵，例如 $\mathbf{A}=\mathrm{diag}(A_1,\ldots,A_n)$、$\mathbf{A}'=(A_1\ldots A_n)$。矩阵 A 的最大和最小正奇异值分别记为$\sigma_{\max}(A)$和 $\sigma_{\min^+}(A)$。对称矩阵 A 的最大和最小正特征值分别记为 $\lambda_{\max}(A)$和 $\lambda_{\min^+}(A)$。对于任意矩阵A，我们定义 $\kappa_A=\frac{\sigma^2_{\max}(A)}{\sigma^2_{\min^+}(A)}$。当涉及一组矩阵 A时，我们将该参数视为针对块对角矩阵定义，即 $\kappa_A=\max_i\sigma^2_{\max}(A_i)/\min_i\sigma^2_{\min^+}(A_i)$。对于给定的线性子空间 $\mathcal{L}$，我们还定义相应的投影算子 $\mathbf{P}_{\mathcal{L}}$。

论文结构。我们首先在第 2 节概述带仿射约束的凸优化方法。随后，我们在以下各节中将第 2 节介绍的算法应用于问题（P）相应特殊情形的具体重新表述。第 3 节概述该领域已有的研究结果；第 4 节给出光滑强凸情形的复杂度结果（最优算法及其对应下界），并在表 1 中汇总。类似地，第 5 节介绍非光滑且非强凸情形的研究结果。最后，第 6 节给出总结性评述。

2. 带仿射约束的优化

本文考虑的所有问题均可表述为带仿射等式约束的凸优化问题：

$$\min_{u\in\mathcal{U}} G(u) \quad \text{s.t.} \quad Bu = b, \tag{5}$$

其中 $G\colon \mathcal{U}\to\mathbb{R}$ 是在集合$\mathcal{U}\subseteq\mathbb{R}^d$上定义的凸函数，矩阵 $B\in\mathbb{R}^{s\times d}$ 非零，且 b 属于 B的列空间。根据目标函数 G的结构性质，我们采用不同的优化技术来求解问题（5）。

2.1. 目标函数的假设

我们提出一组假设，以便为几类凸函数推导收敛保证和复杂度界，从而求解问题（5）。

假设 2.1（强凸性）。函数 $G(u)$ 在 $\mathcal{U}$ 上是 μ-强凸的，对于 $\mu\geq 0$，即对于任意 $u,\ u'\in\mathcal{U}$，

$$G(u')\geq G(u)+\langle\nabla G(u),u'-u\rangle+\frac{\mu}{2}\|u'-u\|^2.$$

当 $\mu=0$ 时，函数 G 被称为（非强）凸函数。

假设2.2（光滑性）。函数 $G(u)$ 在 $\mathcal{U}$ 上是 L-光滑的，即它在 $\mathcal{U}$ 上可微，且对于任意$u,\ u'\in\mathcal{U}$，有

$$G(u')-G(u)-\langle\nabla G(u),u'-u\rangle\leq\frac{L}{2}\|u'-u\|^2.$$

假设 2.3（有界次梯度）。函数 G 在 $\mathcal{U}$ 上具有有界次梯度，即存在常数 $M>0$，使得对于任意 $u\in\mathcal{U}$ 和任意 $G'(u)\in\partial G(u)$，

$$\|G'(u)\|\leq M.$$

2.2. 光滑目标函数

具有光滑目标函数（即满足假设 2.2）的问题（5）可使用加速近端交替预测校正算法（APAPC）求解，该算法由 Salim 等人 于 2022 年针对强凸问题提出。当 $\mathcal{U}=\mathbb{R}^d$ 时，这是求解带仿射约束凸优化问题的一种最优方法。为将该方法用于非强凸目标函数，我们引入一种基于正则化的约化技术；由于存在仿射约束，其分析并不常规。APAPC 的理论复杂度结果总结于定理 2.4。

定理 2.4。设 $G(u)$ 满足假设*2.1* 和 *2.2*，其中 $0<\mu<L$。令 $\|u^0-u^*\|^2\leq R^2$ 并引入精度 $\varepsilon>0$。

1. 强凸情形 $\mu>0$ *(Salim 等人，2022，命题 1）*。存在一组适用于算法 *1*的参数，使得经过 $N=O\left(\kappa_B\sqrt{L/\mu}\log(1/\varepsilon)\right)$ 次迭代后，该算法得到 u^N，满足 $\|u^N-u^*\|_2^2\leq\varepsilon$。

2. 凸情形 $\mu=0$（新结果，附录 *D*）。将算法 *1* 应用于正则化函数 $G^\nu(u)=G(u)+\nu/2\|u^0-u\|^2$，其中 $\nu=\varepsilon/R^2$，经过 $N=O\left(\kappa_B\sqrt{LR^2/\varepsilon}\log(1/\varepsilon)\right)$ 次迭代后，我们得到 u^N，其满足 $G(u^N)-G(u^*)\leq\varepsilon$，且 $\|Bu^N-b\|^2\leq O(\sigma_{\max}^2(B)\varepsilon^2)$。

算法 1 APAPC

输入： $u^0\in\mathbb{R}^d$
参数： $\eta,\theta,\alpha>0,\tau\in(0,1),\ N\in\mathbb{N}$
设置 $u_f^0=u^0,\ z^0=0\in\mathbb{R}^d$
对于 $k=0,1,2,\ldots,N-1$ **执行**
  $u_g^k:=\tau u^k+(1-\tau)u_f^k$
  $u^{k+\frac{1}{2}}:=(1+\eta\alpha)^{-1}(u^k-\eta(\nabla G(u_g^k)-\alpha u_g^k+z^k))$
  $z^{k+1}:=z^k+\theta B^\top(Bu^{k+\frac{1}{2}}-b)$
  $u^{k+1}:=(1+\eta\alpha)^{-1}(u^k-\eta(\nabla G(u_g^k)-\alpha u_g^k+z^{k+1}))$
  $u_f^{k+1}:=u_g^k+\frac{2\tau}{2-\tau}(u^{k+1}-u^k)$
结束循环
输出： u^N

2.3. 非光滑目标函数

本节介绍一种求解问题（5）的方法，其中目标函数 G 是非光滑的。按照（Dvinskikh &Gasnikov，2021b；Gorbunov 等人，2019），我们通过引入惩罚目标函数来处理仿射等式约束：

$$\min_{u\in\mathcal{U}} H_r(u):=G(u)+\frac{r}{2}\|Bu-b\|^2, \tag{6}$$

其中 $r>0$ 是惩罚系数。

通过适当选择 r，求解原问题可转化为最小化 H_r（详见附录E）。由于 H_r 由一个非光滑项 $G(u)$ 和一个光滑项组成

二次惩罚，因此我们可以应用滑移技术，分别获得各个分量的复杂度界。具体而言，我们采用梯度滑移算法，见（Lan，2019）。为描述该算法，我们定义 H_r的如下二次模型：

$$\Phi_{H_r}(w,u_1,u_2,u_3,\beta,\eta)=\langle G'(u_1)+rB^\top(Bu_2-b),w\rangle + \frac{\beta}{2}\|u_3-w\|^2+\frac{\beta\eta}{2}\|u_1-w\|^2, \quad (7)$$

其中 β，η为参数，且 w，u_1，u_2，$u_3\in U$。随后，我们推导出用于在非光滑情形下求解问题 (5) 的算法 2 。在每次迭代中，该算法求解类型为 (7) 的子问题。

算法 2 梯度滑动法

输入： $u^0\in U$
参数： $\{\gamma_k\}_{k=1}^\infty$，$\{\eta_k\}_{t=1}^\infty$，$\{\theta_t\}_{t=1}^\infty\subseteq\mathbb{R}_{++}$，$N\in\mathbb{N}$, $\{T_k\}_{k=1}^N\subseteq\mathbb{N}^N$
$\overline{u}^0:=u^0$
对于 $k=1,2,\ldots,N$ **do**
  $u^k:=\gamma_k u^{k-1}+(1-\gamma_k)u^{k-1}$
  $u_0^k:=\tilde{u}_0^k$, $\tilde{u}_0^k:=u^{k-1}$
  对于 $t=1,2,\ldots,T_k$ **执行**
    $u_t^k=$ 使……最小 $\Phi_{H_r}(w,u_{t-1}^k,u^k,u^{k-1},\beta_k,\eta_t)$
    $\tilde{u}_t^k=\theta_t u_t^k+(1-\theta_t)\tilde{u}_{t-1}^k$
  结束循环
  $u^k:=u_{T_k}^k$, $\tilde{u}^k:=\tilde{u}_{T_k}^k$
  $\overline{u}^k:=\gamma_k\tilde{u}^k+(1-\gamma_k)\tilde{u}^{k-1}$
结束循环
输出： $\overline{u}^N$

定理 2.5给出了算法 2 达到 $(\varepsilon,\ \delta)$ 精度所需的梯度计算次数和矩阵乘法次数上界，这意味着$G(u)-G(u^*)\le\varepsilon$ 且 $\|Bu-b\|\le\delta$。详情请参阅附录 E。

定理 2.5(Lan 等人 (2020))。 设 G 满足假设 *2.1* ，其中 $\mu\ge0$；假设*2.3* 中 $M>0$，并设 $\left\|u^0-u^*\right\|^2\le R^2$。给定 $\varepsilon>0$，考虑使用算法(6)求解惩罚问题2。

1. 凸情形（$\mu=0$）。该算法需要$N_B=O\left(\frac{MR}{\varepsilon}\sqrt{\kappa_B}\right)$ 次乘法运算，分别与 B、$B^\top$ 和 $N_{G'}=$ 相乘，并需要 $O\left(\frac{M^2R^2}{\varepsilon^2}+N_B\right)$ 次对 G 的次梯度进行求值，以生成输出 u^N，使得 $G(u^N)-G(u^*)\le\varepsilon$，且$\left\|Bu^N-b\right\|\le O(\varepsilon)$。

2. 强凸情形（$\mu>0$）。通过重启，该算法返回 u^N，满足相同的精度保证，所需的 B乘法运算为 $N_B=O\left(\frac{M}{\sqrt{\mu}\varepsilon}\sqrt{\kappa_B}\right)$ 次，$B^\top$ 和 $N_{G'}=O\left(\frac{M^2}{\mu\varepsilon}+N_B\right)$ 次对 G 的次梯度评估。

2.4. 切比雪夫加速

切比雪夫加速（Salim 等人，2022；Scaman 等，2017 年；Auzinger 和 Melenk，2011 年）最初作为一种理论工具，用于降低光滑共识优化中的通信复杂度；它能有效改善仿射约束矩阵的条件数。切比雪夫加速的思想是将仿射约束 $Bu=b$ 等价地改写为 $Kx=b'$，其中 $K=P_B(B^\top B)$、$b'=P_B(B^\top B)B^\top b$，且 P_B 是一个多项式，使得 $P_B(\sigma_i^2(B))=$ 当且仅当 $\sigma_i(B)=0$。采用适当缩放和平移的切比雪夫多项式后，改写形式的条件数为 $\kappa_K=O(1)$，而矩阵-向量乘法 Kx 可通过O 次对 $B^\top u$ 和 Bu 的矩阵-向量乘法实现，因此无需显式计算并存储矩阵多项式 K。我们将切比雪夫加速的应用记为 $B\to P_B(B^\top B)$。切比雪夫迭代的伪代码见附录 C。

3. 带仿射约束的去中心化优化中的已有结果

本节讨论去中心化优化领域的已知结果及其收敛性质。我们回顾了先前研究中相关的去中心化优化问题表述及其收敛性质。这些表述是问题（P）的特例。

3.1. 符号与假设

去中心化通信通常表示为矩阵与向量的乘法。具体而言，我们假设存在一个通信矩阵 $W\in\mathbb{R}^{n\times n}$，它与通信网络 $\mathcal{G}=(\mathcal{V},\mathcal{E})$相关联，并具有以下性质。

假设 3.1. 该通信矩阵 W 满足：
1. W 是对称半正定的。2. $W_{ij}\ne0$ 当且仅当 $i=j$ 或（i，j）$\in\mathcal{E}$。3. $Wx=0$ 当且仅当 $x_1=\ldots=x_n$。

这种矩阵的一个例子是拉普拉斯矩阵 $L=D-A$，其中 A 是网络 $\mathcal{G}$的邻接矩阵，D 是度矩阵。在全文中，我们将$\mathbf{W}=W\otimes I$记为，其中单位矩阵 I 的维数由上下文决定。

我们考虑去中心化优化中目标函数的以下标准假设。

假设 3.2。对于某个 $L_f\ge\mu_f\ge0$，每个函数 f_i都是 L_f-光滑且 μ_f-强凸的。

当 $\mu_f>0$时，我们将局部函数的条件数记为 $\kappa_f=L_f/\mu_f$。

假设 3.3。每个函数 f_i 的次梯度有界，

常数为 $M_f > 0$，并且对于某个 $\mu_f \geq 0$，它是 μ_f-强凸的。

3.2. 共识优化

此情形对应于不存在仿射等式约束、仅涉及公共变量的场景。当假设 3.1 和 3.2 成立且 $\mu_f > 0$时，梯度调用和通信轮次的最优收敛率分别为 $O\left(\sqrt{\kappa_f}\log(1/\varepsilon)\right)$ 和 $O\left(\sqrt{\kappa_f}\sqrt{\kappa_W}\log(1/\varepsilon)\right)$，如（Kovalev 等人，2020）所确立。

在非光滑且非强凸情形下，即当假设 3.1和3.3成立且 $\mu_f = 0$时，次梯度计算的次数上界为$O\left(M_f^2R^2/\varepsilon\right)$，而通信轮次的数量则由 $O\left(\sqrt{\kappa_W}M_fR/\varepsilon\right)$ 给出，其中 R表示可行域的半径。这些界已在（Scaman 等人，2018）中得到并证明是最优的。

3.3. 相同的局部约束

考虑仅有公共变量 $\tilde{x}$ 和~ ~相同局部约束的情形，即对所有 $i = 1, \ C_i = C$ 和 $\tilde{c}_i = \tilde{c}, \ldots\ldots, n$。在（Rogozin 等人，2022）中，作者在假设 3.1 和 3.2 成立时，为该问题形式推导出了以下上界： $O(\sqrt{\kappa_f}\log(1/\varepsilon))$ 次梯度调用，$\sim N_{\widetilde{C}} = O(\sqrt{\kappa_f}\sqrt{\kappa_{\widetilde{C}}}\log(1/\varepsilon))$ 次乘以 C和$\widetilde{C}^\top$，以及$N_{\overline{W}} = O\left(\sqrt{\kappa_f}\sqrt{\kappa_W}\log(1/\varepsilon)\right)$ 次通信。正如我们将在第 4.1 节中所展示的，当局部约束不相同时，复杂度界会发生显著变化。

3.4. 耦合约束

现在考虑每个节点在本地持有一部分仿射约束 A_i的情形。

$$\min_{x_1\in\mathbb{R}^{d_1},\ldots,x_n\in\mathbb{R}^{d_n}} \sum_{i=1}^n f_i(x_i) \quad \text{s.t.} \quad \sum_{i=1}^n (A_ix_i - b_i) = 0. \tag{8}$$

对于光滑且强凸目标函数，这个问题已在（Yarmoshik 等人，2024b）中得到研究。作者引入了一种新的条件数，用以刻画耦合约束下的收敛率。

定义 3.4。对于一组矩阵（$B_1, \ldots\ldots, B_n$），引入一个交互矩阵

$$S_B = \frac{1}{n}\sum_{i=1}^n B_iB_i^\top, \tag{9}$$

以及混合条件数

$$\widehat{\kappa}_B = \frac{\max\limits_{i=1,\ldots,n}\lambda_{\max}(B_iB_i^\top)}{\lambda_{\min^+}(S_B)} = \frac{\max\limits_{i=1,\ldots,n}\lambda_{\max}(B_iB_i^\top)}{\lambda_{\min^+}\left(\frac{1}{n}\sum_{i=1}^n B_iB_i^\top\right)}.$$

注：3.5。 请注意， $\widehat{\kappa}_B$ 的转置顺序不确定性很重要。令

$$\widehat{\kappa}_{B^\top} = \frac{\max\limits_{i=1,\ldots,n}\lambda_{\max}(B_i^\top B_i)}{\lambda_{\min^+}(S_{B^\top})} = \frac{\max\limits_{i=1,\ldots,n}\lambda_{\max}(B_i^\top B_i)}{\lambda_{\min^+}\left(\frac{1}{n}\sum_{i=1}^n B_i^\top B_i\right)}.$$

一般而言 $\widehat{\kappa}_{B^\top} \neq \widehat{\kappa}_B$。例如，令 $B_i = e_i$，其中 e_i是第 i 个坐标向量。则 $\widehat{\kappa}_B = n$且 $\widehat{\kappa}_{B^\top} = 1$。

注3.6。请注意，当矩阵相等时 $B_1 = \ldots = B_n = B$混合条件数 $\widehat{\kappa}_B$自然退化为通常条件数 κ_B，因此 $\widehat{\kappa}_B = \widehat{\kappa}_{B^\top} = \kappa_B$。

当假设 3.1 和 3.2在 $\mu_f > 0$下成立时，(Yarmoshik 等人，2024b) 中达到的最优收敛率为 $O\left(\sqrt{\kappa_f}\log(1/\varepsilon)\right)$ 次梯度调用，$O(\sqrt{\kappa_f}\sqrt{\widehat{\kappa}_A}\log(1/\varepsilon))$ 次矩阵-向量乘法 $\mathbf{A}$,$\mathbf{A}^\top$ 以及 $O(\sqrt{\kappa_f}\sqrt{\widehat{\kappa}_A}\sqrt{\kappa_W}\log(1/\varepsilon))$ 次通信轮次。

4. 光滑强凸问题的最优算法

在推导问题 (P) 的算法和复杂度界之前，我们先考虑两种特殊情形。第一种情形是公共变量仅受局部约束，第二种情形是同时存在耦合约束和局部约束。

4.1. 共享变量约束

我们考虑每个节点都受到其自身针对全局变量的局部仿射约束的情形。具体而言，我们研究如何将 (Rogozin 等人, 2022) 中的表述从相同约束推广到非相同约束。该优化问题可写为

$$\min_{\tilde{x}\in\mathbb{R}^{\tilde{d}}} \sum_{i=1}^n f_i(\tilde{x}) \quad \text{s.t.} \quad \widetilde{C}_i\tilde{x} = \tilde{c}_i \quad \forall i \in \{1,\ldots,n\}. \tag{S}$$

引入变量的局部副本，可以将问题重新表述为分块矩阵形式

$$\min_{\tilde{\mathbf{x}}\in\left(\mathbb{R}^{\tilde{d}}\right)^n} F(\tilde{\mathbf{x}}) := \sum_{i=1}^n f_i(\tilde{x}_i) \ \text{ s.t. } \ \mathbf{W}\tilde{\mathbf{x}} = 0, \ \widetilde{\mathbf{C}}\mathbf{x} = \tilde{\mathbf{c}}, \tag{10}$$

其中 ~$\mathbf{x}$ $= \text{col}\{\tilde{x}_1,\ldots,\tilde{x}_n\}$、~$\mathbf{c}$ $= \text{col}\{\tilde{c}_1,\ldots,\tilde{c}_n\}$, ~$\mathbf{W}$ $= W \otimes I_d$对共识进行编码，矩阵 $\mathbf{C} = \text{diag}(\widetilde{C}_1, \ldots, \widetilde{C}_n) \in \mathbb{R}^{\tilde{p}\times\tilde{d}n}$汇集局部约束，其中 $\tilde{p} = \sum_{i=1}^n \tilde{p}_i$。

