用于凸复合优化的网络无关步长去中心化非精确近端梯度法

郭路遥、施新立，*IEEE*高级会员，曹进德，*IEEE*会士，万子豪 g.

arXiv:2302.03238v2 [math.OC] 4 Mar 2023

摘要——本文提出一种基于 CTA（组合后适应）策略的新型去中心化算法，用于求解无向连通网络上的凸复合优化问题。这类问题中的局部损失函数包含光滑项和非光滑项。所提出的算法采用非协调的网络无关常数步长，只需近似求解一系列近端映射；当非光滑损失函数的近端映射可能没有解析解时，这一特点使其特别适用于求解去中心化复合优化问题。对于一般凸情形，我们证明了所提算法具有 $O(1/k)$ 收敛速率；若精确求解近端映射，则可将收敛速率提高至 $o(1/k)$。此外，在度量次正则性条件下，我们证明了该算法具有线性收敛率。数值实验验证了该算法的有效性。

索引词—**去中心化优化、非精确近端梯度方法、线性收敛、网络无关。**

I. 引言

本文考虑由 N 个计算智能体组成的网络上的以下去中心化复合优化问题:

$$\mathrm{x}^* \in \arg\min_{\mathrm{x}\in\mathbb{R}^n} \Big\{\sum_{i=1}^{N}(f_i(\mathrm{x})+g_i(\mathrm{x}))\Big\}, \tag{1}$$

其中，局部损失函数 $f_i : \mathbb{R}^n \to \mathbb{R}$ 和 $g_i : \mathbb{R}^n \to \mathbb{R}\cup\{\infty\}$ 均为凸函数，且只能由相应的智能体访问。我们假设 f_i 是 L_i-光滑的，并且g_i 适当、闭，但不一定光滑。这种设定具有一般性，常见于多种应用领域，包括多智能体控制、无线通信和去中心化机器学习 [1]。

可以说，解决去中心化场景的直接方法是将经典梯度下降法与网络结构相结合 [2]–[5]。一方面，在 [2]–[4],中，混合矩阵通过对邻居状态进行加权平均，用于实现局部状态的共识。这些算法可被视为针对等式共识约束的二次罚方法；当步长固定时，这会导致非精确收敛 [3]。为获得精确的最优解，采用了递减步长 [4]。另一方面， [5] 不通过混合矩阵对局部迭代量进行平均，而是利用相对状态的符号，提出了一种基于精确罚方法的算法。然而，这些原始方法在恒定步长下无法实现精确收敛。

原始-对偶领域近期的一系列研究提出了多种替代方法，这些方法在固定步长下能够精确收敛到最优解。当 $g_i = 0$ 和 f_i 为强凸时，已有多种线性收敛算法，包括 D-ADMM [6], DLM [7], EXTRA[8], DIGing [9], Harnessing [10], AugDGM [11], [12], 和精确扩散法 [13]。当 $g_i \neq 0$ 以及所有智能体的非光滑项均相同时，在假设 f_i 为强凸的条件下， [14]–[17] 针对问题 (1) 提出了多种线性收敛算法。当 $g_i \neq 0$ 且这些智能体的非光滑项可能各不相同时，已有 IC-ADMM [18],PG-ADMM [19], PG-EXTRA [20], NIDS [21] 和 PAD [22]。文献 [18],证明，在 f_i强凸的条件下，IC-ADMM 收敛。文献 [19]，表明，PG-ADMM 具有 $O(1/k)$遍历收敛率。基于 EXTRA，文献 [20]和 [21],分别提出了 PG-EXTRA 和 NIDS。对于 PG-EXTRA，研究表明该算法具有 $O(1/k)$ 收敛率；当光滑部分消失时，则具有 $o(1/k)$ 收敛率。对于具有网络无关步长的 NIDS，研究证明其具有 $o(1/k)$ 收敛率。文献 [22],通过二次罚方法和线性化 ADMM，提出了一种称为 PAD 的非精确收敛算法。对于 PAD，研究证明，当罚参数 $1/\tilde{\epsilon} > 0$固定时，问题 (1) 的解与 PAD 的极限点之间的距离为 $O(\tilde{\epsilon}^{\frac{1}{2}})$ 阶。此外，研究还证明，在一般凸性条件下，其具有$O(1/k)$ 遍历收敛率；在 f_i强凸的条件下，则具有线性收敛率。

根据其特征 [15]，这些算法可分为基于ATC（先适应后组合）和CTA（先组合后适应）的去中心化算法。对于基于ATC的去中心化算法，已有若干采用与网络无关的常数步长的算法，例如AugDGM [11], [12],Exact Diffusion [13], ABC-Algorithm[15], NEXT [16]/SONATA [17], 以及NIDS [21] （参见 [15] 以了解

本研究得到国家自然科学基金委员会项目（编号：61833005和62003084）以及江苏省自然科学基金项目（编号：BK20200355）的资助。（通讯作者：曹进德。）L. Guo现就职于东南大学数学学院和移动信息通信与安全前沿科学中心，中国江苏省南京市，邮编210096（电子邮箱：guo luyaoo@163.com）。X. Shi和Z. Wang现就职于东南大学网络空间安全学院，中国江苏省南京市，邮编210096（电子邮箱：xinli shi@seu.edu.cn；wwzh0328@seu.edu.cn）。J. Cao现就职于东南大学数学学院、移动信息通信与安全前沿科学中心，中国江苏省南京市，邮编210096；紫金山实验室，中国江苏省南京市，邮编211111；同时还就职于延世大学延世前沿实验室，韩国首尔03722（电子邮箱：jdcao@seu.edu.cn）。