约束矩阵具有块结构：~$\mathbf{B}^\top$ $= (\sim\mathbf{C}^\top \ \gamma\mathbf{W})$。根据（Yarmoshik 等人，2024b）中的引理 2，我们选择适当的缩放参数 γ 并

推导定理 4.1 中针对共享变量的含仿射等式约束的去中心化优化问题的上界。

定理 4.1 (新结果， 附录 G)。 将算法 *1* 应用于问题 (10) 在 $\mathbf{W} \to P_W(\mathbf{W})$ 以及 $\widetilde{\mathbf{B}} \to P_C(\widetilde{\mathbf{B}}^\top \widetilde{\mathbf{B}})$ 条件下的重新表述，可得到一种复杂度如表 *1*所示的方法。对于具有局部约束的问题，该下界在一个自然定义的去中心化一阶算法类中是最优的。

与耦合情形的一个显著区别在于，无需正则化即可保证强凸性，而且预处理器的设计有更大的自由度。具体而言，$\mathbf{C}$ 可以独立于 $\mathbf{W}$ 进行预处理；而对于耦合约束，$\mathbf{A}$ 和 $\mathbf{W}$ 则不可避免地相互纠缠。事实上，对于任意不含常数项的多项式 P，只要对 $\widetilde{\mathbf{C}}^\top \widetilde{\mathbf{C}}$ 的所有非零特征值 λ 都满足 $P(\lambda) \neq 0$，则 $(\widetilde{\mathbf{C}}^\top \mathbf{W})^\top x = 0$ 当且仅当 $(P(\widetilde{\mathbf{C}}^\top \widetilde{\mathbf{C}})^\top \mathbf{W})^\top x = 0$。这一性质使得预处理策略更加丰富。尽管如此，尽管有这种额外的灵活性，上文得到的上界最终仍是紧的。可以用类似于耦合约束下界论证的方式证明其最优性（见附录 G）。

4.2. 耦合约束与局部约束

个体变量通过耦合约束相互关联，并受局部约束局部限制。

$$\min_{x_1 \in \mathbb{R}^{d_1}, \ldots, x_n \in \mathbb{R}^{d_n}} \sum_{i=1}^n f_i(x_i) \qquad \text{(C\&L)}$$
$$\text{s.t.} \sum_{i=1}^n (A_i x_i - b_i) = 0,\ C_i x_i = d_i.$$

对于这一表述，仿射约束的去中心化友好矩阵为 $\mathbf{B} = \begin{pmatrix} \mathbf{A} & \alpha \mathbf{W} \\ \beta \mathbf{C} & 0 \end{pmatrix}$ 其中 $\mathbf{A} = \operatorname{diag}(A_1, \ldots, A_n)$ 且 $\mathbf{C} = \operatorname{diag}(C_1, \ldots, C_n)$，参见附录 F.2。

现在我们为这一设置引入自然的复杂度参数。

定义 4.2。 考虑两个矩阵集（B_1，…，B_n）和（D_1，…，D_n）。令 $\mathbf{B}'$ =（B_1，…，B_n）和 $\mathbf{D}$ =diag（D_1，…，D_n）。我们定义

$$\widetilde{\mu}_{BD} := \frac{1}{n} \sigma^2_{\min^+}(\mathbf{B}' \mathbf{P}_{\ker \mathbf{D}}),$$
$$\widetilde{\kappa}_{BD} := \begin{cases} \dfrac{\max\limits_{i=1,\ldots,n} \lambda_{\max}(B_i B_i^\top)}{\widetilde{\mu}_{BD}}, & \text{if } \widetilde{\mu}_{BD} > 0, \\ 1, & \text{if } \widetilde{\mu}_{BD} = 0. \end{cases}$$

为方便起见，引入 $\widetilde{\kappa}_{BD} = 1$ 在 $\widetilde{\mu}_{BD} = 0$ 情况下的定义，因为在这种情况下，$\widetilde{\kappa}_{BD}$ 不会影响复杂度（见引理 4.4 和定理4.5）。

注 4.3。 注意，如果 $\widetilde{\mu}_{BD} > 0$，那么我们有 $\widetilde{\kappa}_{BD} \leq \widehat{\kappa}_B$。事实上，对于任意矩阵 M 以及维数相容的线性子空间 $\mathcal{L}$，都有 $\ker^\perp M\mathbf{P}_{\mathcal{L}} = \ker^\perp M \cap \mathcal{L}$，因此 $\sigma_{\min^+}(M\mathbf{P}_{\mathcal{L}}) = \min_{h \in \ker^\perp M, h \in \mathcal{L}} \|Mh\|/\|h\| \geq \min_{h \in \ker^\perp M} \|Mh\|/\|h\| = \sigma_{\min^+}(M)$。当与 $\sigma_{\min^+}(\mathbf{B}')$ 对应的奇异向量属于 $\ker\mathbf{D}$ 时，等号成立。特别地，当 $\mathbf{D} = 0$ 时，我们有 $\widetilde{\kappa}_{BD} = \widehat{\kappa}_B$。

下面的引理给出了 κ_B 的上界，这对于复杂度上界至关重要。它还刻画了分布式仿射约束与通信网络之间的互连如何影响优化性能。

引理 4.4 (新， 附录 H)。存在常数 $\alpha > 0$和 $\beta > 0$，使得条件数 $\kappa_{\mathbf{B}}$ 满足

$$\kappa_{\mathbf{B}} = \begin{cases} O\left(\widetilde{\kappa}_{AC}\kappa_W + \widetilde{\kappa}_{AC}\kappa_C\right), & \widetilde{\mu}_{AC} > 0, \\ O\left(\kappa_W + \kappa_C\right), & \widetilde{\mu}_{AC} = 0. \end{cases} \qquad (12)$$

定理 4.5(新， 附录 I)。将算法 *1* 应用于问题的重新表述(C&L)，其中 $\mathbf{C} \to P_C(\mathbf{C}^\top \mathbf{C})$，$\mathbf{W} \to P_W(\mathbf{W})$，然后 $\mathbf{B} \to P_B(\mathbf{B}^\top \mathbf{B})$，我们得到一种复杂度如表 *1* 所示的方法。这些复杂度界因相应的下界而达到最优。

当没有局部约束（$\mathbf{C} = 0$）时，我们有 $\widetilde{\kappa}_{AC} = \widehat{\kappa}_A$（见备注 4.3）。在这种情况下，本小节得到的复杂度界与定理 F.4中的界相同，从而有力地推广了（Yarmoshik 等人，2024b）的主要结果。

4.3. 混合约束

我们引入 K =diag(B,B)、z =col(x, y)、v = col(b, c, 0)，并定义可行集 $\mathcal{Z} = \mathbb{R}^d \times (\mathbb{R}^m)^n \times (\mathbb{R}^d)^n$。然后，问题 (P) 可重新表述为

$$\min_{\mathbf{z} \in \mathcal{Z}} G(\mathbf{z}) := \sum_{i=1}^n f_i(x_i, \tilde{x}_i) \quad \text{s.t.} \quad \mathbf{K}\mathbf{z} = \mathbf{v}. \qquad (13)$$

我们现在汇集此前的全部结果，并说明带有共享变量约束的去中心化优化（第4.1节）以及带有耦合约束和局部约束的去中心化优化（第4.2节）的复杂度界如何结合进表述为（P）的一般混合设置中。定理 4.6给出了问题（P）的上界。

定理 4.6(新增， 附录 J)。 应用算法 *1* 求解问题 (13)，可得到一种方法，其复杂度如表 *1*所示。在局部约束相同的情况下，

~矩阵 C_i 相等时，复杂度 $N_{\widetilde{C}}$ 和 N_W 变为

$$N_{\widetilde{C}} = O\left(\sqrt{\kappa_f}\sqrt{\kappa_{\widetilde{C}}}\log\left(\frac{1}{\varepsilon}\right)\right)$$
$$N_W = O\left(\sqrt{\kappa_f}\sqrt{\widetilde{\kappa}_{AC}}\sqrt{\kappa_W}\log\left(\frac{1}{\varepsilon}\right)\right).$$

因此，当耦合约束和局部约束同时作用于不同的变量集合时，我们的统一框架可得出紧的复杂度界。这些结果涵盖了纯耦合和纯局部两种情形，表明整体难度由**A**、**C**的联合条件数以及网络拓扑 W共同决定。这些界是最优的，其依据是：在局部约束相同的情况下，定理F.2 和F.4 的下界与一般情形下定理4.1 和F.4 的下界相结合。

5. 非强凸与非光滑优化的扩展

5.1. 光滑凸优化

利用2.4节中的定理2，我们推导出光滑、非强凸情形下问题（P）的迭代复杂度上界，如定理5.1所述。

定理 5.1（**新，**附录K）。假设*3.1*和*3.2* 成立，且 $L_f > \mu_f = 0$。考虑对问题*1* 应用带正则化技术的算法（13），该问题是问题（P）的重新表述。则所得方法的复杂度如表*2*所示。

由此定理，我们可以直接得到其他类型约束下的结果，汇总于表 2中。解释见附录 K。注意，由于对数因子的存在，这些上界并非最优。

5.2. 非光滑凸优化

类似地，我们推导出了非光滑凸情形下的结果（见表3）。

定理5.2 (新结果，见附录L)。 假设 *3.3* 成立，且 $\mu_f = 0$ 和 $M_f > 0$。对问题(13) 应用带切比雪夫加速的梯度滑动法，可得到一种方法，其复杂度如表*3*所示。

5.3. 非光滑强凸优化

在强凸且非光滑的情形下，我们考虑~有界集上的问题（P），否则，关于 $X_1 \times \cdots \times X_n \times X$的假设 3.3无法成立。记 $\mu_f > 0$， $\mathcal{X} = X_1 \times \cdots \times X_n$。引理 5.3保证了 $\widetilde{\mathcal{X}} = \left(\widetilde{X}\right)^n$的惩罚重构是强凸的。 G

引理 5.3 (新结果， 附录 M.1)。 假设假设*2.1* 在 $\mu_f > 0$ 下成立。令 α 和 ε 满足以下条件：

$$\alpha^2 = \frac{\mu_{\mathbf{A}} + L_{\mathbf{A}}}{\mu_{\mathbf{W}}}, \quad \varepsilon \leq \frac{4r^2\mu_{\mathbf{A}}}{\mu_f}. \tag{14}$$

那么， $G(\mathbf{x}, \mathbf{y}\ \tilde{}\mathbf{x})$的惩罚函数，其中 G 定义于 (13)，在 $\frac{\mu_f}{2}$ 上是$\mathcal{X} \times \mathcal{L}_m^{\perp} \times \widetilde{\mathcal{X}}$-强凸的。

通过综合第 2.3 节、第 3 节以及引理5.3 的结果，我们得出了混合仿射约束下去中心化优化的迭代复杂度上界。结果汇总见表 4。

定理 5.4 (新，附录 M.2)。**若满足假设** *3.3*，且 $\mu_f > 0$ 和 $M_f > 0$成立，则应用带切比雪夫加速的重启式 *GradientSliding*，可得到一种复杂度如表 *4*所示的方法。

6. 结论

我们将去中心化优化领域的一系列结果统一到混合仿射约束的问题表述之下。我们的表述具有一般性，可涵盖水平和垂直联邦学习、分布式系统控制以及分布式机器学习中的其他主题。我们的目标是系统地分析该问题。为此，我们证明了复杂度下界，并针对问题表述的不同变体提供了相应的最优方法。

本文在逻辑上接下来要讨论不同问题类别如何相互转化（例如，见表 1 中的结果）。如果将某个优化问题改写为不同形式，能否降低其复杂度?

另一个未来方向是考虑约束$\sum_{i=1}^n (A_i x_i + \tilde{A}_i \tilde{x}_i - b_i) = 0$。我们假设此类约束能够涵盖新的机器学习表述，即联邦学习与分布式自监督学习中的全局与局部模型的混合。

参考文献

Auzinger, W.与Melenk, J. 大型线性系统的迭代解。维也纳工业大学讲义，2011年。

Boyd, S.、Parikh, N.、Chu, E.、Peleato, B.、Eckstein, J. 等。通过交替方向乘子方法进行的分布式优化与统计学习。机器学习基础与趋势®，3(1):1–122，2011年。

Chang, T.-H. 一种用于多智能体约束优化的近端对偶共识 ADMM 方法。《*IEEE* 信号处理汇刊》，64(14):3719–3734，2016。

Chen, T.、Jin, X.、Sun, Y. 与 Yin, W. Vafl：一种纵向异步联邦学习方法。*arXiv* 预印本 *arXiv:2007.06081*，2020。

Doan, T. T. 和 Olshevsky, A. 动态网络上的分布式资源分配：二次时间算法。《系统与控制快报》，99:57–63，2017年。

Du, K. 和 Meng, M. 带耦合不等式约束的分布式聚合优化的线性收敛。*arXiv* 预印本 *arXiv:2306.06700*，2023年。

Du, K. 和 Meng, M. 带仿射耦合约束的分布式聚合优化。《神经网络》，184：107085，2025年。

Dvinskikh, D. 和 Gasnikov, A. 用于随机凸规划问题的去中心化并行原始与对偶加速方法。《反问题与不适定问题杂志》，29(3):385–405，2021a。

Dvinskikh, D. 和 Gasnikov, A. 用于随机凸规划问题的去中心化并行原始与对偶加速方法。《反问题与不适定问题杂志》，29(3):385–405，2021b。

Falsone, A.、Notarnicola, I.、Notarstefano, G. 和 Prandini, M. 用于分布式约束耦合优化的跟踪交替方向乘子法。*Automatica*，117:108962，2020年。

Gong, K. 和 Zhang, L. 用于带耦合约束凸优化的去中心化近端乘子法。*arXiv* 预印本 *arXiv:2310.15596*，2023年。Gorbunov, E.、Dvinskikh, D. 和 Gasnikov, A. 用于随机凸优化的最优去中心化分布式算法。*arXiv* 预印本 *arXiv:1911.07363*，2019年。

Gutknecht, M. H. 和 R¨ollin, S. 重新审视切比雪夫迭代。《并行计算》，28(2):263–283，2002年。

Hanzely, F. 和 Richt´arik, P. 全局与局部模型混合的联邦学习。*arXiv* 预印本 *arXiv:2002.05516*，2020年。

Hanzely, F.、Hanzely, S.、Horv´ath, S. 和 Richt´arik, P. 个性化联邦学习的下界与最优算法。《神经信息处理系统进展》，33:2304–2315，2020年。

Kairouz, P.、McMahan, H. B.、Avent, B.、Bellet, A.、Bennis, M.、Bhagoji, A. N.、Bonawitz, K.、Charles, Z.、Cormode, G.、Cummings, R. 等。联邦学习的进展与未决问题。*arXiv* 预印本 *arXiv:1912.04977*，2019年。Kairouz, P.、McMahan, H. B.、Avent, B.、Bellet, A.、Bennis, M.、Bhagoji, A. N.、Bonawitz, K.、Charles, Z.、Cormode, G.、Cummings, R. 等。联邦学习的进展与未决问题

联邦学习。机器学习基础与趋势®，14(1–2):1–210，2021年。Koloskova, A., Loizou, N., Boreiri, S., Jaggi, M., and Stich, S. U. 具有变化拓扑和本地更新的去中心化 SGD 统一理论。*ICML 2020*，*arXiv* 预印本 *arXiv:2003.10422*，2020。Koloskova, A., Lin, T., and Stich, S. U. 去中心化机器学习中梯度跟踪的改进分析。神经信息处理系统进展，34，2021。Kovalev, D., Salim, A., and Richt´arik, P. 光滑强凸去中心化优化的最优实用算法。神经信息处理系统进展，33，2020。Kovalev, D., Gasanov, E., Gasnikov, A., and Richtarik, P. 时变网络上光滑强凸去中心化优化的下界与最优算法。神经信息处-理系统进展，34，2021a。Kovalev, D., Shulgin, E., Richt´arik, P., Rogozin, A. V., and Gasnikov, A. ADOM：面向时变网络的加速去中心化优化方法。载于国际机器学习会议，第 5784–5793 页。PMLR，2021b。Kovalev, D., Borodich, E., Gasnikov, A., and Feoktistov, D. 时变网络上非光滑凸去中心化优化的下界与最优算法。神经信息处理系统进展，37:96566–96606，2024年。Lan, G. 机器学习优化方法讲义。电子版预印本，2019。Lan, G. 机器学习的一阶与随机优化方法。Springer，2020。Lan, G., Lee, S., and Zhou, Y. 去中心化与随机优化的通信高效算法。数学规划，180(1):237–284，2020。Li, H. and Lin, Z. 时变图上的去中心化优化加速梯度跟踪。*arXiv* 预印本 *arXiv:2104.02596*，2021。Li, H., L¨u, Q., Liao, X., and Huang, T. 时变一般有向图下分布式约束优化的加速收敛算法。*IEEE* 系统、人类与控制论汇刊：系统，50 (7):2612–2622，2018。Lian, X., Zhang, C., Zhang, H., Hsieh, C.-J., Zhang, W., and Liu, J. 去中心化算法能否胜过集中式算法？去中心化并行随机梯度下降案例研究

随机梯度下降。载于神经信息处理系统进展，第 5330–5340 页，2017年。

Liang, S., Yin, G., 等。带耦合约束的分布式光滑凸优化。*IEEE* 自动控制汇刊，65(1):347–353，2019。

Liu, Y., Kang, Y., Zou, T., Pu, Y., He, Y., Ye, X., Ouyang, Y., Zhang, Y.-Q., 和 Yang, Q. 纵向联邦学习：概念、进展与挑战。*IEEE* 知识与数据工程汇刊，36(7):3615–3634，2024年。

Makhija, D., Ho, N., 和 Ghosh, J. 面向异构客户端的联邦自监督学习。*arXiv* 预印本 *arXiv:2205.12493*，2022。

McMahan, B.、Moore, E.、Ramage, D.、Hampson, S. 与 y Arcas, B. A. 基于去中心化数据的深度网络通信高效学习。载于人工智能与统计，第 1273–1282 页。PMLR，2017年。Necoara, I. 与 Nedelcu, V. 分布式对偶梯度方法与误差界条件。*arXiv* 预印本 *arXiv:1401.4398*，2014年。Necoara, I. 与 Nedelcu, V. 关于线性收敛：面向线性约束可分凸问题的分布式对偶梯度算法。*Automatica*，55:209–216，2015年。Necoara, I.、Nedelcu, V. 与 Dumitrache, I. 面向网络中估计与控制的并行与分布式优化方法。过程控制期刊，21(5):756–766，2011年。Nedi´c, A. 与 Ozdaglar, A. 面向多智能体优化的分布式次梯度方法。*IEEE* 自动控制汇刊，54(1):48–61，2009年。Nedic, A.、Ozdaglar, A. 与 Parrilo, P. A. 多智能体网络中的约束一致性与优化。*IEEE* 自动控制汇刊，55(4):922–938，2010年。Nedic, A.、Olshevsky, A. 与 Shi, W. 在时变图上实现分布式优化的几何收敛。*SIAM* 优化期刊，27(4): 2597–2633，2017年。Nedi´c, A.、Olshevsky, A. 与 Shi, W. 分布式资源分配的改进收敛速率。载于 *2018* 年 *IEEE* 决策与控制会议（*CDC*），第 172–177 页。IEEE，2018年。Rogozin, A.、Yarmoshik, D.、Kopylova, K. 与 Gasnikov, A. 带仿射约束的去中心化强凸优化：原始与对偶方法。载于 国际优化与应用会议，第 93–105 页。Springer，2022。

Salim, A., Condat, L., Kovalev, D., 和 Richtárik, P. 强凸最小化在仿射约束下的最优算法。在 *人工智能与统计会议论文集*，第 4482–4498 页。PMLR，2022。Scaman, K., Bach, F., Bubeck, S., Lee, Y. T. 和 Massouli´e, L. 网络中光滑强凸分布式优化的最优算法。载于第 *34* 届国际机器学习会议论文集，第 *70* 卷，第 3027–3036 页。JMLR. org，2017。Scaman, K., Bach, F., Bubeck, S., Massouli´e, L. 和 Lee, Y. T. 网络中非光滑分布式优化的最优算法。载于 神经信息处理系统进展，第 2740–2749 页，2018。Stanko, S., Karimullin, T., Beznosikov, A. 和 Gasnikov, A. 用于横向和纵向联邦学习的压缩加速方法：S. stanko 等。优化理论与应用杂志，208(2):68，2026。Uribe, C. A., Lee, S., Gasnikov, A. 和 Nedi´c, A. 网络分布式优化最优算法的对偶方法。优化方法与软件，第 1–40 页，2020。Vepakomma, P., Gupta, O., Swedish, T. 和 Raskar, R. 医疗健康领域的拆分学习：无需共享患者原始数据的分布式深度学习。*arXiv* 预印本 *arXiv:1812.00564*，2018。Wang, J. 和 Hu, G. 基于对偶近端梯度方法的多簇网络耦合约束分布式优化。*arXiv* 预印本 *arXiv:2203.00956*，2022。Wang, J., Kolar, M. 和 Srerbo, N. 分布式多任务学习。载于人工智能与统计，第 751–760 页。PMLR，2016。Wu, X., Wang, H. 和 Lu, J. 耦合约束下的分布式优化。*IEEE* 自动控制汇刊，68(3):1847–1854，2022。Xie, C., Chen, P.-Y., Li, Q., Nourian, A., Zhang, C. 和 Li, B. 通过高效通信和交替方向乘子法改进隐私保护纵向联邦学习。载于 *2024*年 *IEEE* 安全可信机器学习会议 *(SaTML)*，第 443–471 页。IEEE，2024。Yang, Q., Liu, Y., Chen, T. 和 Tong, Y. 联邦机器学习：概念与应用。*ACM* 智能系统与技术汇刊 *(TIST)*，10(2):1–19，2019。

Yarmoshik, D., Rogozin, A. 和 Gasnikov, A. 时变网络上的仿射约束分布式优化。计算管理科学，21(1):10，2024a。

Yarmoshik, D.、Rogozin, A.、Kiselev, N.、Dorin, D.、Gasnikov, A. 和 Kovalev, D. 去中心化的耦合约束优化。*arXiv* 预印本 *arXiv:2407.02020*，2024b。Zhang, B.、Gu, C. 和 Li, J. 时变有向图上带耦合约束的分布式凸优化。工业与管理优化期刊，17(4)：2119–2138，2021。Zhang, S.、Choromanska, A. E. 和 LeCun, Y. 采用弹性平均随机梯度下降的深度学习。神经信息处理系统进展，28，2015。Zhu, M. 和 Martinez, S. 不等式约束和等式约束下的分布式凸优化。*IEEE* 自动控制汇刊，57(1)：151–164，2011。

A. 补充符号说明

用 $\mathcal{L}_m$ 表示所谓的共识空间，其定义为 $\mathcal{L}_m = \{(y_1, \ldots, y_n) \in (\mathbb{R}^m)^n : y_1, \ldots, y_n \in \mathbb{R}^m$ 以及 $y_1 = \cdots = y_n\}$，而 $\mathcal{L}_m^{\perp}$ 表示 $\mathcal{L}_m$的正交补，其定义为

$$\mathcal{L}_m^{\perp} = \{(y_1, \ldots, y_n) \in (\mathbb{R}^m)^n : y_1, \ldots, y_n \in \mathbb{R}^m \text{ and } y_1 + \cdots + y_n = 0\}. \tag{15}$$

除非另有说明，对于任意矩阵 A，我们分别用 L_A 和 μ_A 表示其最大和最小正奇异值平方的某个上界和下界：

$$\lambda_{\max}(A^\top A) = \sigma_{\max}^2(A) \le L_A, \qquad \mu_A \le \sigma_{\min^+}^2(A) = \lambda_{\min^+}(A^\top A). \tag{16}$$

当提及一组矩阵 $A = (A_1 \ldots A_n)$ 时，我们将这些参数解释为针对块对角矩阵定义的参数，例如 $L_A = \max_i \sigma_{\max}^2(A_i)$、$\mu_A = \min_i \sigma^2{}_{\min^+}(A_i)$。

B. 混合约束的问题表述

局部模型与全局模型的混合。 与共识优化的区别在于，由局部持有的模型权重可能因节点而异。

$$\min_{z,x_1,\ldots,x_n} \sum_{i=1}^{n} f_i(x_i) + \frac{\lambda}{2}\|x_i - z\|^2$$
$$\text{s.t. } x_1 + \cdots + x_n = nz$$

变量 z 存储在第一个节点上，因此我们得到一个带有耦合约束的问题。

分布式多任务学习（多任务学习）。在（Wang等人，2016）中，作者提出了一种分布式多任务学习。每个节点都训练自己的模型，但以逐坐标的方式施加联合（不可分离）正则化项。

$$\min_{x_1,\ldots,x_n} \sum_{i=1}^{n} f_i(x_i) + \lambda \sum_{j=1}^{d} r(x^{(j)}),$$

其中， $x^{(j)}$ 是 $\mathbb{R}^n$ 中的一个向量，由 x_1，……， x_n的第 j个坐标组成。正则化项选为 $r(x) = \|x\|_2$ 或$r(x) = \|x\|_\infty$。该问题可以改写为耦合约束形式。引入大小为 $n \times d$ 的矩阵 Q_{ij}，其中除位置（i， j）上的一个元素为 1 外，其余元素均为零。乘法 $y = Q_{ij}x$ 返回向量 $y \in \mathbb{R}^n$，其第 i个元素等于 x的第 j个元素，其他分量均为零。

$$\min_{\substack{x_1,\ldots,x_n \\ y_1,\ldots y_d}} \sum_{i=1}^{n} f_i(x_i) + \lambda \sum_{j=1}^{d} r(y_j)$$
$$\text{s.t. } y_j = \sum_{i=1}^{n} Q_{ij}x_i.$$

这里 $y_j = x^{(j)}$ 是收集局部持有的 x_i 的第 j个分量的向量。

联邦自监督学习（自监督学习）。

在分布式自监督学习中，一组智能体致力于学习局部持有数据的表示；若相应数据具有共同特征或样本，则会与其他智能体同步。各方共享一个表示对齐数据集，以对齐其表示（Makhija等人，2022）。相应问题可表述为带共识约束的优化问题。设每个节点持有一个权重为 $x_i \in \mathbb{R}^{d_i}$ 的模型（局部模型可能具有不同的架构）。为简单起见，我们假设模型是线性的，且其损失函数写作 $\|C_i x_i - c_i\|^2$。共识约束施加于内部表示，我们假设其形式为 $A_i x_i$。联邦自监督学习问题可写作

$$\min_{\substack{x_1,\ldots,x_n \\ z_1,\ldots,z_n}} \sum_{i=1}^{n} \|C_i x_i - c_i\|^2 + \frac{\mu}{2}\|A_i x_i - z_i\|^2$$
$$\text{s.t. } z_1 = \ldots = z_n.$$

我们看到，这个问题归结为共识优化。

C. 切比雪夫迭代

算法3 切比雪夫(v, B, b) (Gutknecht & R¨ollin, 2002)

Input: v, B, b
$n \leftarrow \left\lceil \sqrt{\frac{\sigma^2_{\max}(B)}{\sigma^2_{\min^+}(B)}} \right\rceil$
$\rho \leftarrow \left(\sigma^2_{\max}(B) - \sigma^2_{\min^+}(B)\right)^2/16, \quad \nu \leftarrow \left(\sigma^2_{\max}(B) + \sigma^2_{\min^+}(B)\right)/2$
$\delta^0 \leftarrow -\nu/2$
$p^0 \leftarrow -B^\top(Bv - b)/\nu$
$v^1 \leftarrow v + p^0$
for $i = 1, \ldots, n-1$ **do**
  $\beta^{i-1} \leftarrow \rho/\delta^{i-1}$
  $\delta^i \leftarrow -(\nu + \beta^{i-1})$
  $p^i \leftarrow \left(B^\top(Bv^i - b) + \beta^{i-1}p^{i-1}\right)/\delta^i$
  $v^{i+1} \leftarrow v^i + p^i$
end for
Output: v^n

D. 定理的证明 2.4

D.1. 强凸情形（$\mu > 0$）

为证明这一情形，我们只需回顾一条关于 APAPC 收敛的定理，该定理出自（Salim 等人，2022 ）.

定理 D.1. （（*Salim 等人，2022*））。假设 *2.1* 和 *2.2* 成立。存在一组算法 *1*的参数，使得得到 u^N 满足 $\left\|u^N - u^*\right\|_2^2 \leq \varepsilon$ 需要 $N = O\left(\kappa_B \sqrt{L/\mu} \log(1/\varepsilon)\right)$ 次迭代。

D.2. 凸情形（$\mu = 0$）

正则化可用于将为强凸目标函数设计的方法适配到非强凸问题。我们将正则化函数定义如下：

$$G^\nu(u) = G(u) + \frac{\nu}{2}\|u^0 - u\|^2. \tag{17}$$

以下引理描述了求解正则化问题所需的精度。

引理 *D.2*（新增）。令 $G: \mathbb{R}^d \to \mathbb{R}$ 为凸且 L-光滑的函数，并假设存在解 $u^* \in \operatorname{Arg\,min} G(u)$ 和 $u_\nu^* = \operatorname{arg\,min} G^\nu(u)$。定义 $D = G(u^*) - \min_{u: Bu=b} G(u)$。假设 $\|u^0 - u^*\|^2 \leq R^2$。回顾定理 *2.4* 中的 u: $Bu=b$ $u: Bu=b$ $\nu = \varepsilon/R^2$，并令

$$\delta = \frac{\varepsilon^2}{32\left(D + \frac{\varepsilon}{2}\right)\left(L + \frac{\varepsilon}{R^2}\right)}. \tag{18}$$

若有 $\|u - u_\nu^*\|^2 \leq \delta$，则

$$G(u) - G(u^*) \leq \varepsilon,\ \|Bu - b\|^2 \leq \delta \sigma_{\max}^2(B). \tag{19}$$

换言之，只需以 $\delta = O(\varepsilon^2)$ 的精度求解正则化问题（以自变量的收敛性衡量），即可得到初始问题的 ε-解。

证明。我们有

$$\begin{aligned}
G(u) - G(u^*) &\overset{(a)}{\leq} G^\nu(u) - G^\nu(u_\nu^*) + \frac{\nu}{2}\|u^0 - u^*\|^2 \\
&\overset{(b)}{\leq} \langle \nabla G^\nu(u_\nu^*), u - u_\nu^* \rangle + \frac{L+\nu}{2}\|u - u_\nu^*\|^2 + \frac{\nu}{2}\|u^0 - u^*\|^2 \\
&\overset{(c)}{\leq} \|\nabla G^\nu(u_\nu^*)\| \cdot \|u - u_\nu^*\| + \frac{L+\nu}{2}\|u - u_\nu^*\|^2 + \frac{\nu}{2}\|u^0 - u^*\|^2 \\
&\overset{(d)}{\leq} \sqrt{2(L+\nu)\left(G^\nu(u_\nu^*) - \min_u G^\nu(u)\right)\|u - u_\nu^*\|^2} + \frac{L+\nu}{2}\|u - u_\nu^*\|^2 + \frac{\nu}{2}\|u^0 - u^*\|^2 \\
&\overset{(e)}{\leq} \sqrt{2(L+\nu)\left(G^\nu(u^*) - \min_u G(u)\right)\|u - u_\nu^*\|^2} + \frac{L+\nu}{2}\|u - u_\nu^*\|^2 + \frac{\nu}{2}\|u^0 - u^*\|^2 \\
&\overset{(f)}{=} \sqrt{2(L+\nu)\left(G(u^*) - \min_u G(u) + \frac{\nu}{2}\|u^0 - u^*\|^2\right)\|u - u_\nu^*\|^2} + \frac{L+\nu}{2}\|u - u_\nu^*\|^2 + \frac{\nu}{2}\|u^0 - u^*\|^2 \\
&\overset{(g)}{=} \sqrt{2(L+\nu)\left(D + \frac{\nu R^2}{2}\right)\delta} + \frac{(L+\nu)\delta}{2} + \frac{\nu R^2}{2} \\
&\overset{(h)}{\leq} \sqrt{2(L+\nu)\left(D + \frac{\varepsilon}{2}\right)\delta} + \frac{(L+\nu)\delta}{2} + \frac{\nu R^2}{2} \\
&\overset{(i)}{\leq} \frac{\varepsilon}{64} + \frac{\varepsilon}{4} + \frac{\varepsilon}{2} \\
&< \varepsilon.
\end{aligned}$$

其中，(a) 根据正则化函数 G^ν 的定义；(b) 利用了 G^ν 的 $(L+\nu)$-光滑性；(c) 利用了柯西–施瓦茨不等式；(d) 利用了 G^ν 的凸性和光滑性；(e) 利用了以下事实：对于所有

u；(f) 利用了 G^ν的定义；(g) 利用了 R、 D 的定义以及 $\|u-u_\nu^*\| \le \delta$这一假设；(h) 根据 ν的定义；(i) 利用了 δ 在（18）中的定义。

此外，

$$\|Bu-b\|^2 = \|Bu-Bu_\nu^*\|^2 \le \sigma_{\max}^2(B)\cdot\|u-u_\nu^*\|^2 \le \sigma_{\max}^2(B)\cdot\delta = O(\sigma_{\max}^2(B)\varepsilon^2).$$

□

根据命题 1（Salim 等人，2022）以及引理 D.2，为达到 $G(u^k)-G(u^*) \le \varepsilon$，APAPC（算法 1）需要 $\mathcal{O}\left(\kappa_B\sqrt{\frac{L+\nu}{\nu}}\log\left(\frac{1}{\delta}\right)\right)$ 次迭代。

由 δ 在（18）中的定义以及 ν 在引理D.2中的定义，我们有：

$$\frac{1}{\delta} = \mathcal{O}\left(\frac{\left(M+\frac{\varepsilon}{2}\right)\left(L+\frac{\varepsilon}{R^2}\right)}{\varepsilon^2}\right),\quad \frac{L+\nu}{\nu} = 1+\frac{L}{\nu} = 1+\frac{LR^2}{\varepsilon} = \mathcal{O}\left(\frac{LR^2}{\varepsilon}\right)$$

因此，得到的复杂度为 $\mathcal{O}\left(\kappa_B\sqrt{\frac{LR^2}{\varepsilon}}\log\left(\frac{\left(M+\frac{\varepsilon}{2}\right)\left(L+\frac{\varepsilon}{R^2}\right)}{\varepsilon^2}\right)\right)$ 次迭代。

E. 关于梯度滑动法的讨论

E.1. 预备知识：梯度滑动法

下面讨论用于分解由两项组成的优化问题复杂度的确定性梯度滑动（GS）方法（Lan，2020，算法 8.1）；（Lan 等人，2020），考虑问题

$$\min_{u\in U} K(u) := K_s(u) + K_n(u). \tag{20}$$

定理 *E.1*（*Lan，2020*）。设 K_s 满足假设 *2.1*，强凸性参数为 $\mu = 0$，且满足假设 *2.2*；$K_n(u)$ 满足假设 *2.1*（$\mu = 0$）和假设 *2.3*，且 $\left\|u^0 - u^*\right\|^2 \le R^2$。将梯度滑移算法应用于问题 *(20)*，需要对 K_s 的梯度进行 $N_s = O\left(\sqrt{\frac{LR^2}{\varepsilon}}\right)$ 次调用，并对K_n 的次梯度进行 $N_n = O\left(\frac{M^2R^2}{\varepsilon^2} + N_s\right)$ 次调用，从而得到满足 $K(\widehat{u}) - K^* \le \varepsilon$ 的 $\widehat{u}$。

通过拉格朗日 乘子法，问题 (5) 可等价地写成以下鞍点问题 ：

$$\min_{u\in U} \max_{v\in\mathbb{R}^p} [G(u) + \langle v, Bu - b\rangle] \tag{21}$$

引理 E.2 (Lan 等人 (2020))**.** 设 u^* 是 (5)的一个最优解。那么存在一个最优对偶乘子 v^*，使得对于(21)有

$$\|v^*\| \le R_{\text{dual}} := \frac{M}{\sigma_{\min^+}(B)}. \tag{22}$$

L引理 E.3 (Gorbunov 等人 (2019))**.** 设 $\varepsilon > 0$ 和 $r = \frac{2R^2}{\text{对偶}\varepsilon}$。若 $\widehat{u}$是 ε-解，属于问题 (6)，即 .

$$H_r(\widehat{u}) - \min_{u\in U} H_r(u) \le \varepsilon,$$

then

$$G(\widehat{u}) - \min_{Bu-b=0} G(u) \le \varepsilon, \quad \|B\widehat{u} - b\| \le \frac{2\varepsilon}{R_{\text{dual}}}.$$

尽管 (Lan 等人，2020) 和 (Gorbunov 等人，2019) 仅研究了共识优化问题 ($B = \mathbf{W}$，$b = 0$)，参考文献中引理 E.2 和 E.3 的证明无需修改即可适用于任意 B 和 $b \in \operatorname{Im} B$。

E.2. 定理2.5 的证明

证明。记 $K_n(u) = G(u)$ 与 $K_s(u) = \frac{R^2}{\text{对偶}\varepsilon}\|Bu\|_2^2$，从而 $L = \frac{2R^2}{\text{对偶}\varepsilon}\frac{\sigma^2 \quad B}{} \le \frac{2M^2}{\varepsilon\sigma^2}\frac{\sigma^2}{\max()\ \max(B)\ \min^+(B)}$ $= \frac{2M^2}{\varepsilon}\kappa_B$，其中该不等式由引理E.2得到。应用定理E.1 即得所需结果。

当 $\mu > 0$ 时，我们只需将 $L = \frac{2M^2}{\varepsilon}\kappa_B$ 代入 R-Sliding 的复杂度（Lan，2020，定理 8.3） $N_s = O\left(\sqrt{\frac{L}{\mu}}\log\frac{1}{\varepsilon}\right)$，$N_n = \overline{O}(\frac{M^2}{\mu\varepsilon} + N_s)$。需要注意的是，出于某种原因，（Lan，2020，第 8.1.3.1 节）仅考虑光滑分量为强凸的情形。但从（Lan，2020，定理 8.3）的证明可以看出，其中只用到了光滑项与非光滑项之和 $K(u)$ 的强凸性，因此我们可以将其应用于 $K_n(u)$ 为强凸的情形，正如（Dvinskikh 和 Gasnikov，2021a；Uribe 等，2020）所做的那样。 □

第3节的辅助定理与引理

F.1. 相同的局部约束

设约束矩阵 C_i 相等。则仿射约束的块对角矩阵具有如下形式：$\mathbf{C} = I_n \otimes C$。关于 $\mathbf{B}^\top = [\mathbf{C}^\top \ \gamma\mathbf{W}]$的谱性质，可以参照引理 F.1重新考察。

引理 F.1（（Rogozin 等人，2022，引理 1））。

$$\sigma^2_{\max}(\mathbf{B}) = \sigma^2_{\max}(C) + \gamma^2\sigma^2_{\max}(W),$$
$$\sigma^2_{\min^+}(\mathbf{B}) = \min\left\{\sigma^2_{\min^+}(C), \gamma^2\sigma^2_{\min^+}(W)\right\}.$$

A因此，我们得到了关于定理 4.1 的通信次数的增强界 .