表 I 网络上凸复合优化去中心化算法的收敛性质

算法	f_i	g_i	步长条件	Rate	精确性
PUDA [14]	强凸	$g_1=\cdots=g_N$; g_i的近端映射可高效计算	$\tau<\frac{2-\sigma_M(I-W)}{L_F}$	线性	精确
ABC-Algorithm [15]			$\tau<\frac{1}{L_F+\mu_F}$		
NEXT [16]/SONATA [17]			$\tau<\min\{\frac{\tilde{\mu}_m}{\sum_{i=1}^N L_i}\ \varpi\}$		
IC-ADMM [18]	强凸	g_i的近端映射可高效计算	$\tau<\frac{\beta\sigma_m(\mathcal{L}_+)}{L_F^2/\mu_F^2}$	收敛	精确
PG-ADMM [19]	凸		$\tau_i<\frac{1}{L_i+\beta d_i}$	遍历: $O(\frac{1}{k})$	
PG-EXTRA [20]			$\tau<\frac{1+\sigma_m(W)}{L_F}$	非遍历: $O(\frac{1}{k})_-$	
NIDS [21]			$\tau_i<\frac{2}{L_i}$, $0<\beta\le 0.5/\max_i\tau_i$	非遍历: $o(\frac{1}{k})$	
PAD [22]	凸	g_i的近端映射可高效计算	$\tau<\frac{1}{L_F+\beta\sigma_M(I-W)}$	遍历: $O(\frac{1}{k})$	非精确
	强凸		$\tau<\frac{1}{L_F^2/\mu_F+\beta\sigma_M(I-W)}$	线性	
D-iPGM	凸		$\tau_i<\frac{2}{L_i}$, $0<\beta\le 0.5/$最大值$_i\tau_i$	遍历: $O(\frac{1}{k})$ 非遍历: $o(\frac{1}{k})_-$	精确
	度量次正则			线性	

τ和 β分别是原始变量更新和对偶更新的步长。 τ_i是本地代理的原始步长。L_i是梯度的 Lipschitz 常数 ∇f_i和 $L_F=$最大值$_i\{L_i\}$。 μ_F是函数 f_i的强凸常数。$\tilde{\mu}_m$是 f_i的逐次凸近似替代函数的强凸参数。 $\varpi>0$是一个给定常数。$\mathcal{L}_+$是无符号拉普拉斯矩阵。关于 W， $\sigma_m(\cdot)$和 $\sigma_M(\cdot)$的定义可参见假设1和记号部分。

更多详情）。然而，据我们所知，现有基于CTA的去中心化算法的步长选择与网络拓扑有关。此外，在上述用于复合优化的算法中，要求 g_i 的近端映射具有解析解，或能够被高效计算。然而，在许多应用中， g_i的近端映射并不存在解析解。这类问题包括广义LASSO正则项[23], 融合 LASSO 正则项 [24], 回归八角选择与聚类算法（ OSCAR）正则项[25], 以及组 LASSO 正则项 [26]。因此，我们考虑引入非精确内层求解器，以计算近端映射的近似值。已有若干研究聚焦于非精确版本的增广拉格朗日法 [27], 原始-对偶混合梯度法 [28], 交替方向乘子法 [29]和去中心化增广拉格朗日法 [30], [31]。据我们所知，针对问题（1）的相应结果相对较少。此外，对于问题（1），当非光滑部分 g_i 彼此不同时，在某些假设下，相应算法的线性收敛率尚未得到确立。尽管 [14]表明，当 f_i 是强凸光滑函数，且g_i 是具有近端映射闭式解的非光滑凸函数时，对于复合优化问题（1），无法达到与维数无关的精确全局线性收敛，但这并不意味着在某些条件下不可能获得与维数相关的线性收敛率。因此，自然引出了以下三个问题。

问题 1。能否给出一种具有与网络无关的常数步长的基于CTA的去中心化算法?

问题 2。当 g_i 的近端映射没有解析解时，能否给出一种去中心化算法?

问题 3。能否证明所提出的去中心化算法对问题（1）具有线性收敛性?

本文旨在针对以下问题 1、2 和 3，在无向连通网络上求解问题（1）。我们提出了一种新颖的基于 CTA 的去中心化非精确近端梯度法（D-iPGM），其采用与网络无关的常数步长。在 D-iPGM 的实现中，允许对近端映射进行可计算近似，这意味着当 g_i 的近端映射没有闭式解时，所提出的算法可以降低计算成本。在收敛性分析方面，我们证明了其在一般凸性下的 $O(1/k)$ 收敛速率。当 g_i 的近端映射能够精确求解时，我们证明了 $o(1/k)$ D-iPGM 的非遍历收敛性。我们进一步证明，在度量次正则性条件下，D-iPGM 具有线性收敛率，而度量次正则性弱于强凸性。表 I 比较了 D-iPGM 与现有用于网络上凸复合优化问题的去中心化算法。

本文结构安排如下。第II 节将问题（1）转化为约束形式，并基于该重新表述提出 D-iPGM。此外，还讨论了 D-iPGM 与 PG-EXTRA [20]和 NIDS [21] 之间的一些关系。随后，第III节研究一般凸性和度量次正则性下的收敛性质。最后，第IV 节开展数值实验，并在第V节给出结论。

记号： $\mathbb{R}^n$ 表示配备内积 $\langle\cdot,\cdot\rangle$的 n维向量空间。 $\|\cdot\|_1$、$\|\cdot\|$ 和 $\|\cdot\|_\infty$分别表示 l_1范数、 l_2范数和 l_∞范数。0、I和1分别表示相应维度的零矩阵、单位矩阵和全1矩阵。对于 $\mathbf{x}_i\in\mathbb{R}^n, i=1\ldots,m$, 令col$\{\mathbf{x}1,\cdots,\mathbf{x}_N\}=[\mathbf{x}_1^\intercal,\cdots,\mathbf{x}_N^\intercal]^\intercal$。对于矩阵$A\in\mathbb{R}^{m\times n}$， $\sigma_m(A)$和 $\sigma_M(A)$分别表示 A 的最小和最大非零奇异值。若 A 为对称矩阵，则$A\succ 0$ 表示 A 为正定。令 $\mathcal{H}\succ 0$，且 $\mathcal{D}$ 为 $\mathbb{R}^n$的非空闭凸子集。对于任意$\mathbf{x}\in\mathbb{R}^n$，定义 $\|\mathbf{x}\|_{\mathcal{H}}^2:=\langle\mathbf{x}\ \mathcal{H}\mathbf{x}\rangle$， dist$_{\mathcal{H}}(\mathbf{x}\ \mathcal{D}):=\min_{\mathbf{x}'\in\mathcal{D}}\|\mathbf{x}\ -\mathbf{x}'\|_{\mathcal{H}}$,