定理 F.2（（Rogozin 等人，2022））。将算法 *1* 应用于问题 (10) 并结合 $C_1 = \ldots = C_n = C$，应用$\mathbf{W} \to P_W$（$\mathbf{W}$）和 $\mathbf{C} \to P_C$（$\mathbf{C}^\top\mathbf{C}$）后，我们得到了一种所需通信次数为

$$N_{\nabla f} = O\left(\sqrt{\kappa_f}\log\left(\frac{1}{\varepsilon}\right)\right) \textit{ gradient calls},$$
$$N_C = O\left(\sqrt{\kappa_f}\sqrt{\kappa_C}\log\left(\frac{1}{\varepsilon}\right)\right) \textit{ mul. by } \mathbf{C} \textit{ and } \mathbf{C}^\top,$$
$$N_W = O\left(\sqrt{\kappa_f}\sqrt{\kappa_W}\log\left(\frac{1}{\varepsilon}\right)\right) \textit{ communications}.$$

这些上界由于相应的下界而是最优的。

F.2. 耦合约束

令 $A_i \in \mathbb{R}^{m\times d_i}$。耦合约束的适用于去中心化的改写为 $\mathbf{Ax} + \mathbf{Wy} = \mathbf{b}$，其中 $\mathbf{y} \in \mathbb{R}^{mn}$。于是问题可写为

$$\min_{\mathbf{x}\in\mathbb{R}^d} F(\mathbf{x}) := \sum_{i=1}^n f_i(x_i) \quad \text{s.t.} \quad \mathbf{Ax} + \mathbf{Wy} = \mathbf{b}. \tag{C}$$

事实上，由于 $\mathbf{W}$ 的值域为 $\{\mathbf{x}\colon x_1 + x_2 + \ldots x_n = 0\}$，我们有 $\mathbf{Ax} - \mathbf{b} \in$ 值域 $\mathbf{W} \Leftrightarrow \sum_{i=1}^n (A_i x_i - b_i) = 0$。引入变量 $\mathbf{y}$ 是为了对值域 $\mathbf{W}$进行参数化。然而，这样做的代价是，加入 $\mathbf{y}$后，目标函数 $F(\mathbf{x}, \mathbf{y})$不再是强凸的。因此，利用强凸性的关键技术是在目标函数中加入适当缩放的增广拉格朗日型惩罚项（例如，参见引理 I.1）。

以下定义用于描述具有耦合约束的问题的去中心化算法的收敛率。

引理 F.3 ((Yarmoshik 等人，2024b，引理 2))。记$\mathbf{B} = \begin{pmatrix}\mathbf{A} & \beta\mathbf{W}\end{pmatrix}$。设定 $\beta^2 = \frac{\lambda_{\min^+}(S_A)+\sigma^2_{\max}(\mathbf{A})}{\sigma^2_{\min^+}(\mathbf{W})}$，我们得到

$$\sigma^2_{\max}(\mathbf{B}) \le \sigma^2_{\max}(\mathbf{A}) + (\sigma^2_{\max}(\mathbf{A}) + \lambda_{\min^+}(S_A))\kappa^2_W,$$
$$\sigma^2_{\min^+}(\mathbf{B}) \ge \frac{\lambda_{\min^+}(S_A)}{2}.$$

定理 F.4 ((Yarmoshik 等人，2024b)，定理 1 和 2)。对 (C) 施加惩罚并应用切比雪夫加速$\mathbf{W} \to P_W(\mathbf{W})$，然后 $\mathbf{B} \to P_B(\mathbf{B}^\top\mathbf{B})$ 并应用算法 *1*，我们得到一种所需的计算方法

$$N_{\nabla f} = O\left(\sqrt{\kappa_f}\log\left(\frac{1}{\varepsilon}\right)\right) \textit{ gradient calls},$$
$$N_A = O\left(\sqrt{\kappa_f}\sqrt{\widehat{\kappa}_A}\log\left(\frac{1}{\varepsilon}\right)\right) \textit{ mul. by } \mathbf{A} \textit{ and } \mathbf{A}^\top,$$
$$N_W = O\left(\sqrt{\kappa_f}\sqrt{\widehat{\kappa}_A}\sqrt{\kappa_W}\log\left(\frac{1}{\varepsilon}\right)\right) \textit{ communications}.$$

对于具有耦合约束的问题，这一界在一个自然定义的去中心化一阶算法类别中是最优的。

G. 证明 共享变量约束的复杂度界（定理 4.1）

G.1. 上界

如定理陈述中所述，我们首先对矩阵 2.4（第 $\mathbf{W}$ 节）应用切比雪夫预条件，得到矩阵 $\mathbf{W}' = P_W(\mathbf{W})$，满足 $\kappa_{\mathbf{W}'} = O(1)$。然后，根据引理 F.3，按照其陈述选择 γ，对于$\widetilde{\mathbf{B}}^\top = \begin{pmatrix} \widetilde{\mathbf{C}} & \gamma \mathbf{W}' \end{pmatrix}$

$$\kappa_{\widetilde{\mathbf{B}}} = \frac{\sigma_{\max}^2(\widetilde{\mathbf{B}})}{\sigma_{\min^+}^2(\widetilde{\mathbf{B}})} = \frac{\sigma_{\max}^2(\widetilde{\mathbf{B}}^\top)}{\sigma_{\min^+}^2(\widetilde{\mathbf{B}}^\top)} \leq \frac{\sigma_{\max}^2(\widetilde{\mathbf{C}}^\top) + \left[\sigma_{\max}^2(\widetilde{\mathbf{C}}^\top) + \lambda_{\min^+}(S_{C^\top})\right] \kappa_{\mathbf{W}'}^2}{\frac{1}{2}\lambda_{\min^+}(S_{C^\top})} = O(\kappa_{\mathbf{W}'}^2 \widehat{\kappa}_{\widetilde{\mathbf{C}}}) = O(\widehat{\kappa}_{\widetilde{\mathbf{C}}}). \quad (23)$$

最后，根据定理 D.1，算法1 应用于问题时的迭代复杂度为

$$\min_{\mathbf{x} \in (\mathbb{R}^d)^n} F(\mathbf{x}) := \sum_{i=1}^n f_i(x_i) \text{ s.t. } \widetilde{\mathbf{B}}' x = \tilde{\mathbf{b}}', \quad (24)$$

is $N = O(\sqrt{\kappa_f} \log \frac{1}{\varepsilon})$，其中 $\widetilde{\mathbf{B}}' = P_{\widetilde{B}}(\widetilde{\mathbf{B}}^\top \widetilde{\mathbf{B}})$，$\kappa_{\widetilde{\mathbf{B}}'} = O(1)$ 和 $\tilde{\mathbf{b}}' = P_{\widetilde{B}}(\widetilde{\mathbf{B}}^\top \widetilde{\mathbf{B}}) \widetilde{\mathbf{B}}^\top \widetilde{\mathbf{B}} \widetilde{\mathbf{B}}^\top \begin{pmatrix} \tilde{\mathbf{c}} \\ 0 \end{pmatrix}$。

在这种情况下，算法 1 的每次迭代需要进行 1 次计算、∇f、$O(\deg P_{\widetilde{\mathbf{B}}})$ 次 $= O(\sqrt{\kappa_{\widetilde{\mathbf{B}}}}) = O(\sqrt{\widehat{\kappa}_{\widetilde{\mathbf{C}}}})$ 乘法运算（乘以 $\widetilde{\mathbf{B}}\ \widetilde{\mathbf{B}}^\top$），每次都需要 $O(1)$ 次乘以 $\widetilde{\mathbf{C}}\ \widetilde{\mathbf{C}}^\top$ 的乘法运算，以及 $O(\deg P_W) = O(\sqrt{\kappa_W})$ 次乘以 $\mathbf{W}$ 的乘法运算，即通信轮次，由此得到定理的第一部分。

G.2. 下界

这个证明是对（Yarmoshik 等人，2024b，定理 2）证明的修改。主要区别在于，我们如何拆分约束矩阵，以便在 G.2.4 节中将 Nesterov 的坏函数 $h(z) = \frac{1}{2} z^\top \mathbf{M} z + \alpha \|z\|_2^2 - z_1$ 分配到各个节点上。对于原始的耦合约束情形，矩阵 $\mathbf{M}$ 被拆分为两个加项 $\mathbf{M} = \mathbf{E}_1^\top \mathbf{E}_2 + \mathbf{E}_2^\top \mathbf{E}_2$。矩阵 $\mathbf{E}_1, \mathbf{E}_2$ 用于构造耦合约束的矩阵。这里我们取一个非对称的根矩阵 $\mathbf{E}$，使得 $\mathbf{M} = \mathbf{E}^\top \mathbf{E}$，并按行将其拆分为矩阵 $\mathbf{E}_1, \mathbf{E}_2$，以构造局部约束的矩阵。为完整起见，我们在此描述下界的整个构造过程，因为其其他部分（例如 G.2.3 节中定义的目标函数）也需要经过修改（尽管这些修改更为细微且技术性更强），才能应用于此设置。

G.2.1. 对偶问题

证明依赖于将内斯捷罗夫函数作为对偶问题的目标函数。我们构造中的原变量和对偶变量具有相似的分量结构，因此可以利用对偶问题精确解的显式表达式，以及近似解中非零分量数量的上界，推导出原问题近似解精度的上界。

让我们推导对偶问题。考虑约束右端项为零的原问题 ts

$$\begin{aligned} &\min_{x_1, \ldots, x_n \in \ell_2} \sum_{i=1}^n f_i(x_i) \\ &\text{s.t. } \mathbf{C}_i x_i = 0 \quad \forall i = 1, \ldots, n. \\ &x_1 = \ldots = x_n. \end{aligned} \quad (25)$$

简化共识约束，并将仿射约束合并为单个约束 $\mathbf{C}x = 0$（例如，将矩阵 $\mathbf{C}_i$按列堆叠），我们将问题改写为

$$\begin{aligned} &\min_{x \in \ell_2} \sum_{i=1}^n f_i(x) \\ &\text{s.t. } \tilde{\mathbf{C}} x = 0. \end{aligned} \quad (26)$$

对偶问题的形式为

$$\max_z \min_{x\in\ell_2}\left[\sum_{i=1}^n f_i(x) - \left\langle z, \tilde{\mathbf{C}}x\right\rangle\right] = -\min_z F^*(\tilde{\mathbf{C}}^\top z), \tag{27}$$

其中 $F(x) = \sum_{i=1}^n f_i(x)$。

G.2.2. 示例图

图构造与耦合约束下界中的相同。

我们遵循（Kovalev 等人，2021a）中介绍的下界构造原则，并采用（Scaman 等人，2017年）中的示例图。将节点持有的函数组织成一个包含 n 个顶点的路径图，其中 n 可被3整除。图 $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ 的节点分为三组 $\mathcal{V}_1 = \{1, \ldots, n/3\}$、$\mathcal{V}_2 = \{n/3+1, \ldots, 2n/3\}$、$\mathcal{V}_3 = \{2n/3+1, \ldots, n\}$，每组包含 $n/3$ 个顶点。

现在我们回顾（Scaman 等，2017 年）中的构造。路径图的最大和最小特征值形式为 $\lambda_{\max}(W) = 2\left(1+\cos\frac{\pi}{n}\right)$，$\lambda_{\min^+}(W) = 2\left(1-\cos\frac{\pi}{n}\right)$。令 $\beta_n = \frac{1+\cos(\frac{\pi}{n})}{1-\cos(\frac{\pi}{n})}$。由于 $\beta_n \overset{n\to\infty}{\to} +\infty$，存在$n = 3m \geq 3$ 使得 $\beta_n \leq \kappa_W < \beta_{n+3}$。对于这个 n，引入边权重 $w_{i,i+1} = 1 - a\mathbb{I}\{i=1\}$，取相应的加权拉普拉斯矩阵 W_a，并记其条件数为 $\kappa(W_a)$。如果 $a=1$，网络不连通，因此 $\kappa(W_a) = \infty$。如果 $a=0$，则有 $\kappa(W_a) = \beta_n$。根据拉普拉斯谱的连续性可知，对于某个$a \in [0,1)$，有 $\kappa(W_a) = \kappa_W$。注意，$\pi/(n+3) \in [0, \pi/3]$，并且当 $x \in [0, \pi/3]$ 时有 $1-\cos x \geq x^2/4$。我们有

$$\kappa_W \leq \beta_{n+3} = \frac{1+\cos\frac{\pi}{n+3}}{1-\cos\frac{\pi}{n+3}} \leq \frac{72(n+3)^2}{\pi^2} \leq \frac{288n^2}{\pi^2} \leq 32n^2 \quad \Rightarrow \quad \sqrt{\kappa_W} \leq 4\sqrt{2}n = O(n). \tag{28}$$

G.2.3. 示例函数

令 $e_1 = (1\ 0\ \ldots 0)^\top$ 表示第一坐标向量，并令 x 由两个变量块 $x = \begin{pmatrix} p \\ t \end{pmatrix}$ 组成。我们令所有节点的函数 f_i 均相同，并将其定义为

$$f_i(p,t) = \frac{\mu_f}{2}\|p\|^2 + L_f\left\|t + \frac{L'_{\mathbf{C}}}{\mu_f}e_1\right\|_2^2.$$

相应地,

$$f_i^*(u,v) = \frac{1}{2\mu_f}\|u\|_2^2 + \frac{1}{L_f}\|v\|_2^2 - \frac{L'_{\mathbf{C}}}{\mu_f}v_1. \tag{29}$$

Let

$$\mathbf{E} = \begin{pmatrix} 1 & 0 & 0 & 0 & \cdots \\ -1 & 1 & 0 & 0 & \cdots \\ 0 & -1 & 1 & 0 & \cdots \\ 0 & 0 & -1 & 1 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}, \quad \tilde{\mathbf{C}} = \begin{pmatrix} -\sqrt{L'_{\mathbf{C}}}\mathbf{E}^\top & \sqrt{\mu'_{\mathbf{C}}}\mathbf{I} \end{pmatrix} \tag{30}$$

由（27）和（29），对偶问题为

$$\begin{aligned} &\min_z \frac{L'_{\mathbf{C}}}{2\mu_f}\|\mathbf{E}z\|_2^2 + \frac{\mu'_{\mathbf{C}}}{L_f}\|z\|_2^2 - \frac{L'_{\mathbf{C}}}{\mu_f}z_1 \\ =&\min_z \frac{L'_{\mathbf{C}}}{2\mu_f}\langle z, \mathbf{M}z\rangle + \frac{\mu'_{\mathbf{C}}}{L_f}\|z\|_2^2 - \frac{L'_{\mathbf{C}}}{\mu_f}z_1 \\ =&\min_z \frac{L'_{\mathbf{C}}}{\mu_f}\left(\frac{1}{2}\langle z, \mathbf{M}z\rangle + \frac{\mu'_{\mathbf{C}}\mu_f}{L'_{\mathbf{C}}L_f}\|z\|_2^2 - z_1\right), \end{aligned} \tag{31}$$

其中

$$\mathbf{M} = \mathbf{E}^\top \mathbf{E} = \begin{pmatrix} 2 & -1 & 0 & 0 & 0 & \dots \\ -1 & 2 & -1 & 0 & 0 & \dots \\ 0 & -1 & 2 & -1 & 0 & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}.$$

问题（31）恰好是　采用一阶方法进行光滑强凸最小化时，内斯捷罗夫提出的最坏情形问题　s.

引理 G.1（（Yarmoshik 等人，2024b，引理 6））。问题（*31*）的解为 $z^* = \{\rho^k\}_{k=1}^{\infty}$，其中

$$\rho = \frac{\sqrt{\frac{2}{3}\frac{L_{\mathbf{C}} L_f}{\mu_{\mathbf{C}} \mu_f} + 1} - 1}{\sqrt{\frac{2}{3}\frac{L_{\mathbf{C}} L_f}{\mu_{\mathbf{C}} \mu_f} + 1} + 1}.$$

G.2.4. 示例矩阵

~　　~我们将矩阵 $\mathbf{C}$ 拆分为两个矩阵，方法如下：将 $\mathbf{C}$ 的偶数行放入第一个矩阵，奇数行填零；第二个矩阵则以相同方式由 ~$\mathbf{C}$的奇数行构造。形式上，令 $L'_{\mathbf{C}} = \frac{1}{2}L_{\mathbf{C}} - \frac{3}{2}\mu_{\mathbf{C}}$，$\mu'_{\mathbf{C}} = 3\mu_{\mathbf{C}}$，其中$L_{\mathbf{C}} > 0$ 和 $\mu_{\mathbf{C}} > 0$ 是满足 $\widehat{\kappa}_{C^\top} = \frac{L_{\mathbf{C}}}{\mu_{\mathbf{C}}}$的任意参数，并引入

$$\mathbf{C}_i = \begin{cases} \begin{pmatrix} \sqrt{L'_{\mathbf{C}}}\mathbf{E}_1^\top & \sqrt{\mu'_{\mathbf{C}}}\mathbf{I}_1 \end{pmatrix}, & i \in \mathcal{V}_1 \\ \begin{pmatrix} \mathbf{0} & \mathbf{0} \end{pmatrix}, & i \in \mathcal{V}_2 \\ \begin{pmatrix} \sqrt{L'_{\mathbf{C}}}\mathbf{E}_2^\top & \sqrt{\mu'_{\mathbf{C}}}\mathbf{I}_2 \end{pmatrix}, & i \in \mathcal{V}_3 \end{cases},$$

$$\mathbf{E}_1^\top = \begin{pmatrix} 1 & -1 & 0 & 0 & 0 & \cdots \\ 0 & 0 & 0 & 0 & 0 & \cdots \\ 0 & 0 & 1 & -1 & 0 & \cdots \\ 0 & 0 & 0 & 0 & 0 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}, \quad \mathbf{I}_1 = \begin{pmatrix} 1 & 0 & 0 & 0 & \cdots \\ 0 & 0 & 0 & 0 & \cdots \\ 0 & 0 & 1 & 0 & \cdots \\ 0 & 0 & 0 & 0 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}, \tag{32}$$

$$\mathbf{E}_2^\top = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & \cdots \\ 0 & 1 & -1 & 0 & 0 & \cdots \\ 0 & 0 & 0 & 0 & 0 & \cdots \\ 0 & 0 & 0 & 1 & -1 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}, \quad \mathbf{I}_2 = \begin{pmatrix} 0 & 0 & 0 & 0 & \cdots \\ 0 & 1 & 0 & 0 & \cdots \\ 0 & 0 & 0 & 0 & \cdots \\ 0 & 0 & 0 & 1 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}. \tag{33}$$

~显然，如此定义的局部约束等价于$\mathbf{C}x = 0$。

让我们确认，如此选择的$\mathbf{C}_i$ 确实能保证 $\max_{i=1,\dots,n} \frac{\lambda_{\max}(C_i^\top C_i)}{\lambda_{\min^+}(S_{C^\top})} \leq \widehat{\kappa}_{C^\top}$ 满足定理的陈述以及定义 3.4的要求。

对于分子，我们有

$$\max_i \lambda_{\max}(\mathbf{C}_i^\top \mathbf{C}_i) = \max_i \lambda_{\max}(\mathbf{C}_i \mathbf{C}_i^\top) = \lambda_{\max}\left(L'_{\mathbf{C}} \mathbf{E}_1^\top \mathbf{E}_1 + \mu'_{\mathbf{C}} \mathbf{I}\right) = 2L'_{\mathbf{C}} + \mu'_{\mathbf{C}} = L_{\mathbf{C}}. \tag{34}$$

对于分母，直接计算可得

$$\mathbf{C}_1^\top \mathbf{C}_1 = \begin{pmatrix} 1 & -1 & 0 & 0 & \cdots & 1 & 0 & 0 & 0 & \cdots \\ -1 & 1 & 0 & 0 & \cdots & -1 & 0 & 0 & 0 & \cdots \\ 0 & 0 & 1 & -1 & \cdots & 0 & 0 & 1 & 0 & \cdots \\ 0 & 0 & -1 & 1 & \cdots & 0 & 0 & -1 & 0 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \cdots & \vdots & \vdots & \vdots & \vdots & \cdots \\ 1 & -1 & 0 & 0 & \cdots & 1 & 0 & 0 & 0 & \cdots \\ 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 & \cdots \\ 0 & 0 & 1 & -1 & \cdots & 0 & 0 & 1 & 0 & \cdots \\ 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \cdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}, \quad \mathbf{C}_2^\top \mathbf{C}_2 = \begin{pmatrix} 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 & \cdots \\ 0 & 1 & -1 & 0 & \cdots & 0 & 1 & 0 & 0 & \cdots \\ 0 & -1 & 1 & 0 & \cdots & 0 & -1 & 0 & 0 & \cdots \\ 0 & 0 & 0 & 1 & \cdots & 0 & 0 & 0 & 1 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \cdots & \vdots & \vdots & \vdots & \vdots & \cdots \\ 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 & \cdots \\ 0 & 1 & -1 & 0 & \cdots & 0 & 1 & 0 & 0 & \cdots \\ 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 & \cdots \\ 0 & 0 & 0 & 1 & \cdots & 0 & 0 & 0 & 1 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \cdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}. \tag{35}$$

由此可得

$$\begin{aligned} \lambda_{\min^+}\left(\frac{1}{n}\sum_{i=1}^n \mathbf{C}_i^\top \mathbf{C}_i\right) &= \frac{1}{3}\lambda_{\min^+}\begin{pmatrix} L'_{\mathbf{C}}(\mathbf{E}_1\mathbf{E}_1^\top + \mathbf{E}_2\mathbf{E}_2^\top) & \sqrt{L'_{\mathbf{C}}\mu'_{\mathbf{C}}}(\mathbf{E}_1\mathbf{I}_1 + \mathbf{E}_2\mathbf{I}_2) \\ \sqrt{L'_{\mathbf{C}}\mu'_{\mathbf{C}}}(\mathbf{I}_1^\top\mathbf{E}_1 + \mathbf{I}_2^\top\mathbf{E}_2) & \mu'_{\mathbf{C}}(\mathbf{I}_1^\top\mathbf{I}_1 + \mathbf{I}_2^\top\mathbf{I}_2) \end{pmatrix} \\ &= \frac{1}{3}\lambda_{\min^+}\begin{pmatrix} L'_{\mathbf{C}}\mathbf{L}_{\text{path}} & \sqrt{L'_{\mathbf{C}}\mu'_{\mathbf{C}}}\mathbf{E} \\ \sqrt{L'_{\mathbf{C}}\mu'_{\mathbf{C}}}\mathbf{E}^\top & \mu'_{\mathbf{C}}\mathbf{I} \end{pmatrix} \\ &= \frac{1}{3}\lambda_{\min^+}\left(\begin{pmatrix} \sqrt{L'_{\mathbf{C}}}\mathbf{E} \\ \sqrt{\mu'_{\mathbf{C}}}\mathbf{I} \end{pmatrix}\begin{pmatrix} \sqrt{L'_{\mathbf{C}}}\mathbf{E}^\top & \sqrt{\mu'_{\mathbf{C}}}\mathbf{I} \end{pmatrix}\right) \\ &= \frac{1}{3}\lambda_{\min^+}\left(\begin{pmatrix} \sqrt{L'_{\mathbf{C}}}\mathbf{E}^\top & \sqrt{\mu'_{\mathbf{C}}}\mathbf{I} \end{pmatrix}\begin{pmatrix} \sqrt{L'_{\mathbf{C}}}\mathbf{E} \\ \sqrt{\mu'_{\mathbf{C}}}\mathbf{I} \end{pmatrix}\right) \\ &= \frac{1}{3}\lambda_{\min^+}\left(L'_{\mathbf{C}}\mathbf{M} + \mu'_{\mathbf{C}}\mathbf{I}\right) = \frac{\mu'_{\mathbf{C}}}{3} = \mu_{\mathbf{C}}, \end{aligned}$$

其中 $\mathbf{L}_{\text{path}} = \mathbf{E}\mathbf{E}^\top$ 是无限（单向）路径图的拉普拉斯矩阵，与$\mathbf{M}$ 的区别仅在于第一个对角分量$\mathbf{L}_{\text{path}}[1,1] = 1$。

G.2.5. 精度界定

我们考虑如下定义的、用于具有局部仿射约束问题的一阶去中心化算法类**定义 G.2** ((Yarmoshik 等人, 2024a, 定义 1))。 记 $\mathcal{M}_i(k)$，其中 $\mathcal{M}_i(0) = \{x_i^0\}$表示第 i个节点在步骤 k的局部记忆。一阶去中心化算法在步骤 k 允许采取的行动仅限于以下三种

1. 局部计算： $\mathcal{M}_i(k) = \text{span}\left(\{x, \nabla f_i(x)\ \nabla f_i^*(x) : x \in \mathcal{M}_i(k)\}\right)$;
2. 与直接邻居进行去中心化通信： $\mathcal{M}_i(k) = \text{span}\left(\{\mathcal{M}_j(k) : \text{边 } (i,\ j) \in E\}\right)$。
3. 矩阵乘法： $\mathcal{M}_i(k) = \text{span}\left(\{b_i, \mathbf{C}_i^\top \mathbf{C}_i x : x \in \mathcal{M}_i(k)\}\right)$。

每一步之后 k，算法 必须给出当前的近似解 $x_i^k \in \mathcal{M}_i(k)$ 和集合 $\mathcal{M}_i(k+1) = \mathcal{M}_i(k)$。

不失一般性，我们可以假设 $x_i^0 = 0$。回想一下，我们将 x 分成两个变量块 p 和 t。从矩阵 $\mathbf{C}_1^\top \mathbf{C}_1$ 的结构可知，在第 k 步，节点 $i \in \mathcal{V}_1$ 只能在 p_i^k 或 t_i^k 中最后一个非零分量的索引为奇数时，使 p_i^k 或 t_i^k 中的非零分量数量增加一个。类似地，根据 $\mathbf{C}_2^\top \mathbf{C}_2$, 的结构，来自 $\mathcal{V}_2$ 的节点只有在下一个零分量为奇数时才能将其“解锁”。因此，要使 p_i^k 或 t_i^k 中的非零分量数量增加两个，信息必须从 $\mathcal{V}_1$ 传输到 $\mathcal{V}_2$ 再传回来（或反向传输），这需要进行 $2n/3 = \Omega(\sqrt{\kappa_W})$ 个去中心化通信轮次和两次矩阵乘法。

由于 强对偶，问题（25）的解可由其对偶（27）的解得到a s

$$x^*(z^*) = \begin{pmatrix} p^* \\ t^* \end{pmatrix}(\mathbf{C}^\top z^*) = \begin{pmatrix} \frac{\sqrt{L'_{\mathbf{C}}}}{\mu_f}\mathbf{E}z^* \\ \frac{\sqrt{\widehat{\mu}_{\mathbf{A}}}}{2L_f}(z^* - \frac{L'_{\mathbf{C}}}{\mu_f}e_1) \end{pmatrix}.$$

因此 t^* 只是 z^* 除第一个分量外的缩放版本。引理G.1 以及标准计算（见（Yarmoshik 等人，2024b，附录 C.4））随后给出了个数 q 的如下界，即 t_i^k 中达到精度 $\left\|x_i^k - x_i^*\right\|_2^2 \leq \varepsilon$所需的非零分量：

$$q \geq \Omega\left(\sqrt{\frac{L_{\mathbf{C}} L_f}{\mu_{\mathbf{C}} \mu_f}} \log\left(\frac{1}{\varepsilon}\right)\right), \tag{36}$$

这意味着所需的矩阵乘法次数

$$N_{\mathbf{C}} \geq \Omega\left(\sqrt{\frac{L_{\mathbf{C}} L_f}{\mu_{\mathbf{C}} \mu_f}} \log\left(\frac{1}{\varepsilon}\right)\right), \tag{37}$$

以及去中心化通信轮次

$$N_{\mathbf{W}} \geq \Omega\left(\sqrt{\kappa_W}\sqrt{\frac{L_{\mathbf{C}} L_f}{\mu_{\mathbf{C}} \mu_f}} \log\left(\frac{1}{\varepsilon}\right)\right). \tag{38}$$

梯度计算次数的下界采用与（Yarmoshik 等人，2024b）中相同的求和技巧获得：对于每个 $f_i(x_i)$，我们添加一个具有适当参数的 Nestrov 的最坏函数的独立副本 $h_i(w_i)$（变量 w_i 是新的独立变量，不同节点之间不存在任何耦合）。

H. 引理4.4 的证明

首先，分块矩阵的最大奇异值的平方上界由其各个块的最大奇异值平方之和给出，因此

$$\sigma_{\max}^2(\mathbf{B}) \le \sigma_{\max}^2(\mathbf{A}) + \gamma^2\sigma_{\max}^2(\mathbf{W}) + \beta^2\sigma_{\max}^2(\mathbf{C}).$$

我们将系数 α 和 β 设为

$$\alpha^2 = \frac{1}{\mu_W}\begin{cases} L_A + \frac{1}{4}\widetilde{\mu}_{AC}, & \widetilde{\mu}_{AC} > 0 \\ 2L_A, & \widetilde{\mu}_{AC} = 0 \end{cases}, \quad \beta^2 = \frac{1}{\mu_C}\begin{cases} L_S + \frac{1}{2}\widetilde{\mu}_{AC}, & \widetilde{\mu}_{AC} > 0 \\ L_S + 2L_A, & \widetilde{\mu}_{AC} = 0 \end{cases}, \tag{39}$$

其中 $L_S = \frac{1}{n}\sigma_{\max}^2(\mathbf{A}')$。

我们将证明 $\mathbf{B}$ 的最小正奇异值满足以下界：

$$\sigma_{\min^+}^2(\mathbf{B}) \ge \begin{cases} \frac{1}{4}\widetilde{\mu}_{AC}, & \widetilde{\mu}_{AC} > 0, \\ L_A, & \widetilde{\mu}_{AC} = 0. \end{cases} \tag{40}$$

$\sigma^2{}_{\min^+}(\mathbf{B})$ 下界的证明

由于 $\sigma^2{}_{\min^+}(\mathbf{B}) = \sigma^2{}_{\min^+}(\mathbf{B}^\top)$，我们将对后者进行界定。图像 我们有 $\ker^\perp \mathbf{B}^\top = \operatorname{Im}\mathbf{B} = \operatorname{Im}\begin{pmatrix}\mathbf{A} & \mathbf{C}\end{pmatrix} + \begin{pmatrix}\mathcal{L}_m^\perp \\ 0\end{pmatrix}$。

任取 $z \in \ker^\perp \mathbf{B}^\top$。利用 $\operatorname{Im}\mathbf{W} = \mathcal{L}_m^\perp$，我们可以将 z 表示为 $z = \begin{pmatrix}\mathbf{A} & \mathbf{C}\end{pmatrix}\xi + \begin{pmatrix}\mathbf{P}^{\mathcal{L}_m^\perp} \\ 0\end{pmatrix}\eta$，其中 ξ, η。

证明的关键技巧是将 z 分解为三个正交分量 $z = \begin{pmatrix}u \\ 0\end{pmatrix} + \begin{pmatrix}v \\ 0\end{pmatrix} + \begin{pmatrix}0 \\ w\end{pmatrix}$，其中$u = \mathbf{P}^{\mathcal{L}_m^\perp}(\mathbf{A}\xi + \mathbf{P}^{\mathcal{L}_m^\perp}\eta)$、 $v = \mathbf{P}^{\mathcal{L}_m}(\mathbf{A}\xi + \mathbf{P}^{\mathcal{L}_m^\perp}\eta) = \mathbf{P}^{\mathcal{L}_m}\mathbf{A}\xi$ 以及 $w = \mathbf{C}\xi$。

以下关系可直接由分解的定义得到：

$$u \in \mathcal{L}_m^\perp \tag{41a}$$
$$v \in \mathcal{L}_m \tag{41b}$$
$$u + v \in \operatorname{Im}\mathbf{A} + \mathcal{L}_m^\perp \tag{41c}$$
$$w \in \operatorname{Im}\mathbf{C} \tag{41d}$$
$$\begin{pmatrix}v \\ w\end{pmatrix} \in \operatorname{Im}\begin{pmatrix}\mathbf{P}^{\mathcal{L}_m}\mathbf{A} \\ \mathbf{C}\end{pmatrix} = \ker^\perp\begin{pmatrix}\mathbf{A}^\top\mathbf{P}^{\mathcal{L}_m} & \mathbf{C}^\top\end{pmatrix} \tag{41e}$$

根据（Yarmoshik 等人，2024b，引理 2）并利用上述关系，我们得到 $\|\mathbf{B}^\top z\|_2^2$ 满足

$$\|\mathbf{B}^\top z\|_2^2 \overset{(a)}{=} \left\|\begin{pmatrix}\mathbf{A}^\top(u+v) + \mathbf{C}^\top w \\ \mathbf{W}u\end{pmatrix}\right\|_2^2 \overset{(b)}{\ge} -L_A\|u\|_2^2 + \frac{1}{2}\|\mathbf{A}^\top v + \mathbf{C}^\top w\|_2^2 + \mu_W\|u\|_2^2, \tag{42}$$

其中，(a) 由（41b）得出；(b) 由杨氏不等式、（41a）以及 L_A、μ_W（16）的定义得出。

现在考虑右侧的第二项。由（41b）可得 $\mathbf{A}^\top v = \mathbf{A}^\top\mathbf{P}^{\mathcal{L}_m}v$。接着，记 $\mathbf{J} = \begin{pmatrix}\mathbf{P}^{\mathcal{L}_m}\mathbf{A} & \mathbf{C}\end{pmatrix}$，并利用

（41e）可得

$$\|\mathbf{A}^\top v + \mathbf{C}^\top w\|_2^2 = \|\mathbf{A}^\top\mathbf{P}^{\mathcal{L}_m}v + \mathbf{C}^\top w\|_2^2 \ge \mu_{\mathbf{J}}\left\|\begin{pmatrix}v \\ w\end{pmatrix}\right\|_2^2. \tag{43}$$

为了估计 $\mu_{\mathbf{J}}$，我们考虑向量 $t \in \ker^\perp\mathbf{J} = \operatorname{Im}\mathbf{A}^\top\mathbf{P}^{\mathcal{L}_m} + \operatorname{Im}\mathbf{C}^\top$，并将其分解为正交分量 $t = y + q$，其中 $y = \mathbf{P}^{\ker^\perp\mathbf{C}}t$ 且 $q = \mathbf{P}^{\ker\mathbf{C}}t$。然后应用杨氏不等式

$$\|\mathbf{J}t\|_2^2 = \|\mathbf{P}^{\mathcal{L}_m}\mathbf{A}(y+q)\|_2^2 + \|\mathbf{C}t\|_2^2 \ge -\sigma_{\max}^2(\mathbf{P}^{\mathcal{L}_m}\mathbf{A})\|y\|_2^2 + \frac{1}{2}\|\mathbf{P}^{\mathcal{L}_m}\mathbf{A}q\|_2^2 + \mu_C\|y\|_2^2. \tag{44}$$

我们来估计右端的第一项和第二项。首先，对于任意 y，有 $\|\mathbf{P}_{\mathcal{L}_m}\mathbf{A}y\|_2^2 = \left\|\frac{1}{n}\mathbf{1}_n\otimes\mathbf{A}'y\right\|_2^2 = \frac{1}{n}\|\mathbf{A}'y\|_2^2 \le \frac{1}{n}\sigma_{\max}^2(\mathbf{A}')\|y\|_2^2$。由于 $\frac{1}{n}\sigma_{\max}^2(\mathbf{A}') = \sigma_{\max}\left(\frac{1}{n}\mathbf{1}_n\mathbf{1}_n^\top\otimes\sum_{i=1}^n\mathbf{A}_i\mathbf{A}_i^\top\right) = \sigma_{\max}(\mathbf{S})$，我们将该系数记为 L_S。其次，根据 q和 t的定义，有 $q = \mathbf{P}_{\ker\mathbf{C}}t \in \operatorname{Im}\mathbf{P}_{\ker\mathbf{C}}\mathbf{A}^\top\mathbf{P}_{\mathcal{L}_m} = \operatorname{Im}\mathbf{P}_{\ker\mathbf{C}}\mathbf{A}'^\top = \ker^\perp\mathbf{A}'\mathbf{P}_{\ker\mathbf{C}}$。因此，$\|\mathbf{P}_{\mathcal{L}_m}\mathbf{A}q\|_2^2 = \left\|\frac{1}{n}\mathbf{1}_n\otimes\mathbf{A}'q\right\|_2^2 = \left\|\frac{1}{n}\mathbf{1}_n\otimes\mathbf{A}'\mathbf{P}_{\ker\mathbf{C}}q\right\|_2^2 \ge \frac{1}{n}\sigma^2\min^+(\mathbf{A}'\mathbf{P}_{\ker\mathbf{C}})\|q\|_2^2 = \widetilde{\mu}_{AC}\|q\|_2^2$。这里，如果 $\mathbf{A}'\mathbf{P}_{\ker\mathbf{C}}$是零矩阵，我们允许 $\widetilde{\mu}_{AC}$等于零。综上可得

$$\|\mathbf{J}t\|_2^2 \ge -L_S\|y\|_2^2 + \frac{1}{2}\widetilde{\mu}_{AC}\|q\|_2^2 + \mu_C\|y\|_2^2. \tag{45}$$

现在利用缩放系数的性质（39）。若 $\widetilde{\mu}_{AC} = 0$，则有 $t = y$且 $\|\mathbf{J}t\|_2^2 \ge (\mu_C - L_S)\|t\|_2^2$，从而保证 $\mu_C \ge L_S + 2L_A$我们得到 $\mu_J \ge 2L_A$。否则，若 $\widetilde{\mu}_{AC} > 0$，对于 $\mu_C \ge L_S + \frac{1}{2}\widetilde{\mu}_{AC}$我们得到 $\mu_J = \frac{1}{2}\widetilde{\mu}_{AC}$，此时通过缩放$\mathbf{C}$无法进一步增大。

将其代入（42）可得

$$\sigma_{\min^+}^2(\mathbf{B}) \ge \begin{cases} L_A, & \widetilde{\mu}_{AC} = 0, \\ \frac{1}{4}\widetilde{\mu}_{AC}, & \widetilde{\mu}_{AC} > 0, \end{cases} \tag{46}$$

其中假设 $\mu_W \ge L_A + \begin{cases} L_A, & \widetilde{\mu}_{AC} = 0 \\ \frac{1}{4}\widetilde{\mu}_{AC}, & \widetilde{\mu}_{AC} > 0 \end{cases}$。