以及 $\mathcal{P}_{\mathcal{D}}^{\mathcal{H}}(\mathbf{x}) := \arg\min_{\mathbf{x}'\in\mathcal{D}} \|\mathbf{x}-\mathbf{x}'\|_{\mathcal{H}}$。对于给定的闭适当凸函数 θ，θ 相对于 $\|\cdot\|_{\mathcal{H}}$ 的近端映射定义为 $\mathbf{prox}_{\theta}^{\mathcal{H}}(\mathbf{x}) := \arg\min_{\mathbf{y}}\{\theta(\mathbf{y}) + \frac{1}{2}\|\mathbf{y}-\mathbf{x}\|_{\mathcal{H}}^2\}$。当 $\mathcal{H}=\mathbf{I}$ 时，我们在这些记号中省略它。$\mathcal{B}_r(\bar{\mathbf{x}}) := \{\mathbf{x}: \|\mathbf{x}-\bar{\mathbf{x}}\| < r\}$ 表示以 $\mathbf{x}$ 为中心、半径为 $r>0$ 的开欧几里得范数球。

II. 去中心化非精确近端梯度法

考虑一个无向连通网络 $\mathcal{G}(\mathcal{V},\mathcal{E})$，其中 $\mathcal{V}=\{v_1,\ldots,v_N\}$ 表示顶点集，边集 $\mathcal{E}\subseteq\mathcal{V}\times\mathcal{V}$ 描述网络中的连通关系，即当且仅当 $(i,j)\in\mathcal{E}$ 时，智能体 i 与 j 之间存在通信链路。令 $\mathbf{x}_i\in\mathbb{R}^n$ 为智能体 i 持有的决策变量。为将问题 (1) 转化为共识约束形式，并确保 $\mathbf{x}_1=\cdots=\mathbf{x}_N$，我们引入全局混合矩阵 $\mathbf{W}=W\otimes\mathbf{I}_n$，其中 $W=[W_{ij}]\in\mathbb{R}^{N\times N}$，并在去中心化优化中作出以下标准假设。

假设 1. 混合矩阵 W 是对称且双随机的，并且 $\mathbf{I}-\mathbf{W}$ 的零空间为 $\mathrm{Span}(\mathbf{1})$。此外，若 $(i,j)\notin\mathcal{E}$ 且 $i\neq j$, $W_{ij}=W_{ji}=0$；否则，$W_{ij}>0$。

此外，对于该混合矩阵，与 [20]类似，我们引入另一个混合矩阵 $\tilde{\mathbf{W}}=\frac{1}{2}(\mathbf{I}+\mathbf{W})$。根据假设1，约束 $\mathbf{x}_1=\cdots=\mathbf{x}_N$ 等价于恒等式 $(\mathbf{I}-\mathbf{W})\mathbf{x}=\mathbf{0}$，而这又等价于 $\sqrt{\mathbf{I}-\mathbf{W}}\mathbf{x}=\mathbf{0}$。于是，问题（1）可等价地表述为：

$$\min_{\mathbf{x}} \underbrace{\sum_{i=1}^{N} f_i(\mathbf{x}_i)}_{:=F(\mathbf{x})} + \underbrace{\sum_{i=1}^{N} g_i(\mathbf{x}_i)}_{:=G(\mathbf{x})},\ \text{s.t.}\ \sqrt{\mathbf{I}-\mathbf{W}}\mathbf{x}=\mathbf{0}, \tag{2}$$

其中 $\mathbf{x}=[\mathbf{x}_1^{\mathsf{T}},\cdots,\mathbf{x}_N^{\mathsf{T}}]^{\mathsf{T}}\in\mathbb{R}^{Nn}$。为保证（1）的适定性，需要假设（1）的最优解存在。全文中，我们作出如下假设。

假设2。 局部函数 f_i 是凸的且 L_i 光滑。

根据 [32, 定理5。8]和假设2，对于任意 $\mathbf{x}$，$\mathbf{y}$，$\mathbf{z}\in\mathbb{R}^{Nn}$，有以下常用不等式

$$\langle\nabla F(\mathbf{x})-\nabla F(\mathbf{y}),\mathbf{x}-\mathbf{y}\rangle \geq \|\nabla F(\mathbf{x})-\nabla F(\mathbf{y})\|^2_{\mathbf{L}_F^{-1}}, \tag{3}$$

$$\langle\mathbf{x}-\mathbf{y},\nabla F(\mathbf{z})\rangle \leq F(\mathbf{x})-F(\mathbf{y})+\frac{1}{2}\|\mathbf{y}-\mathbf{z}\|^2_{\mathbf{L}_F}, \tag{4}$$

$$\langle\mathbf{x}-\mathbf{y},\nabla F(\mathbf{z})-\nabla F(\mathbf{x})\rangle \leq \frac{1}{4}\|\mathbf{y}-\mathbf{z}\|^2_{\mathbf{L}_F}, \tag{5}$$

其中 $\mathbf{L}_F=\mathbf{diag}\{L_1\mathbf{I}_n,\cdots,L_N\mathbf{I}_n\}$。

A. 算法开发

回顾 [15]中提出的ABC统一框架

$$\mathbf{s}^k = \mathrm{A}\mathbf{x}^k - \tau\mathrm{B}\nabla F(\mathbf{x}^k) - \boldsymbol{\lambda}^k, \tag{6a}$$

$$\boldsymbol{\lambda}^{k+1} = \boldsymbol{\lambda}^k + \mathrm{C}\mathbf{s}^k, \tag{6b}$$

$$\mathbf{x}^{k+1} = \mathrm{prox}_{\tau G}(\mathbf{s}^k), \tag{6c}$$

其中，A、B、C是适当选取的权重矩阵，$\tau>0$ 是原始步长，且非光滑函数 g_i 对所有智能体都是相同的。令 $\mathrm{C}=\frac{1}{2}(\mathbf{I}-\mathbf{W})$。根据这一统一框架，当 $\mathrm{A}=\mathrm{B}\neq\mathbf{I}$ 时，通过消除对偶变量 $\boldsymbol{\lambda}$，可得

$$\begin{aligned}\mathbf{x}^{k+1} &= \mathrm{prox}_{\tau G}(\mathbf{s}^k),\\ \mathbf{s}^{k+1} &= \tilde{\mathbf{W}}\mathbf{s}^k + \underbrace{\mathrm{A}\underbrace{(\mathbf{x}^{k+1}-\mathbf{x}^k+\tau\nabla F(\mathbf{x}^k)-\tau\nabla F(\mathbf{x}^{k+1}))}_{\text{First: Adapt}}}_{\text{Then: Combine}},\end{aligned}$$

而这些算法是基于 ATC 的去中心化算法。类似地，当 $\mathrm{B}=\mathbf{I}$ 时，可以推导出

$$\begin{aligned}\mathbf{x}^{k+1} &= \mathrm{prox}_{\tau G}(\mathbf{s}^k),\\ \mathbf{s}^{k+1} &= \tilde{\mathbf{W}}\mathbf{s}^k + \underbrace{\mathrm{A}\underbrace{(\mathbf{x}^{k+1}-\mathbf{x}^k)}_{\text{First: Combine}}+\tau\nabla F(\mathbf{x}^k)-\tau\nabla F(\mathbf{x}^{k+1})}_{\text{Then: Adapt}},\end{aligned}$$

而这些算法是基于 CTA 的去中心化算法。对于基于 ATC 的算法，需要通信状态信息和局部梯度信息。采用这种策略，基于 ATC 的算法的步长条件通常与网络拓扑无关，例如 NIDS[21]，它交换梯度适配估计，即 $2\mathbf{x}^{k+1}-\mathbf{x}^k-\tau(\nabla F(\mathbf{x}^{k+1})-\nabla F(\mathbf{x}^k))$，NIDS 的步长条件为 $0<\tau<2/L_F$，其中 $L_F=\max_{i\in\mathcal{V}}\{L_i\}$。对于基于 ATC 的算法，它们只需共享状态信息；但是，其步长条件通常与网络拓扑有关，例如 PG-EXTRA [20]，它仅交换状态信息，而 PG- EXTRA 的步长条件为 $0<\tau<(1+\sigma_m(\mathbf{W}))/L_F$。为了结合这两类算法各自的优势，我们针对问题（1）提出以下去中心化算法。

$$\tilde{\mathbf{x}}^k = \mathrm{prox}_G^{\Gamma^{-1}}(\mathbf{x}^k-\Gamma(\nabla F(\mathbf{x}^k)-\boldsymbol{\lambda}^k)), \tag{7a}$$

$$\boldsymbol{\lambda}^{k+1} = \boldsymbol{\lambda}^k - \beta(\mathbf{I}-\mathbf{W})\tilde{\mathbf{x}}^k, \tag{7b}$$

$$\mathbf{x}^{k+1} = \tilde{\mathbf{x}}^k - \Gamma(\boldsymbol{\lambda}^k-\boldsymbol{\lambda}^{k+1}), \tag{7c}$$

其中 $\Gamma=\mathrm{diga}\{\tau_1\mathbf{I}_n,\ \cdots,\ \tau_N\mathbf{I}_n\}$ 是原始步长矩阵，$\beta>0$ 是对偶步长，$\boldsymbol{\lambda}^k=\sqrt{\mathbf{I}-\mathbf{W}}\boldsymbol{\alpha}^k$ 是“缩放的”拉格朗日乘子，而 $\boldsymbol{\alpha}^k$ 是拉格朗日乘子。为了实现网络无关的步长，我们采用简单的原始校正步（7c）。在第 III节中，我们将证明，当 $0<\tau_i<2/L_i$，$i\in\mathcal{V}$，$0<\beta\leq 0.5/\max_i\tau_i$ 时，算法（7）收敛。消去对偶变量 $\boldsymbol{\lambda}$ 后，有

$$\begin{aligned}\tilde{\mathbf{x}}^k &= \mathrm{prox}_G^{\Gamma^{-1}}(\mathbf{s}^k),\\ \mathbf{x}_c^k &= \beta\Gamma(\mathbf{I}-\mathbf{W})\tilde{\mathbf{x}}^k,\\ \mathbf{x}^{k+1} &= \tilde{\mathbf{x}}^k - \mathbf{x}_c^k,\\ \mathbf{s}^{k+1} &= \mathbf{s}^k - \mathbf{x}^k + \tilde{\mathbf{x}}^k - 2\mathbf{x}_c^k + \Gamma\nabla F(\mathbf{x}^k) - \Gamma\nabla F(\mathbf{x}^{k+1}).\end{aligned}$$

显然，所提出算法的实现只需要邻居变量 $\tilde{\mathrm{x}}_j^k$，并且每次迭代中只需进行一轮通信。与（6）相比，我们知道，所提出的算法确实是一种基于CTA的去中心化算法，并且独立于这一原始-对偶算法框架。

算法1D-iPGM

输入： 混合 矩阵 $W, 0 < \tau_i < 2/L_i$, 和 $0 < \beta \leq 0.5/$最大值$_i \tau_i$。令 $\{\varepsilon_k\}$ 为一个可求和序列非负数。

1: 初始化 $\mathbf{x}^0 \in \mathbb{R}^{Nn}$, $\boldsymbol{\lambda}^0 = \mathbf{0}$.