κ_B 的界

最后，为了简化$\mathbf{B}$ 的条件数表达式，我们使用以下界：

$$\begin{aligned}\widetilde{\mu}_{AC} &= \frac{1}{n}\sigma_{\min^+}^2(\mathbf{A}'\mathbf{P}_{\ker\mathbf{C}}) \le \frac{1}{n}\sigma_{\max}^2(\mathbf{A}'\mathbf{P}_{\ker\mathbf{C}}) \le \frac{1}{n}\sigma_{\max}^2(\mathbf{A}')\sigma_{\max}^2(\mathbf{P}_{\ker\mathbf{C}}) \\ &= \frac{1}{n}\sigma_{\max}^2(\mathbf{A}') \le \frac{1}{n}\sum_{i=1}^n\sigma_{\max}^2(\mathbf{A}_i) \le \sigma_{\max}^2(\mathbf{A}) = L_A,\end{aligned} \tag{47}$$

and

$$L_S = \frac{1}{n}\sigma_{\max}^2(\mathbf{A}') \le L_A. \tag{48}$$

对于 $\widetilde{\mu}_{AC} > 0$，这给出

$$\kappa_B = \frac{L_B}{\mu_B} \le \frac{4L_A}{\widetilde{\mu}_{AC}} + \frac{4L_A + \widetilde{\mu}_{AC}}{\widetilde{\mu}_{AC}}\frac{L_W}{\mu_W} + \frac{4L_S + 2\widetilde{\mu}_{AC}}{\widetilde{\mu}_{AC}}\frac{L_C}{\mu_C} = O\left(\widetilde{\kappa}_{AC}\kappa_W + \widetilde{\kappa}_{AC}\kappa_C\right), \tag{49}$$

以及对于 $\widetilde{\mu}_{AC} = 0$

$$\kappa_B = \frac{L_B}{\mu_B} \le \frac{L_A}{L_A} + \frac{L_A + L_A}{L_A}\frac{L_W}{\mu_W} + \frac{L_S + 2L_A}{L_A}\frac{L_C}{\mu_C} = O\left(\kappa_W + \kappa_C\right). \tag{50}$$

I. 证明　耦合约束和局部约束的复杂度界（定理 4.5）

I.1. 上界

由于耦合约束需要引入辅助变量 y，而目标函数 $F(\mathbf{x})$ 不依赖于该变量，因此相对于整个变量集合并非强凸，无法应用算法1。因此，我们引入一个正则化/惩罚化的目标函数（采用增广拉格朗日的形式）

$$G(\mathbf{x},\mathbf{y}) = \sum_{i=1}^{n} f_i(x_i) + \frac{r}{2}\left\|\mathbf{A}\mathbf{x} + \alpha\mathbf{W}\mathbf{y} - \mathbf{b}\right\|_2^2, \quad r = \frac{\mu_f}{2L_A}. \tag{51}$$

下一个引理表明，$G(\mathbf{x},\ \mathbf{y})$ 确定了强凸性问题。

引理 I.1（本质上是（Yarmoshik 等人，2024b，引理 1））。 $G(\mathbf{x},\ \mathbf{y})$ 在$\mu_G = \mu_f/4)$上是 $\mathbb{R}^d \times \mathcal{Y}$强凸的，且 $(L_G = 2L_f\kappa_W^2)$-光滑，其中 $\mathcal{Y}$是所有$\mathbf{y} \in R^{mn}$ 的子空间（回忆 $A_i \in \mathbb{R}^{m\times d_i}$），满足 $\langle \mathbf{y}\ \mathbf{1}_{mn}\rangle = 0$。

证明。 令 $D_G(\mathbf{x}'、\mathbf{y}';\ \mathbf{x}、\mathbf{y})$ 表示 G 的 Bregman 散度：

$$D_G(\mathbf{x}',\mathbf{y}';\mathbf{x},\mathbf{y}) = G(\mathbf{x}',\mathbf{y}') - G(\mathbf{x},\mathbf{y}) - \langle\nabla_x G(\mathbf{x},\mathbf{y}),\mathbf{x}'-\mathbf{x}\rangle - \langle\nabla_y G(\mathbf{x},\mathbf{y}),\mathbf{y}'-\mathbf{y}\rangle. \tag{52}$$

μ_G 的值可如下得到：

$$\begin{aligned}
D_G(x',y';x,y) &= D_F(x';x) + \frac{r}{2}\left\|\mathbf{A}(x'-x) + \alpha\mathbf{W}(y'-y)\right\|_2^2 \\
&\overset{(a)}{\geq} \frac{\mu_f}{2}\left\|x'-x\right\|_2^2 + \frac{r}{4}\left\|\alpha\mathbf{W}(y'-y)\right\|_2^2 - \frac{r}{2}\left\|\mathbf{A}(x'-x)\right\|_2^2 \\
&\overset{(b)}{\geq} \frac{\mu_f}{2}\left\|x'-x\right\|_2^2 + \frac{r\alpha^2\mu_{\mathbf{W}}}{4}\left\|y'-y\right\|_2^2 - \frac{rL_{\mathbf{A}}}{2}\left\|x'-x\right\|_2^2 \\
&\geq \frac{\mu_f}{4}\left\|x'-x\right\|_2^2 + \frac{\mu_f\alpha^2\mu_{\mathbf{W}}}{8L_{\mathbf{A}}}\left\|y'-y\right\|_2^2, \\
&\overset{(c)}{\geq} \frac{\mu_f}{8}\left\|\begin{pmatrix}x'-x\\ y'-y\end{pmatrix}\right\|^2,
\end{aligned}$$

其中，(a) 由杨氏不等式得出；(b) 由 $y'-y \in \mathcal{Y}$ 和 $\ker^{\perp}\mathbf{W} = \mathcal{Y}$ 得出；(c) 是因为 $\alpha^2 \geq \frac{L_A}{\mu_W}$ 根据其在（39）中的定义。

L_G 的值可如下得到：

$$\begin{aligned}
D_G(x',y';x,y) &= D_F(x';x) + \frac{r}{2}\left\|\mathbf{A}(x'-x) + \alpha\mathbf{W}(y'-y)\right\|_2^2 \\
&\overset{(a)}{\leq} \frac{L_f}{2}\left\|x'-x\right\|_2^2 + r\left\|\alpha\mathbf{W}(y'-y)\right\|_2^2 + r\left\|\mathbf{A}(x'-x)\right\|_2^2 \\
&\leq \frac{L_f}{2}\left\|x'-x\right\|_2^2 + r\alpha^2 L_{\mathbf{W}}\left\|y'-y\right\|_2^2 + rL_{\mathbf{A}}\left\|x'-x\right\|_2^2 \\
&\leq \frac{L_f+\mu_f}{2}\left\|x'-x\right\|_2^2 + \frac{\mu_f\alpha^2 L_{\mathbf{W}}}{2L_{\mathbf{A}}}\left\|y'-y\right\|_2^2, \\
&\overset{(b)}{\leq} L_f\frac{L_{\mathbf{W}}}{\mu_{\mathbf{W}}}\left\|\begin{pmatrix}x'-x\\ y'-y\end{pmatrix}\right\|^2,
\end{aligned}$$

其中（ a) 由杨氏不等式得出；(b) 是因为 $\alpha^2 \leq 2\frac{L_A}{\mu_W}$ 根据其在（39）中的定义，以及 $\mu_f \leq L_f$. □

因此，我们重新 将 $\sum_{i=1}^n f_i(x_i)$ 与 $G(\mathbf{x},\ \mathbf{y})$ 代入问题 (C&L) 的适用于去中心化的改写中)

$$\begin{aligned}
&\min_{\mathbf{x},\mathbf{y}} \sum_{i=1}^{n} f_i(x_i) \\
&\text{s.t. } \mathbf{B}\begin{pmatrix}\mathbf{x}\\ \mathbf{y}\end{pmatrix} = \begin{pmatrix}\mathbf{b}\\ \beta\mathbf{c}\end{pmatrix}, \quad \mathbf{B} = \begin{pmatrix}\mathbf{A} & \alpha\mathbf{W}\\ \beta\mathbf{C} & 0\end{pmatrix}.
\end{aligned}$$

请注意，这不会改变该问题的最小化解。

接下来，我们应用切比雪夫预条件（第 2.4 节），并相应地将矩阵 W 和 C 替换为 $\mathbf{W}' = P_W(\mathbf{W})$, $\kappa_{W'} = O(1)$ 和 $\mathbf{C}' = P_C(\mathbf{C}^\top\mathbf{C})$, $\kappa_{C'} = O(1)$，使得 $\mathbf{B} = \begin{pmatrix} \mathbf{A} & \alpha \\ \beta & \mathbf{W}'\mathbf{C}' \end{pmatrix}$。然后，根据引理 4.4，选取按式 (39) 定义的 α 和 β，可得 $\kappa_B = O(\widetilde{\kappa}_{AC}\kappa_{W'} + \widetilde{\kappa}_{AC}\kappa_{C'}) = O(\widetilde{\kappa}_{AC})$，并由引理 I.1 $\kappa_G = \frac{8L_f\kappa_{W'}^2}{\mu_f} = O(\kappa_f)$。

最后，由于定理 D.1，算法 1 应用于问题（C&L）的以下等价重构时，其迭代复杂度为

$$\min_{\mathbf{x}\in\mathbb{R}^d, \mathbf{y}\in\mathbb{R}^{mn}} G(\mathbf{x},\mathbf{y}) \text{ s.t. } \mathbf{B}'\begin{pmatrix}\mathbf{x}\\ \mathbf{y}\end{pmatrix} = \mathbf{b}', \tag{53}$$

其中 $\mathbf{B}' = P_B(\mathbf{B}^\top\mathbf{B})$、$\kappa_{\mathbf{B}'} = O(1)$ 和 $\mathbf{b}' = \frac{P_B}{(\mathbf{B}^\top\mathbf{B})\,\mathbf{B}^\top\mathbf{B}\,\mathbf{B}^\top}\begin{pmatrix}\mathbf{b}\\ \mathbf{c}'\end{pmatrix}$、$\mathbf{c}' = \beta\frac{P_C}{(\mathbf{C}^\top\mathbf{C})\,\mathbf{C}^\top\mathbf{C}\,\mathbf{C}^\top\mathbf{c}}$ 为 $N = O(\sqrt{\kappa_G}\log\frac{1}{\varepsilon}) = O(\sqrt{\kappa_f}\log\frac{1}{\varepsilon})$.

在此情况下，算法 1 的每次迭代都需要进行 1 次 ∇f, $O(\deg P_{\mathbf{B}}) = O(\sqrt{\kappa_{\mathbf{B}}}) = O(\sqrt{\widetilde{\kappa}_{AC}})$ 乘以 $\mathbf{B}$、$\mathbf{B}^\top$ 的乘法运算，每次此类运算都需要 $O(1)$ 次乘以 A、$\mathbf{A}^\top$ 的乘法运算，$O(\deg P_C) = O(\sqrt{\kappa_C})$ 次乘以 C、$\mathbf{C}^\top$ 的乘法运算，以及 $O(\deg P_W) = O(\sqrt{\kappa_W})$ 次乘以 W 的乘法运算，即通信轮次，这给出了定理的第一部分。

注意，尽管条件 I.1 $\mathbf{y} \in \mathcal{Y}$，我们在此处仍正确地应用了引理 $\mathbf{y} = 0$算法 1 会使所有迭代值 $\mathbf{y}^k$ 保持在子空间 $\mathcal{Y}$ 内，这是由于 $\mathrm{Im}\mathbf{W} = \mathrm{Im}\mathbf{W}$。

I.2. 下界

我们的最坏问题示例属于情形 $\widetilde{\mu}_{AC} > 0$。

我们采用耦合约束下界的证明（Yarmoshik 等人，2024b，定理 2），并结合了局部约束下界证明中所用的双层通信图思路（Yarmoshik 等人，2024a，定理 2）。

令 $W \in \mathbb{R}^{n\times n}$ 的定义与附录 G.2.2 中相同。令 W_C 也是一个拉普拉斯矩阵，其 $\frac{\lambda_{\max}(W_C)}{\lambda_{\min}+(W_C)} = \kappa_C$ 按附录 G.2.2 中的方式构造，并将相应路径图中的顶点数记为 l，因此 $W_C \in \mathbb{R}^{l\times l}$。

考虑以下目标函数

$$f_i(p_i,t_i) = \frac{\mu_f}{2}\frac{1}{l}\sum_{j=1}^{l}\left\|p_{ij} + \frac{\sqrt{L_A'}}{2\mu_f}e_1\right\|_2^2 + \frac{L_f}{2}\frac{1}{l}\sum_{j=1}^{l}\|t_{ij}\|_2^2, \tag{54}$$

其中 $x_i = (p_i,\ t_i)$、p_i 和 t_i 为 $\in \ell_2^l$，而 $e_1 = (1\,0\,\ldots 0)^\top$ 表示第一坐标向量。显然，每个 f_i 都是 L_f/l-光滑且 μ_f/l-强凸的，因此 $\sum_{i=1}^n f_i(x_i)$ 的条件数确为 $\kappa_f = \frac{L_f}{\mu_f}$。对所有 $i \in \{1,\ldots,n\}$定义 $C_i = \begin{pmatrix}\sqrt{W_C}\otimes I_{\ell_2} & 0\\ 0 & \sqrt{W_C}\otimes I_{\ell_2}\end{pmatrix}$ 和 $b_i = 0$。这便得到所需的条件数：$\frac{\sigma^2\,\text{最大值}(C)\,\sigma^2}{\text{最小值}+(C)} = \frac{\lambda_{\max}(W_C)}{\lambda_{\min}+(W_C)} = \kappa_C$。局部约束 $C_i\begin{pmatrix}p_i\\ t_i\end{pmatrix} = 0$ 表达了两个共识约束：事实上，第一个块等价于 $(W_C\otimes I_{\ell_2})p_i = 0$，进而等价于 $p_{i,\,1} = \ldots = p_{i,\,l}$，因为 $\ker W_C = \{\alpha 1_l \mid \alpha\in\mathbb{R}\}$；同理，第二个块给出 $t_{i,\,1} = \ldots = t_{i,\,l}$。

请注意，根据构造，n 和 l 均可被 3 整除。我们定义索引集 $\mathcal{V}_1 = \{1,\ldots,n/3\}$、$\mathcal{V}_2 = \{n/3+1,\ldots,2n/3\}$、$\mathcal{V}_3 = \{2n/3+1,\ldots,n\}$，类似地，$\mathcal{U}_1 = \{1,\ldots,l/3\}$、$\mathcal{U}_2 = \{l/3+1,\ldots,2l/3\}$、$\mathcal{U}_3 =$

$\{2l/3+1,\ldots,l\}$。令

$$E_1 = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & \ldots \\ 0 & 1 & -1 & 0 & 0 & \ldots \\ 0 & 0 & 0 & 0 & 0 & \ldots \\ 0 & 0 & 0 & 1 & -1 & \ldots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}, \quad E_2 = \begin{pmatrix} 1 & -1 & 0 & 0 & 0 & \ldots \\ 0 & 0 & 0 & 0 & 0 & \ldots \\ 0 & 0 & 1 & -1 & 0 & \ldots \\ 0 & 0 & 0 & 0 & 0 & \ldots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix},$$

$$A_i = \begin{cases} \begin{pmatrix} \operatorname{diag}(\delta_{\mathcal{U}_1}) \otimes \sqrt{L'_A} E_1^\top & \operatorname{diag}(\delta_{\mathcal{U}_1}) \otimes \sqrt{\mu'_A} I_{\ell_2^l} \end{pmatrix}, & i \in \mathcal{V}_1 \\ 0, & i \in \mathcal{V}_2 \\ \begin{pmatrix} \operatorname{diag}(\delta_{\mathcal{U}_3}) \otimes \sqrt{L'_A} E_2^\top & \operatorname{diag}(\delta_{\mathcal{U}_3}) \otimes \sqrt{\mu'_A} I_{\ell_2^l} \end{pmatrix}, & i \in \mathcal{V}_3 \end{cases},$$

其中 $\delta_{\mathcal{U}_k} \in \mathbb{R}^l$ 和 $[\delta_{\mathcal{U}_k}]_j = \begin{cases} 1, & j \in \mathcal{U}_k \\ 0, & \text{否则。} \end{cases}$

注意，如果我们预先求解局部共识约束，就会得到与（Yarmoshik 等人，2024b，附录 C.3）中完全相同的问题。因此，此处考虑的问题的解，就是（Yarmoshik 等人，2024b，附录 C）中问题的解在每个顶点处复制 l 次。根据定义 4.2

$$\widetilde{\kappa}_{AC} = \frac{\max\limits_{i=1,\ldots,n} \lambda_{\max}(A_i A_i^\top)}{\frac{1}{n}\sigma^2_{\min^+}(\mathbf{A}'\mathbf{P}_{\ker \mathbf{C}})} \tag{55}$$

为了估计 $\max_{i=1,\ldots,n} \lambda_{\max}(A_i A_i^\top)$，我们重新排列 A_i 的列，使其成为一个分块对角矩阵

$$A_i = \begin{cases} \operatorname{diag}(\delta_{\mathcal{U}_1}) \otimes \begin{pmatrix} \sqrt{L'_A} E_1^\top & \sqrt{\mu'_A} I_{\ell_2^l} \end{pmatrix}, & i \in \mathcal{V}_1 \\ 0, & i \in \mathcal{V}_2 \\ \operatorname{diag}(\delta_{\mathcal{U}_3}) \otimes \begin{pmatrix} \sqrt{L'_A} E_2^\top & \sqrt{\mu'_A} I_{\ell_2^l} \end{pmatrix}, & i \in \mathcal{V}_3 \end{cases} =: \begin{cases} \operatorname{diag}(\delta_{\mathcal{U}_1}) \otimes \bar{A}_1, & i \in \mathcal{V}_1 \\ 0, & i \in \mathcal{V}_2 \\ \operatorname{diag}(\delta_{\mathcal{U}_3}) \otimes \bar{A}_2, & i \in \mathcal{V}_3 \end{cases} \tag{56}$$

并且根据（Yarmoshik 等人，2024b，定理 2）的证明，其各块的最大奇异值平方有上界 $\sigma^2_{\max}(\bar{A}_k) \le 2L'_A + \mu'_A = L_A$，$k \in \{1,2\}$ 如所需，只要令 $L'_A = \frac{1}{2}L_A - \frac{9}{2}\mu_A$，$\mu'_A = 9\widetilde{\mu}_{AC}$。

对于（55）的分母，有核 $C_i = \{(p_i, t_i) : p_{i1} = \ldots = p_{il}, \quad t_{i1} = \ldots = t_{il}\}$，$\mathbf{P}_{\ker C_i} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \otimes \left(\frac{1}{L} 1_l 1_l^\top \otimes I_{\ell_2}\right)$，因此，计算 $\mathbf{A}'\mathbf{P}_{\ker \mathbf{C}}$ 并按照（56）重新排列其列后，我们得到

$$\mathbf{A}'\mathbf{P}_{\ker \mathbf{C}} = \begin{pmatrix} 1^\top_{|\mathcal{V}_1|} \otimes \frac{1}{l} 1_{|\mathcal{U}_1|} 1_l^\top \otimes \bar{A}_1 & 0 & 0 \\ 0 & 0 & 1^\top_{|\mathcal{V}_2|} \otimes \frac{1}{l} 1_{|\mathcal{U}_2|} 1_l^\top \otimes \bar{A}_2 \end{pmatrix}. \tag{57}$$

例如，考虑第一个非零对角块

$$\frac{1}{n}\sigma^2_{\min^+}\left(1^\top_{|\mathcal{V}_1|} \otimes \frac{1}{l} 1_{|\mathcal{U}_1|} 1_l^\top \otimes \bar{A}_1\right) = \frac{|\mathcal{V}_1||\mathcal{U}_1| l}{nl^2}\sigma^2_{\min^+}(\bar{A}_1) = \frac{1}{9}\sigma^2_{\min^+}(\bar{A}_1) \ge \frac{1}{9}\mu'_A = \widetilde{\mu}_{AC}. \tag{58}$$