2: **对于** $k = 1, 2, \ldots\ldots, K$ **执行**

3: 原始变量更新：

$$\tilde{\mathrm{x}}_i^k \approx \tilde{\mathrm{y}}_i^k = \mathrm{prox}_{\tau_i g_i}(\mathrm{x}_i^k - \tau_i(\nabla f_i(\mathrm{x}_i^k) - \lambda_i^k)), \quad (8)$$

使得存在一个向量

$$\mathrm{d}_i^k \in \partial g_i(\tilde{\mathrm{x}}_i^k) + \nabla f_i(\mathrm{x}_i^k) - \lambda_i^k + \frac{1}{\tau_i}(\tilde{\mathrm{x}}_i^k - \mathrm{x}_i^k), \quad (9a)$$

$$\text{satisfying } \|\mathrm{d}_i^k\| \leq \varepsilon_k. \quad (9b)$$

4: 智能体 i将 $\tilde{\mathbf{x}}_i^k$ 与邻居交换。

5：对偶更新：

$$\lambda_i^{k+1} = \lambda_i^k - \beta(\tilde{\mathrm{x}}_i^k - \sum_{j=1}^N W_{ij}\tilde{\mathrm{x}}_j^k). \quad (10)$$

6：校正步骤：$\mathbf{x}_i^{k+1} = \tilde{x}_i^k - \tau_i(\lambda_i^k - \lambda_i^{k+1})$.

7: **en d for**

输出： $\mathbf{x}^K$.

另一方面，当 g_i 的近端映射不存在解析解时，非常精确地求解子问题（7a）是一种浪费。更高效的做法是制定适当的条件，并在满足条件后停止子问题求解过程，即内层迭代。通过为子问题（7a）引入绝对可求和误差准则，我们提出了 D-iPGM（见算法 1）。

在算法1中，为了节省计算开销，在 D-iPGM 的第 k次迭代中，允许采用非精确近端梯度下降，即我们可以通过求得一个近似解来非精确地求解子问题（8）。这意味着我们必须找到一种合适的内层迭代方法来非精确地求解该子问题。对于大多数应用而言，这是可行的。例如，对于一般的 LASSO 正则项 [23], 融合 LASSO 正则项 [24],OCSAR 正则项 [25],以及组 LASSO 正则项[26],，其中 g_i具有如下形式：$g_i(\mathrm{x}) = r_{1,i}(\mathrm{x}) + r_{2,i}(\mathrm{D}_i\mathrm{x}_i)$，其中 $r_{1,i}$、$r_{2,i}$ 均为非光滑函数，D_i 为给定矩阵，我们可以使用算子分裂方法 [33] 或加速线性化ADMM [34]，结合内层停止准则（9）求解（8）。

注1。 由于 g_i 是凸、闭且适当的，根据近端映射的定义，每个子问题 (8)都是强凸的，因此每个 y_i^k 都由$(\mathrm{x}_i^k, \lambda_i^k)$唯一确定。由 *[29]*, 可知，对于任意 $\varepsilon_k \geq 0$，均可找到某个 $\tilde{\mathrm{x}}^k$ 使绝对可求和误差准则 (9) 成立。因此，算法 *1* 定义良好。此外，根据 *[28,*引理*3*，*2]*, 算法*1* 的内层循环总是在有限次迭代后终止。

注2。 由（9a）和近端映射的定义可知，$\tilde{\mathrm{x}}_i^k = \tilde{\mathrm{y}}_i^k$ 当 $\mathrm{d}_i^k = \mathbf{0}$时。因此，$\tilde{\mathrm{x}}_i^k$ 可视为精确解 $\tilde{\mathrm{y}}_i^k$的近似解。为简化收敛性分析，我们给出内层停止准则 （9）。然而，在实践中，我们也可以使用 $\|\mathrm{x}_i^{k,l+1} -$, $\mathrm{x}_i^{k,l}\| \leq \varepsilon_0\varepsilon_k$ 作为内层停止准则，其中 $\varepsilon_0 > 0$ 为常数，而 $\mathrm{x}_i^{k,l}$ 是 l内层迭代中 $\tilde{\mathrm{y}}_i^k$的估计值，因为$\mathrm{d}_i^{k,l+1} \in \partial g_i$、$(\mathrm{x}_i^{k,l+1}$、$)+\nabla f_i(\mathrm{x}_i^k)-\lambda_i^k+\frac{1}{\tau_i}(\mathrm{x}_i^{k,l+1}-$、$\mathrm{x}_i^k)$可以由 $\|\mathrm{x}_i^{k,l+1}-$、$\mathrm{x}_i^{k,l}\|$控制，即存在常数ε_0 使得 $\|\mathrm{d}_i^{k,l+1}\| \leq \varepsilon_0\|$、$\mathrm{x}_i^{k,l+1}-$、$\mathrm{x}_i^{k,l}\|$。我们将在第*IV*节对此进行具体说明。

注3。 令 $\mathbf{d}^k =$列向量$\{\mathrm{d}_1^k, \cdots, \mathrm{d}_N^k\}$。不同于 *[28]*，我们要求 $\|\mathrm{d}_i^k\| \leq \varepsilon_k$，而非 $\|\mathbf{d}^k\| \leq \varepsilon_k \max\{\beta, \|\tilde{\mathbf{x}}^k\|\}$。因此，通过这种设定，所有智能体都有各自独立的子问题，以及针对这些子问题的独立误差准则，这意味着这些子问题可以通过去中心化方式求解。此外，为了控制残差d_i^k，我们引入一个可求和序列 $\{\varepsilon_k\}$，这意味着我们可以通过 $\varepsilon_k = \varepsilon_0/(k+1)^r$ 取任意 $\varepsilon_0 > 0$和 $r > 1$，从而选取 ε_k；或者在 $0 < r < 1$时通过 $\varepsilon_k = r^k$ 选取。显然，当 g_i 的近端映射具有闭式表示时，我们可以令 $\varepsilon_k \equiv 0, k \geq 0$。在这种情况下，我们使用精确近端梯度下降进行原始变量更新，因此其复杂度与现有近端梯度法 *[14]–[22]* 相同。