对于带有 A_2的块，同样成立，因此令 $\widetilde{\mu}_{AC} = 1$ 并令 $L_A = \widetilde{\kappa}_{AC}$， A 的条件数确实为 $\widetilde{\kappa}_{AC}$。

由此，我们验证了所考虑的问题实例满足关于目标函数取值以及约束矩阵条件数的假设。现在，让我们推导使用任意一阶方法求解该问题实例的预言机复杂度下界。

问题的解具有这样的性质：近似解中的非零分量数量 $\nu = \max\{k : \exists i,j : [p_{i,j}]_k > 0\}$以 ε 与精确解 x^* 之间的平方距离 x 为下界，即 $\nu \ge \Omega\left(\sqrt{\kappa_A \kappa_f} \log\left(\frac{1}{\varepsilon}\right)\right)$(Yarmoshik 等人, 2024b, 定理 2)。不失一般性地假设算法的初始点为零。显然，从问题的结构（尤其是 E_1的结构）可以看出， $\mathcal{V}_1$中的节点无法仅凭自身执行任何$x_{i,j}$，将其最大非零分量数量增加超过 1，所执行的操作

一阶方法允许的任何操作，即原始/对偶梯度计算、乘以 $C_i, C_i^\top, A_i, A_i^\top$以及去中心化通信（ $\mathcal{V}_3$中的节点也一样）；$\mathcal{V}_2$ 中的节点无法通过局部操作增加 ν。因此，为了将非零分量数量增加 $O(1)$，必须将信息从 $\mathcal{V}_1$ 传递到 $\mathcal{V}_3$（或反向传递），这需要 $n/3 \geq \Omega\left(\sqrt{\kappa_W}\right)$ 个通信轮次。这部分证明与耦合约束的情形相同。

现在，局部约束带来的区别在于，对于 $\mathcal{V}_1$，通过矩阵 E_1 进行乘法运算（这是它们增加 ν 的唯一方式）仅针对 $j \in \mathcal{U}_1$执行，而 E_2 则适用于 $j \in \mathcal{U}_3$。因此，任何算法都需要在 $\mathcal{U}_1$ 和 $\mathcal{U}_3$之间传递信息，才能使 ν增加 1，这需要进行 $l/3 \geq \Omega\left(\sqrt{\kappa_C}\right)$ 次乘法运算 C_i，$C_i^\top$。

综上所述，我们得到

$$N_A \geq \Omega\left(\sqrt{\widetilde{\kappa}_{AC}\kappa_f}\log\left(\frac{1}{\varepsilon}\right)\right) \tag{59}$$

$$N_W \geq \Omega\left(\sqrt{\kappa_W\widetilde{\kappa}_{AC}\kappa_f}\log\left(\frac{1}{\varepsilon}\right)\right) \tag{60}$$

$$N_C \geq \Omega\left(\sqrt{\kappa_C\widetilde{\kappa}_{AC}\kappa_f}\log\left(\frac{1}{\varepsilon}\right)\right) \tag{61}$$

通过将每个节点上带有适当参数的 Nestrov 最坏函数 $h_i(w_i)$ 的独立副本与所构造的目标函数 $f_i(x_i)$ 相加，可以得到 $N_{\nabla f}$ 的界（变量 w_i 是新的独立变量，不同节点之间不存在任何耦合）。

J. 混合约束上界的证明（定理 4.6）

定义 $\mathbf{K} = \text{diag}(\mathbf{B}\ \widetilde{\mathbf{B}})$，其中 $\mathbf{B}$ 是耦合约束和局部约束的矩阵 $\mathbf{B} = \begin{pmatrix} \mathbf{A} & \beta & \alpha W \otimes I_m \\ & & \mathbf{C}\ 0 \end{pmatrix}$，$\mathbf{A} = \text{diag}(A_1, \ldots\ldots, A_n)$，而 $\widetilde{\mathbf{B}}$ 是共识约束和局部约束的矩阵 $\widetilde{\mathbf{B}} = \begin{pmatrix} \widetilde{\mathbf{C}} \\ \gamma W \otimes I_d \end{pmatrix}$。

由于$\mathbf{K}$ 是块对角矩阵，我们分别对 $\mathbf{B}_1$ 和I.1节中所述的$\mathbf{B}_2$ 以及附录G.1中所述的，对其进行预条件化，从而得到其预条件化结果。

J.1. 非同一局部约束

证明。我们应用切比雪夫预条件（第2.4节），并将矩阵 W和 C 分别替换为 $W' = P_W(W)$、$\kappa_{W'} = O(1)$ 和 $\mathrm{C}' = P_C(\mathrm{C}^\top\mathrm{C})$、$\kappa_{C'} = O(1)$，使得 $\mathrm{B}_1 = \begin{pmatrix} \mathrm{A} & \beta & \alpha W \otimes I_m \\ & & \mathrm{C}' \quad 0 \end{pmatrix}$ 和 $\mathrm{B}_2^\top = \begin{pmatrix} \widetilde{\mathrm{C}}^\top & \gamma W' \otimes I_d \end{pmatrix}$。然后，我们将 B_1 替换为 $\mathrm{B}'_1 = P_{\mathbf{B}_1}(\mathrm{B}_1^\top\mathrm{B}_1)$、$\kappa_{\mathbf{B}'_1} = O(1)$，并将 B_2 替换为 $\mathrm{B}'_2 = P_{\mathbf{B}_2}(\mathrm{B}_2^\top\mathrm{B}_2)$、$\kappa_{\mathbf{B}'_2} = O(1)$。通过这种预条件化，我们得到 $\sigma_{\min^+}(\mathrm{B}'_1)\ \sigma_{\min^+}(\mathrm{B}'_2) \geq 11/15$ 和 $\sigma_{\max}(\mathrm{B}'_1)\ \sigma_{\max}(\mathrm{B}'_2) \leq 19/15$（Salim 等人，2022，第6节.3.2），因此对于 $\mathrm{K} = \text{diag}(\mathrm{B}'_1, \mathrm{B}'_2)$，我们得到 $\kappa_K = O(1)$。

为解决耦合约束带来的强凸性问题（参见第 I.1 节），我们对目标函数进行惩罚/正则化：$G(\mathbf{x}\ \tilde{\mathbf{x}}, \mathbf{y}) = \sum_{i=1}^n f_i(x_i, \tilde{x}_i) + \frac{r}{2}\|\mathbf{A}\mathbf{x} + \alpha\mathbf{W}'\mathbf{y} - \mathbf{b}\|_2^2$。根据（Yarmoshik 等人，2024b，引理 1），对于 $r = \frac{\mu_f}{2L_A}$，我们有 $\kappa_G = O(\kappa_f)$。

现在，由定理 D.1 可知，将算法 1 应用于问题的以下等价重构，其迭代复杂度为

(P)

$$\min_{\mathbf{x},\tilde{\mathbf{x}},\mathbf{y}} G(\mathbf{x}, \tilde{\mathbf{x}}, \mathbf{y}) \text{ s.t. } \mathbf{K}\begin{pmatrix} \mathbf{x} \\ \mathbf{y} \\ \tilde{\mathbf{x}} \end{pmatrix} = \begin{pmatrix} \mathbf{b}' \\ \mathbf{c}' \end{pmatrix}, \tag{62}$$

其中 $\mathrm{b}' = P_{\mathrm{B}_1}(\mathrm{B}_1^\top\mathrm{B}_1)\ \mathrm{B}_1^\top\mathrm{B}_1\ \mathrm{B}_1^\top \begin{pmatrix} \mathrm{b} & \mathrm{c} \end{pmatrix}$, $\mathrm{c}' = P_{\mathrm{B}_2}(\mathrm{B}_2^\top\mathrm{B}_2)\ \mathrm{B}_2^\top\mathrm{B}_2\ \mathrm{B}_2^\top \begin{pmatrix} \widetilde{\mathrm{c}} & 0 \end{pmatrix}$ 为 $N = O(\sqrt{\kappa_G}\log\frac{1}{\varepsilon}) = O(\sqrt{\kappa_f}\log\frac{1}{\varepsilon})$。

在这种情况下，算法 1 的每次迭代需要

- 1 ∇f 的计算；
- O（1）次与 $\mathbf{B}'_1$，$\mathbf{B}'^\top_1$ 相乘，相当于 O（度数为 $P_{\mathbf{B}_1}$）$= O(\sqrt{\kappa_{\mathbf{B}_1}}) = O(\sqrt{\tilde{\kappa}_{AC}})$ 次与 $\mathbf{B}_1$，$\mathbf{B}_1^\top$ 相乘，每次需要 O（1）次与 $\mathbf{A}$，$\mathbf{A}^\top$ 相乘，O（度数为 P_W）$= O(\sqrt{\kappa_W})$ 次与 $\mathbf{W}$ 相乘（通信），以及O（度数为 P_C）$= O(\sqrt{\kappa_C})$ 次与 $\mathbf{C}$、$\mathbf{C}^\top$ 相乘；
- O（1）次与 $\mathbf{B}_2$，$\mathbf{B}'^\top_2$ 相乘，相当于 O（度数为 $P_{\mathbf{B}_2}$）$= O(\sqrt{\tilde{\kappa}_{\mathbf{C}^\top}})$ 次与 $\mathbf{B}_2$，$\mathbf{B}_2^\top$ 相乘，每次需要O（1）次与 $\widetilde{\mathbf{C}}$ $\widetilde{\mathbf{C}}^\top$ 相乘，以及 O（度数为 P_W）$= O(\sqrt{\kappa_W})$ 次通信。

统计矩阵乘法的总次数，即可完成证明。 □

J.2. 相同的局部约束

$\widetilde{\quad}$ $\widetilde{\quad}$在这种情况下，我们有 $\widetilde{\mathbf{C}} = I_n \otimes C$。

证明。 切比雪夫预条件的应用方式有所不同：我们不再对 $\mathbf{B}_2$ 取多项式，而是简单地将矩阵 $\widetilde{\mathbf{C}}$ 替换为 $\widetilde{\mathbf{C}}' = P_{\widetilde{C}'} \quad \kappa_{\widetilde{C}'} = O(1) \quad 1 \quad \gamma = 1 \quad \mathbf{B}_2^\top = (\mathbf{C}),$ $\begin{pmatrix} I_n \otimes C'^\top & \gamma W' \otimes I_d \end{pmatrix}$ $\sigma_{\min^+}\ _2\ \sigma_{\max}\ _2$ $\Theta(1)\ _1\ W$ ()；根据引理 F.，这也表明 $\mathbf{B}$ 的 $(\mathbf{B})(\mathbf{B})$ 均有界，且 ()。 $\mathbf{B}$ 和目标函数的处理方式与非相同约束的情况相同。

在这种情况下，算法1 的每次迭代需要

- （与非相同情形相同） ∇f 的 1 次计算；

- （与非相同情形相同） O 次与 $\mathbf{B}_1'$，$\mathbf{B}_1'^{\top}$ 的乘法运算，相当于 O（次数为 $P_{\mathbf{B}_1}$） $= O(\sqrt{\kappa_{\mathbf{B}_1}}) = O(\sqrt{\kappa_A})$ 次与 $\mathbf{B}_1$、$\mathbf{B}_1^{\top}$ 的乘法运算，而每次又需要 $O(1)$ 次与 $\mathbf{A}$、$\mathbf{A}^{\top}$ 的乘法运算，O（次数为 P_W） $= O(\sqrt{\kappa_W})$ 次与 $\mathbf{W}$（通信）的乘法运算，以及 O（次数为 P_C） $= O(\sqrt{\kappa_C})$ 次与 $\mathbf{C}$、$\mathbf{C}^{\top}$ 的乘法运算；
- （复杂度降低了 W） $O(1)$ 次 $\mathbf{B}_2$、$\mathbf{B}_2'^{\top}$乘法运算，每次需要 O（次数为 $P_{\mathbf{C}}$） $= O(\sqrt{\kappa_{\mathbf{C}}})$ 次 $\mathbf{C}$、$\mathbf{C}^{\top}$，以及 O（次数为 P_W） $= O(\sqrt{\kappa_W})$ 次通信。

□

对于 $\mathbf{C} = 0$ 的情况，我们还给出以下有用的引理，它无需使用切比雪夫预条件，而只需对矩阵进行标量缩放，即可精确估计矩阵 $\mathbf{B}$ 的条件数。

引理 *1.* 设 $\widetilde{C}_1 = \ldots = \widetilde{C}_n = \widetilde{C}$。记 $B = diag(\alpha B_1, B_2)$，其中 $B_1 = (A\ \beta W \otimes I_m)$，$A = diag(A_1, \ldots, A_n)$，且 $B_2^{\top} = (I_n \otimes \widetilde{C}^{\top}\ \gamma W \otimes I_d)$。同时回顾定义 *3.4* 中的 S_A 和 $\widehat{\kappa}_A$ 的定义。令 $\alpha^2 = \frac{\sigma^2}{2\sigma^2}\lambda_{\min^+}(S_A)$，$\beta^2 = \frac{\lambda_{\min^+}(S_A)+\sigma^2}{\sigma^2\sigma^2}$，且 $\gamma^2 = \frac{\sigma^2}{\sigma^2\sigma^2}$，则可得

$$\begin{aligned}\sigma_{\max}^2(\mathbf{B}) &\le O\left(\sigma_{\max}^2(\widetilde{C}) + \sigma_{\min^+}^2(\widetilde{C}) \cdot \kappa_W^2 \widehat{\kappa}_A\right),\\ \sigma_{\min^+}^2(\mathbf{B}) &\ge \sigma_{\min^+}^2(\widetilde{C}),\end{aligned}$$

因此 $\kappa_B = O(\kappa_{\widetilde{C}} + \kappa_W^2 \widehat{\kappa}_A)$。

证明。 我们分别估计 $\mathbf{B}_1$ 和 $\mathbf{B}_2$的谱。令 $\beta^2 = \frac{\lambda_{\min^+}(S_A)+\sigma^2}{\sigma^2}$。根据引理 F.3，我们有

$$\begin{aligned}\sigma_{\max}^2(\mathbf{B}_1) &\le \sigma_{\max}^2(\mathbf{A}) + (\sigma_{\max}^2(\mathbf{A}) + \lambda_{\min^+}(S_A))\kappa_W^2,\\ \sigma_{\min^+}^2(\mathbf{B}_1) &\ge \frac{\lambda_{\min^+}(S_A)}{2}.\end{aligned}$$

根据引理 F.1 ，我们得到

$$\begin{aligned}\sigma_{\max}^2(\mathbf{B}_2) &= \sigma_{\max}^2(\widetilde{C}) + \gamma^2\sigma_{\max}^2(W),\\ \sigma_{\min^+}^2(\mathbf{B}_2) &= \min(\sigma_{\min^+}^2(\widetilde{C}), \gamma^2\sigma_{\min^+}^2(W)).\end{aligned}$$

回想引理陈述中 α^2 和 γ^2 的取值，我们得到关于 $\sigma_{\max}^2(\mathbf{B})$ 的估计：

$$\begin{aligned}\sigma_{\max}^2(\mathbf{B}) &\le \max(\alpha^2\sigma_{\max}^2(\mathbf{B}_1), \sigma_{\max}^2(\mathbf{B}_2))\\ &= \max\left(2\sigma_{\min^+}^2(\widetilde{C})\left(\widehat{\kappa}_A + (\widehat{\kappa}_A + 1)\kappa_W^2\right), \sigma_{\max}^2(\widetilde{C}) + \sigma_{\min^+}^2(\widetilde{C})\kappa_W^2,\right)\\ &= O\left(\sigma_{\max}^2(\widetilde{C}) + \sigma_{\min^+}^2(\widetilde{C})\kappa_W^2\widehat{\kappa}_A\right).\end{aligned}$$

类似地，我们得到 $\sigma^2{}_{\min^+}(\mathbf{B})$ 的一个界：

$$\begin{aligned}\sigma_{\min^+}^2(\mathbf{B}) &\ge \min(\alpha^2\sigma_{\min^+}^2(\mathbf{B}_1), \sigma_{\min^+}^2(\mathbf{B}_2))\\ &= \min\left(\frac{\alpha^2\lambda_{\min^+}(S_A)}{2}, \sigma_{\min^+}^2(\widetilde{C}), \gamma^2\sigma_{\min^+}^2(W)\right)\\ &= \sigma_{\min^+}^2(\widetilde{C}).\end{aligned}$$

□

K. 第 5.1 节缺失的证明5.1

概率	预言机	互补
耦合约束 (C)	梯度	$\sqrt{\frac{L_f R^2}{\varepsilon}}$
	MAT.	$\sqrt{\widehat{\kappa}_A}\sqrt{\frac{L_f R^2}{\varepsilon}}$
	通信	$\sqrt{\kappa_W}\sqrt{\widehat{\kappa}_A}\sqrt{\frac{L_f R^2}{\varepsilon}}$
	论文	本文，定理证明。5.1
共享 VAR. 约束 (10)	梯度	$\sqrt{\frac{L_f R^2}{\varepsilon}}$
	矩阵 $\widetilde{C}$	$\sqrt{\widehat{\kappa}_{\widetilde{C}^\top}}\sqrt{\frac{L_f R^2}{\varepsilon}}$
	通信	$\sqrt{\widehat{\kappa}_{\widetilde{C}^\top}}\sqrt{\kappa_W}\sqrt{\frac{L_f R^2}{\varepsilon}}$
	论文	本文，定理证明。5.1
局部 VAR. 约束 (C&L)	梯度	$\sqrt{\frac{L_f R^2}{\varepsilon}}$
	矩阵 A	$\sqrt{\widetilde{\kappa}_{AC}}\sqrt{\frac{L_f R^2}{\varepsilon}}$
	矩阵 C	$\sqrt{\widetilde{\kappa}_{AC}}\sqrt{\kappa_C}\sqrt{\frac{L_f R^2}{\varepsilon}}$
	通信	$\sqrt{\widetilde{\kappa}_{AC}}\sqrt{\kappa_W}\sqrt{\frac{L_f R^2}{\varepsilon}}$
	论文	本文，定理证明。5.1
一般混合约束 (P)	梯度	$\sqrt{\frac{L_f R^2}{\varepsilon}}$
	矩阵 A	$\sqrt{\widetilde{\kappa}_{AC}}\sqrt{\frac{L_f R^2}{\varepsilon}}$
	矩阵 C	$\sqrt{\widetilde{\kappa}_{AC}}\sqrt{\kappa_C}\sqrt{\frac{L_f R^2}{\varepsilon}}$
	矩阵 $\widetilde{C}$	$\sqrt{\widehat{\kappa}_{\widetilde{C}^\top}}\sqrt{\frac{L_f R^2}{\varepsilon}}$
	通信	$\left(\sqrt{\widetilde{\kappa}_{AC}}+\sqrt{\widehat{\kappa}_{\widetilde{C}^\top}}\right)\sqrt{\kappa_W}\sqrt{\frac{L_f R^2}{\varepsilon}}$
	论文	本文，定理证明。5.1

表 2。 带约束去中心化光滑凸优化的收敛率。常数 $\widehat{\kappa}_A$， $\widetilde{\kappa}_{AC}$， $\widehat{\kappa}_{C^\top}$， κ_C的定义与表 1 相同。省略了 对数 $\left(\frac{1}{\varepsilon}\right)$。