B. 与 *PG-EXTRA [20]*和 *NIDS [21]*的比较

在本小节中，我们从算子分裂的角度比较 D-iPGM（$\varepsilon_k = 0$）、PG- EXTRA [20], 和 NIDS [21]。回顾问题(2)，它等价于

$$\min_{\mathbf{x}} F(\mathbf{x}) + G(\mathbf{x}) + \delta_{\{\mathbf{0}\}}(\mathbf{V}\mathbf{x}), \quad (11)$$

其中$\mathbf{V} = \sqrt{\mathbf{I} - \mathbf{W}}$，且 $\delta_{\{\mathbf{0}\}}(\mathbf{V}\mathbf{x})$ 是定义为 $\delta_{\{\mathbf{0}\}}(\mathbf{V}\mathbf{x}) = \mathbf{0}$ 当$\mathbf{V}\mathbf{x} = \mathbf{0}$时的指示函数；否则，$\delta_{\{\mathbf{0}\}}(\mathbf{V}\mathbf{x}) = \infty$，它对约束 $\mathbf{V}\mathbf{x} = \mathbf{0}$进行编码。为使表述简洁，我们定义以下算子

$$\boldsymbol{A} = \begin{pmatrix} \partial G & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}, \boldsymbol{B} = \begin{pmatrix} \mathbf{0} & -\mathbf{V} \\ \mathbf{V} & \mathbf{0} \end{pmatrix}, \boldsymbol{C} = \begin{pmatrix} \nabla F & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}.$$

根据 [35,第11.3节。3], PG-EXTRA 等价于应用于问题（[36], [37] 11）的 C-V分裂算法。此外，它还可以通过在由以下度量定义的度量下进行前向-后向算子分裂而推导得到 [33]

$$\mathbf{H}_{\mathrm{P}} = \begin{pmatrix} \frac{1}{\tau}\mathbf{I} & \mathbf{V} \\ \mathbf{V} & 2\tau\mathbf{I} \end{pmatrix},$$

也就是说，它可以改写为

$$\mathbf{u}^{k+1} = (\mathbf{H}_{\mathrm{P}} + \boldsymbol{A} + \boldsymbol{B})^{-1}(\mathbf{H}_{\mathrm{P}} - \boldsymbol{C})\mathbf{u}^k,$$

其中 $\mathbf{u}^k = (\mathbf{x}^k, \boldsymbol{\alpha}^k)$。由于$\boldsymbol{\lambda}^k = \sqrt{\mathbf{I} - \mathbf{W}}\boldsymbol{\alpha}^k$, PG-EXTRA 的原始-对偶更新为

$$\mathbf{x}^{k+1} = \mathrm{prox}_{\tau G}(\mathbf{x}^k - \tau(\nabla F(\mathbf{x}^k) - \boldsymbol{\lambda}^k)), \quad (12a)$$

$$\boldsymbol{\lambda}^{k+1} = \boldsymbol{\lambda}^k - \frac{1}{2\tau}(\mathbf{I} - \mathbf{W})(2\mathbf{x}^{k+1} - \mathbf{x}^k). \quad (12b)$$

消去对偶变量 $\boldsymbol{\lambda}$后，有

$$\mathbf{x}^{k+1} = \mathrm{prox}_{\tau G}(\mathbf{s}^k), \quad (13a)$$

$$\mathbf{s}^{k+1} = \mathbf{s}^k - \mathbf{x}^{k+1} + \tilde{\mathbf{W}}(2\mathbf{x}^{k+1} - \mathbf{x}^k) + \tau\nabla F(\mathbf{x}^k) - \tau\nabla F(\mathbf{x}^{k+1}). \quad (13b)$$

表 II D-IPGM（$\varepsilon_k = 0$）与PG-EXTRA [20] 和 NIDS [21] 的比较。

算法	通信	步长是否需要协调?	非遍历收敛速率	遍历收敛率	线性收敛率	非精确迭代
PG-EXTRA [20]	$2\mathbf{x}^{k+1}-\mathbf{x}^k$	$\tau<(1+\sigma_m(\mathbf{W}))/L_F$，否	$O(1/k)$	没有结果	没有结果	✗
NIDS [21]	$\tau(\nabla F(\mathbf{x}^k)-\nabla F(\mathbf{x}^{k+1}))$ $+2\mathbf{x}^{k+1}-\mathbf{x}^k$	$\tau_i<2/L_i$，是的	$o(1/k)$	无结果	无结果	✗
D-iPGM，$\varepsilon_k=0$	$\tilde{\mathbf{x}}^k$	$\tau_i<2/L_i$，是	$o\ (1/k)$	$O\ (1/k)$	✓	✓

在 PG-EXTRA 的实现中，仅涉及状态变量 $\mathbf{x}_i$ 的通信。由于$\mathbf{W}$ 的特征值位于 $(-1,1]$，且特征值 1 的重数为 1，因此度量 $\mathbf{H}_{\rm P}$ 是正定的。为确保 PG-EXTRA 的收敛性，要求以下度量矩阵为正定

$$\widehat{\mathbf{H}}_{\rm P}=\mathbf{H}_{\rm P}-\frac{1}{2}{\rm diag}\{L_F\mathbf{I},\mathbf{0}\},$$

其中 $L_F=\max_{i\in\mathcal{V}}\{L_i\}$。因此，PG-EXTRA 的步长条件为 $0<\tau<(1+\sigma_m(\mathbf{W}))/L_F$。