定理 K.1. 的证明5.1
由定理 2.4 可知，要达到 ε-解，算法 1 需要 $N_{\nabla f} = \mathcal{O}\left(\sqrt{\frac{LR^2}{\varepsilon}}\log\left(\frac{1}{\varepsilon}\right)\right)$ 次梯度计算和 $N_{\mathbf{B}} = \mathcal{O}\left(\sqrt{\frac{LR^2}{\varepsilon}}\sqrt{\kappa_{\mathbf{B}}}\log\left(\frac{1}{\varepsilon}\right)\right)$ 次与 $\mathbf{B}$和 $\mathbf{B}^\top$ 相乘的矩阵乘法。

根据引理4.4 并且 α， β 按（39）选取，可得

$$\kappa_{\mathbf{B}} = O\left(\widetilde{\kappa}_{AC}\kappa_W + \widetilde{\kappa}_{AC}\kappa_C\right).$$

对$\mathbf{B}$ 进行切比雪夫预条件需要进行 $O(\sqrt{\widetilde{\kappa}_{AC}})$ 次$\mathbf{A}$ 乘法运算、$\mathbf{A}^\top$， $O(\sqrt{\widetilde{\kappa}_{AC}}\sqrt{\kappa_C})$ 次$\mathbf{C}$ 乘法运算、$\mathbf{C}^\top$，以及 $O(\sqrt{\widetilde{\kappa}_{AC}}\sqrt{\kappa_W})$ 个通信轮次。
将引理 F.3 应用于矩阵 $\widetilde{\mathbf{B}}^\top = (\widetilde{\mathbf{C}}^\top \ \gamma\mathbf{W})$，其中 $\gamma^2 = \frac{\lambda_{\min^+}(\mathbf{S}_{\widetilde{C}^\top})+\sigma^2\max(\widetilde{\mathbf{C}}^\top)}{\sigma^2\min^+(\mathbf{W})}$，可得

$$\kappa_{\widetilde{\mathbf{B}}} = \kappa_{\widetilde{\mathbf{B}}^\top} \le 2\widehat{\kappa}_{\widetilde{C}^\top} + 2\left(\widehat{\kappa}_{\widetilde{C}^\top}+1\right)\kappa_W^2.$$

然后，对矩阵 $\widetilde{\mathbf{B}}$应用切比雪夫加速时，需要进行 $O\left(\sqrt{\widehat{\kappa}_{\widetilde{C}^\top}}\right)$ 次与 $\widetilde{\mathbf{C}}$ $\widetilde{\mathbf{C}}^\top$的乘法运算，以及 $O\left(\sqrt{\kappa_W}\sqrt{\widehat{\kappa}_{\widetilde{C}^\top}}\right)$ 次与 $\mathbf{W}$ 的乘法运算。

因此，总的来说，当对$\mathbf{K}$ 进行切比雪夫预条件需要 $O(\sqrt{\widetilde{\kappa}_{AC}})$ 次$\mathbf{A}$乘法运算、$\mathbf{A}^\top$，

$O(\sqrt{\widetilde{\kappa}_{AC}}\sqrt{\kappa_C})$ 次乘法运算$\mathbf{C}$、$\mathbf{C}^\top$，$O\left(\sqrt{\widetilde{\kappa}_{\widetilde{C}^\top}}\right)$ 次乘法运算 $\widetilde{\mathbf{C}}$ $\widetilde{\mathbf{C}}^\top$ 以及 $O\left((\sqrt{\widetilde{\kappa}_{AC}}+\sqrt{\widetilde{\kappa}_{\widetilde{C}^\top}})\sqrt{\kappa_W}\right)$ 轮通信。由此我们得出了表2中的结果。

当局部矩阵 $\widetilde{C}_i$ 相等时，对于块 $\widetilde{\mathbf{B}}$，预条件化需要 $O(\sqrt{\kappa_C})$ 次$\mathbf{C}$乘法运算、$\mathbf{C}^{\mp}$ 以及$O(\sqrt{\widetilde{\kappa}_{AC}}\sqrt{\kappa_W})$ 个通信轮次。在这种情况下，复杂度 N_C 和 N_W 变为

$$N_{\widetilde{C}} = O\left(\sqrt{\kappa_{\widetilde{C}}}\sqrt{\frac{L_f R^2}{\varepsilon}}\log\left(\frac{1}{\varepsilon}\right)\right)$$
$$N_W = O\left(\sqrt{\widetilde{\kappa}_{AC}}\sqrt{\kappa_W}\sqrt{\frac{L_f R^2}{\varepsilon}}\log\left(\frac{1}{\varepsilon}\right)\right).$$

L. 第 5.2 节中缺失的证明5.2

概率	预言机	互补
耦合约束 (C)	梯度	$\frac{M_f^2R^2}{\varepsilon^2}$
	MAT.	$\sqrt{\widehat{\kappa}_A}\frac{M_fR}{\varepsilon}$
	通信	$\sqrt{\kappa_W}\sqrt{\widehat{\kappa}_A}\frac{M_fR}{\varepsilon}$
	论文	本文，定理. 5.2
共享 VAR. 约束 (S)	梯度	$\frac{M_f^2R^2}{\varepsilon^2}$
	矩阵 $\widetilde{C}$	$\sqrt{\widehat{\kappa}_{\widetilde{C}^\top}}\frac{M_fR}{\varepsilon}$
	通信	$\sqrt{\widehat{\kappa}_{\widetilde{C}^\top}}\sqrt{\kappa_W}\frac{M_fR}{\varepsilon}$
	论文	本文，定理. 5.2
局部 VAR. 约束 (C&L)	梯度	$\frac{M_f^2R^2}{\varepsilon^2}$
	矩阵 A	$\sqrt{\widetilde{\kappa}_{AC}}\frac{M_fR}{\varepsilon}$
	矩阵 C	$\sqrt{\widetilde{\kappa}_{AC}}\sqrt{\kappa_C}\frac{M_fR}{\varepsilon}$
	通信	$\sqrt{\widetilde{\kappa}_{AC}}\sqrt{\kappa_W}\frac{M_fR}{\varepsilon}$
	论文	本文，定理. 5.2
混合约束 (P)	梯度	$\frac{M_f^2R^2}{\varepsilon^2}$
	矩阵 A	$\sqrt{\widetilde{\kappa}_{AC}}\frac{M_fR}{\varepsilon}$
	矩阵 C	$\sqrt{\widetilde{\kappa}_{AC}}\sqrt{\kappa_C}\frac{M_fR}{\varepsilon}$
	矩阵 $\widetilde{C}$	$\sqrt{\widehat{\kappa}_{\widetilde{C}^\top}}\frac{M_fR}{\varepsilon}$
	通信	$\left(\sqrt{\widetilde{\kappa}_{AC}}+\sqrt{\widehat{\kappa}_{\widetilde{C}^\top}}\right)\sqrt{\kappa_W}\frac{M_fR}{\varepsilon}$
	论文	本文，TH. 5.2

表 3. 去中心化非光滑和（非强）凸情形的收敛率。

附录 L.1 证明定理5.2

根据定理 2.5，问题 (13) 的惩罚重构可以通过梯度滑动法求解，需要 $N_{\mathbf{K}} = O\left(\frac{M_fR}{\varepsilon}\sqrt{\kappa_{\mathbf{K}}}\right)$ 次 $\mathbf{K}$ 乘法运算，$\mathbf{K}^\top$ 以及 $\overline{N_{\nabla F}} = O\left(\frac{M_f^2R^2}{\varepsilon^2} + N_{\mathbf{K}}\right)$ 次调用 F 的次梯度。

如附录 K 所示，针对 K 的切比雪夫预条件需要 $O(\sqrt{\widetilde{\kappa}_{AC}})$ 次 A 乘法运算、$\mathsf{A}^\top$，$\overline{O(\sqrt{\widetilde{\kappa}_{AC}}\sqrt{\kappa_C})}$ 次 C 乘法运算、$\mathsf{C}^\top$，$\overline{O\left(\sqrt{\widehat{\kappa}_{\widetilde{C}^\top}}\right)}$ 次 $\widetilde{\mathsf{C}}$ $\widetilde{\mathsf{C}}^\top$ 以及 $O\left((\sqrt{\widetilde{\kappa}_{AC}}+\sqrt{\widehat{\kappa}_{\widetilde{C}^\top}})\sqrt{\kappa_W}\right)$ 次通信轮次。由此，我们推导出表 3 中的结果。

M. 第 5.3 节缺失的证明5.3

概率	预言机	互补
耦合约束 (C)	梯度	$\frac{M_f^2}{\mu_f\varepsilon}$
	矩阵 A	$\sqrt{\widehat{\kappa}_A}\frac{M_f}{\sqrt{\mu_f\varepsilon}}$
	通信	$\sqrt{\kappa_W}\sqrt{\widehat{\kappa}_A}\frac{M_f}{\sqrt{\mu_f\varepsilon}}$
	论文	本文 , TH. 5.4
共享 VAR. 约束 (S)	梯度	$\frac{M_f^2}{\mu_f\varepsilon}$
	矩阵 $\widetilde{C}$	$\sqrt{\widehat{\kappa}_{\widetilde{C}^\top}}\frac{M_f}{\sqrt{\mu_f\varepsilon}}$
	通信	$\sqrt{\widehat{\kappa}_{\widetilde{C}^\top}}\sqrt{\kappa_W}\frac{M_f}{\sqrt{\mu_f\varepsilon}}$
	论文	本文 , TH. 5.4
局部 VAR. 约束 (C&L)	梯度	$\frac{M_f^2}{\mu_f\varepsilon}$
	矩阵 A	$\sqrt{\widetilde{\kappa}_{AC}}\frac{M_f}{\sqrt{\mu_f\varepsilon}}$
	矩阵 C	$\sqrt{\widetilde{\kappa}_{AC}}\sqrt{\kappa_C}\frac{M_f}{\sqrt{\mu_f\varepsilon}}$
	通信	$\sqrt{\widetilde{\kappa}_{AC}}\sqrt{\kappa_W}\frac{M_f}{\sqrt{\mu_f\varepsilon}}$
	论文	本文 , TH. 5.4
混合约束 (P)	梯度	$\frac{M_f^2}{\mu_f\varepsilon}$
	矩阵 A	$\sqrt{\widehat{\kappa}_A}\frac{M_f}{\sqrt{\mu_f\varepsilon}}$
	矩阵 C	$\sqrt{\widetilde{\kappa}_{AC}}\sqrt{\kappa_C}\frac{M_f}{\sqrt{\mu_f\varepsilon}}$
	矩阵 $\widetilde{C}$	$\sqrt{\widehat{\kappa}_{\widetilde{C}^\top}}\frac{M_f}{\sqrt{\mu_f\varepsilon}}$
	通信	$\left(\sqrt{\widetilde{\kappa}_{AC}}+\sqrt{\widehat{\kappa}_{\widetilde{C}^\top}}\right)\sqrt{\kappa_W}\frac{M_f}{\sqrt{\mu_f\varepsilon}}$
	论文	本文 , TH. 5.4

表 4。 不同类型仿射约束下去中心化非光滑强凸优化的收敛率。

~在强凸且非光滑的情形下，我们考虑有界集 8 上的问题 $\mathcal{X}\times\mathcal{X}$，其中 $\mathcal{X}=X_1\times\cdots\times X_n$；否则，假设 2.3 和假设 2.1 与 $\mu>0$ 将无法同时成立。如果一阶算法以 $\mathbf{y}^0=\mathbf{0}$ 为初始点，那么其迭代点 $\mathbf{y}^k$ 将属于线性子空间 $\mathcal{L}_m^\perp$。因此，我们可以将搜索空间限制在该子空间上，并将问题（13）的惩罚表述改写如下

$$\min_{\mathbf{x}\in\mathcal{X},\,\mathbf{y}\in\mathcal{L}_m^\perp,\,\tilde{\mathbf{x}}\in\widetilde{\mathcal{X}}} G(\mathbf{x},\mathbf{y},\tilde{\mathbf{x}})=F(\mathbf{x},\tilde{\mathbf{x}})+\frac{r^2}{\varepsilon}\left\|\mathbf{K}\begin{pmatrix}\mathbf{x}\\ \mathbf{y}\\ \tilde{\mathbf{x}}\end{pmatrix}-\mathbf{v}\right\|_2^2. \tag{63}$$

M.1. 引理证明 5.3

假设 3.3成立，且 $\mu_f\geq 0$。令 α 和 ε 满足以下条件:

$$\alpha^2=\frac{\mu_{\mathbf{A}}+L_{\mathbf{A}}}{\mu_{\mathbf{W}}},\quad \varepsilon\leq\frac{4r^2\mu_{\mathbf{A}}}{\mu_f}. \tag{64}$$

Let

$$\delta=\frac{1}{1+\frac{\mu_f\varepsilon}{4r^2L_{\mathbf{A}}}}\in(0,1). \tag{65}$$

对于任意 $\mathbf{z}=\mathrm{col}(\mathbf{x},\ \mathbf{y}\ \tilde{\mathbf{x}})$ 且 $\mathbf{z}'=\mathrm{col}(\mathbf{x}',\ \mathbf{y}',\ \tilde{\mathbf{x}}')$，其中 $\mathbf{x},\ \mathbf{x}'\in\mathcal{X}$、$\mathbf{y},\ \mathbf{y}'\in\mathcal{L}_m^{\perp}$ 和 $\tilde{\mathbf{x}}\ \tilde{\mathbf{x}}'\in\widetilde{\mathcal{X}}$，我们有 e

$$
\begin{aligned}
D_G(\mathbf{x}',\mathbf{y}',\tilde{\mathbf{x}}';\mathbf{x},\mathbf{y},\tilde{\mathbf{x}}) &= D_F(\mathbf{x}',\tilde{\mathbf{x}}';\mathbf{x},\tilde{\mathbf{x}}) + \frac{r^2}{\varepsilon}\left\|\mathbf{K}(\tilde{\mathbf{z}}'-\mathbf{z})\right\|^2 \\
&\overset{(a)}{\geq} \frac{\mu_f}{2}\left\|\mathbf{x}'-\mathbf{x}\right\|^2 + \frac{\mu_f}{2}\left\|\tilde{\mathbf{x}}'-\tilde{\mathbf{x}}\right\|^2 + \frac{r^2}{\varepsilon}\left\|\mathbf{A}(\mathbf{x}'-\mathbf{x})+\gamma\mathbf{W}(\mathbf{y}'-\mathbf{y})\right\|^2 \\
&\overset{(b)}{\geq} \frac{\mu_f}{2}\left\|\mathbf{x}'-\mathbf{x}\right\|^2 + \frac{\mu_f}{2}\left\|\tilde{\mathbf{x}}'-\tilde{\mathbf{x}}\right\|^2 + \frac{r^2(1-\delta)}{\varepsilon}\left\|\gamma\mathbf{W}(\mathbf{y}'-\mathbf{y})\right\|^2 - \frac{r^2}{\varepsilon}\left(\frac{1}{\delta}-1\right)\left\|\mathbf{A}(\mathbf{x}'-\mathbf{x})\right\|^2 \\
&\overset{(c)}{\geq} \frac{\mu_f}{2}\left\|\mathbf{x}'-\mathbf{x}\right\|^2 + \frac{\mu_f}{2}\left\|\tilde{\mathbf{x}}'-\tilde{\mathbf{x}}\right\|^2 + \frac{r^2(1-\delta)\gamma^2\mu_{\mathbf{W}}}{\varepsilon}\left\|\mathbf{y}'-\mathbf{y}\right\|^2 - \frac{r^2 L_{\mathbf{A}}}{\varepsilon}\left(\frac{1}{\delta}-1\right)\left\|\mathbf{x}'-\mathbf{x}\right\|^2 \\
&\overset{(d)}{=} \frac{\mu_f}{4}\left\|\mathbf{x}'-\mathbf{x}\right\|^2 + \frac{\mu_f}{2}\left\|\tilde{\mathbf{x}}'-\tilde{\mathbf{x}}\right\|^2 + \frac{r^2\mu_f(\mu_{\mathbf{A}}+L_{\mathbf{A}})}{4r^2L_{\mathbf{A}}+\mu_f\varepsilon}\left\|\mathbf{y}'-\mathbf{y}\right\|^2 \\
&\overset{(e)}{\geq} \frac{\mu_f}{4}\left\|\mathbf{x}'-\mathbf{x}\right\|^2 + \frac{\mu_f}{2}\left\|\tilde{\mathbf{x}}'-\tilde{\mathbf{x}}\right\|^2 + \frac{r^2\mu_f(\mu_{\mathbf{A}}+L_{\mathbf{A}})}{4r^2L_{\mathbf{A}}+4r^2\mu_{\mathbf{A}}}\left\|\mathbf{y}'-\mathbf{y}\right\|^2 \\
&= \frac{\mu_f}{4}\left\|\begin{pmatrix}\mathbf{x}'-\mathbf{x}\\ \mathbf{y}'-\mathbf{y}\end{pmatrix}\right\|^2 + \frac{\mu_f}{2}\left\|\tilde{\mathbf{x}}'-\tilde{\mathbf{x}}\right\|^2 \\
&> \frac{\mu_f}{4}\left\|\mathbf{z}'-\mathbf{z}\right\|^2,
\end{aligned}
$$

其中，(a) 由假设 2.1 和 $\mathbf{K}$ 的定义得出；(b) 由杨氏不等式得出；(c) 由 $y'-y\in\mathcal{L}_m^{\perp}$ 得出；(d) 将 δ 代入（65）并将 α^2 代入（64）得出；(e) 由（ε）中 64 的上界得出。

M.2. 定理5.4 的证明

考虑形式为（13）的题目（63），其中 F 满足假设 3.3，且 $\mu_f>0$，β，γ 的定义与第 4 节相同，且 r，α，ε 满足以下条件：

$$
r\leq\frac{M}{\sigma_{\min^+}(\mathbf{B})},\quad \alpha^2=\frac{\mu_{\mathbf{A}}+L_{\mathbf{A}}}{\mu_{\mathbf{W}}},\quad \varepsilon\leq\frac{4r^2\mu_{\mathbf{A}}}{\mu_f}. \tag{66}
$$

如引理 5.3 所示，$G(\mathbf{x},\mathbf{y}\ \tilde{\mathbf{x}})$ 在 $\mathcal{X}\times\mathcal{L}_m^{\perp}\times\widetilde{\mathcal{X}}$ 上是 $\frac{\mu_f}{2}$-强凸的。对问题（63）应用带重启过程的梯度滑动法，并利用定理 2.5，可知使用$N_K=O\left(\frac{M_f}{\sqrt{\mu_f\varepsilon}}\sqrt{\kappa_{\mathbf{K}}}\right)$次 K 和 $\mathrm{K}^{\top}$乘法运算，以及对 F 的次梯度预言机调用 $N_{\nabla F}=O\left(\frac{M_f^2}{\mu_f\varepsilon}+N_K\right)$次，即可找到一个 ε-解。对 K 进行切比雪夫预条件需要进行 $O(\sqrt{\kappa}_{AC})$ 次 A 和 $\mathrm{A}^{\top}$ 乘法运算、$O(\sqrt{\kappa}_{AC}\sqrt{\kappa_C})$ 次 C 和 $\mathrm{C}^{\top}$ 乘法运算、$O\left(\sqrt{\kappa}_{\widetilde{C}^{\top}}\right)$ 次 $\widetilde{\mathrm{C}}$ $\widetilde{\mathrm{C}}^{\top}$ 乘法运算，以及 $O\left((\sqrt{\kappa}_{AC}+\sqrt{\kappa}_{\widetilde{C}^{\top}})\sqrt{\kappa_W}\right)$ 个通信轮次。随后得到矩阵 K'，其满足 $\kappa_{\mathbf{K}'}=O(1)$。因此 $N_{K'}=O\left(\frac{M}{\sqrt{\mu_f\varepsilon}}\right)$。由此得到表 4 中的结果。