对于 NIDS，从 [35, 第 11.3 节3],来看，它等价于原始-对偶三算子分裂算法（PD3O）[38] 应用于问题（11），其度量矩阵为

$$\mathbf{H}_{\rm N}=\begin{pmatrix}\frac{1}{\tau}\mathbf{I} & \mathbf{0}\\ \mathbf{0} & \tau(2\mathbf{I}-\mathbf{V}^2)\end{pmatrix}.$$

NIDS 的原始-对偶更新为

$$\mathbf{x}^{k+1}=\mathrm{prox}_{\tau G}(\mathbf{x}^k-\tau(\nabla F(\mathbf{x}^k)-\boldsymbol{\lambda}^k)),\tag{14a}$$

$$\boldsymbol{\lambda}^{k+1}=\boldsymbol{\lambda}^k-\frac{1}{2\tau}(\mathbf{I}-\mathbf{W})\big(2\mathbf{x}^{k+1}-\mathbf{x}^k+\tau\nabla F(\mathbf{x}^k)-\tau\nabla F(\mathbf{x}^{k+1})\big).\tag{14b}$$

消去对偶变量 $\boldsymbol{\lambda}$后，可得

$$\mathbf{x}^{k+1}=\mathrm{prox}_{\tau G}(\mathbf{s}^k),\tag{15a}$$

$$\mathbf{s}^{k+1}=\mathbf{s}^k-\mathbf{x}^{k+1}+\tilde{\mathbf{W}}\big(2\mathbf{x}^{k+1}-\mathbf{x}^k+\tau\nabla F(\mathbf{x}^k)-\tau\nabla F(\mathbf{x}^{k+1})\big).\tag{15b}$$

与 PG-EXTRA 的原始-对偶版本（$(\nabla F(\mathbf{x}^k)$ 和$\nabla F(\mathbf{x}^{k+1}))$（12）相比，为了获得比 PG-EXTRA 更宽松的收敛条件，对偶更新中增加了两个项。此外，在 NIDS 的实现中，需要通信状态变量 x_i和 ∇f_i。由于 $2\mathbf{I}-\mathbf{V}^2=\mathbf{I}+\mathbf{W}\succ 0$，度量$\mathbf{H}_{\rm N}$ 是正定的。类似地，为确保 NIDS 收敛，度量矩阵

$$\widehat{\mathbf{H}}_{\rm N}=\mathbf{H}_{\rm N}-\frac{1}{2}{\rm diag}\{L_F\mathbf{I},\mathbf{0}\},$$

必须是正定的。因此，NIDS 的步长条件为 $0<\tau<2/L_F$。

受非对称前向-后向-伴随分裂[39], 这一非常通用的三算子分裂方案以及预测-校正框架 [40],的启发，所提出的 D-iPGM 可视为带有原始校正器的三角预条件前向-后向算子分裂算法。

为更清晰地说明这一点，我们定义三角预条件器$\mathbf{Q}_{\rm D}$ 和校正器$\mathbf{M}_{\rm D}$

$$\mathbf{Q}_{\rm D}=\begin{pmatrix}\frac{1}{\tau}\mathbf{I} & \mathbf{V}\\ \mathbf{0} & 2\tau\mathbf{I}\end{pmatrix},\ \mathbf{M}_{\rm D}=\begin{pmatrix}\mathbf{I} & \tau\mathbf{V}\\ \mathbf{0} & \mathbf{I}\end{pmatrix}.$$

令 $\varepsilon_k=0$、$\Gamma=\tau\mathbf{I}$ $\tau>0$ 以及 $\beta=\frac{1}{2\tau}$。根据这些定义，D-iPGM 等价于

$$\text{[D-iPGM]}:\quad \tilde{\mathbf{u}}^k=(\mathbf{Q}_{\rm D}+\boldsymbol{A}+\boldsymbol{B})^{-1}(\mathbf{Q}_{\rm D}-\boldsymbol{C})\mathbf{u}^k,$$
$$\mathbf{u}^{k+1}=\mathbf{u}^k-\mathbf{M}_{\rm D}(\mathbf{u}^k-\tilde{\mathbf{u}}^k).$$

通过消去对偶变量，可得

$$\tilde{\mathbf{x}}^k=\mathrm{prox}_{\tau G}(\mathbf{s}^k),$$
$$\mathbf{x}^{k+1}=\tilde{\mathbf{W}}\tilde{\mathbf{x}}^k,$$
$$\mathbf{s}^{k+1}=\mathbf{s}^k-\mathbf{x}^k-\mathbf{W}\tilde{\mathbf{x}}^k+\tau\nabla F(\mathbf{x}^k)-\tau\nabla F(\mathbf{x}^{k+1}).$$

不同于前向-后向算子分裂，预条件子 $\mathbf{Q}_{\rm D}$ 不要求是正定的。为实现与网络无关的收敛条件，采用校正步骤 $\mathbf{u}^{k+1}=\mathbf{u}^k-\mathbf{M}_{\rm D}(\mathbf{u}^k-\tilde{\mathbf{u}}^k)$，从而使 D-PGM 的度量矩阵

$$\mathbf{H}_{\rm D}=\mathbf{Q}_{\rm D}\mathbf{M}_{\rm D}^{-1}={\rm diag}\{\frac{1}{\tau}\mathbf{I},2\tau\mathbf{I}\}\succ 0,$$

$$\widehat{\mathbf{H}}_{\rm D}=\mathbf{Q}_{\rm D}^{\rm T}+\mathbf{Q}_{\rm D}-\mathbf{M}_{\rm D}^{\rm T}\mathbf{H}_{\rm D}\mathbf{M}_{\rm D}-\frac{1}{2}{\rm diag}\{L_F\mathbf{I},\mathbf{0}\}=\widehat{\mathbf{H}}_{\rm N},$$

为对角矩阵。因此，与 NIDS 类似，D-iPGM 的步长条件为 $0<\tau<2/L_F$。我们在表II中将其与 PG-EXTRA 和 NIDS 进行比较。通过比较可以观察到，D-iPGM 生成的原始序列 $\{\mathbf{x}^k\}$ 不同于 PG-EXTRA 和 NIDS。因此，我们得出结论：D-iPGM 是一种独特的算法，有别于 PG-EXTRA 和 NIDS。为获得区别于 NIDS 的更宽松收敛条件（或与网络无关的步长条件），D-iPGM 采用了额外的原始校正步。每个智能体都可以独立执行该原始校正步，而无需交换任何私有信息。此外，与 PG-EXTRA 相比，额外的计算成本几乎可以忽略不计。值得注意的是，在不存在近端项时，NIDS 和 D-iPGM 生成的序列 $\mathbf{x}^k$ 是相同的。此外，如第 IV-A节所示，D-iPGM（$\varepsilon_k=0$）与 NIDS 在求解复合优化问题时表现相似。

III. 收敛性分析

本节分析 D-iPGM 的收敛性和收敛速率。我们将在变分不等式框架下进行收敛性分析 [40]，，并且

接下来对所提出的 D-iPGM 进行的分析将基于以下基本引理。

引理 *1*（[40]）。设 $\theta_1(x)$ 和 $\theta_2(x)$ 为适当闭凸函数。若 θ_1 可微，且问题 $min\{\theta_1(x)+\theta_2(x): x\in\mathbb{R}^n\}$ 的解集非空，则 $x^*\in arg\ min\{\theta_1(x)+\theta_2(x): x\in\mathbb{R}^n\}$ 当且仅当 $\theta_2(x)-\theta_2(x^*)+\langle x-x^*, \nabla\theta_1(x^*)\rangle\geq 0, \forall x\in\mathbb{R}^n$.

回顾问题（2）的拉格朗日函数：$L(\mathbf{x}\ \boldsymbol{\alpha}) = F(\mathbf{x}) + G(\mathbf{x}) - \langle\boldsymbol{\alpha},\mathbf{V}\mathbf{x}\rangle$，其中 $\mathbf{V}=\sqrt{\mathbf{I}-\mathbf{W}}$ 且 $\boldsymbol{\alpha}\in\mathbb{R}^{Nn}$ 为拉格朗日乘子。根据强对偶性，为求解问题（2），我们可以考虑鞍点问题

$$\min_{\mathbf{x}}\max_{\boldsymbol{\alpha}} F(\mathbf{x})+G(\mathbf{x})-\langle\boldsymbol{\alpha},\mathbf{V}\mathbf{x}\rangle. \tag{16}$$

定义 $\mathcal{M}^*$ 为上述鞍点问题的鞍点集合。成立 $\mathcal{M}^*=\mathcal{X}^*\times\mathcal{Y}^*$，其中 $\mathcal{X}^*$ 是问题（2）的最优解集，而 $\mathcal{Y}^*$ 是其对偶问题的最优解集。将 KKT 映射记为

$$\mathcal{J}(\mathbf{u})=\begin{pmatrix}\partial G(\mathbf{x})+\nabla F(\mathbf{x})-\mathbf{V}\boldsymbol{\alpha}\\ \mathbf{V}\mathbf{x}\end{pmatrix}.$$

当且仅当 $\mathbf{0}\in\mathcal{J}(\mathbf{x}^*,\boldsymbol{\alpha}^*)$ 时，我们有 $(\mathbf{x}^*,\boldsymbol{\alpha}^*)\in\mathcal{X}^*\times\mathcal{Y}^*$。点 $\mathbf{u}^*=(\mathbf{x}^*,\boldsymbol{\alpha}^*)\in\mathcal{M}^*\subseteq\mathbb{R}^{Nn}\times\mathbb{R}^{Nn}:=\mathcal{M}$ 若满足 $L(\mathbf{x}^*,\boldsymbol{\alpha})\leq L(\mathbf{x}^*,\boldsymbol{\alpha}^*)\leq L(\mathbf{x}\ \boldsymbol{\alpha}^*)\ \forall(\mathbf{x}\ \boldsymbol{\alpha})\in\mathcal{M}$，则称其为拉格朗日函数 $L(\mathbf{x}\ \boldsymbol{\alpha})$ 的鞍点；根据引理 1，这也可以改写为以下变分不等式

$$G(\mathbf{x})-G(\mathbf{x}^*)+\langle\mathbf{u}-\mathbf{u}^*,\mathcal{K}_1(\mathbf{u}^*)\rangle\geq 0,\forall\mathbf{u}\in\mathcal{M}, \tag{17}$$

$$\Longleftrightarrow$$

$$\phi(\mathbf{x})-\phi(\mathbf{x}^*)+\langle\mathbf{u}-\mathbf{u}^*,\mathcal{K}_2(\mathbf{u}^*)\rangle\geq 0,\forall\mathbf{u}\in\mathcal{M}, \tag{18}$$

其中 $\phi(\mathbf{x})=G(\mathbf{x})+F(\mathbf{x})$，$\mathbf{u}=$列向量$\{\mathbf{x}\ \boldsymbol{\alpha}\}$，

$$\mathcal{K}_1(\mathbf{u})=\begin{pmatrix}\nabla F(\mathbf{x})-\mathbf{V}\boldsymbol{\alpha}\\ \mathbf{V}\mathbf{x}\end{pmatrix},\ \text{and}\ \mathcal{K}_2(\mathbf{u})=\begin{pmatrix}-\mathbf{V}\boldsymbol{\alpha}\\ \mathbf{V}\mathbf{x}\end{pmatrix}.$$

注意，映射 $\mathcal{K}_2$ 是一个具有斜对称矩阵的仿射映射，因此有

$$\langle\mathbf{u}_1-\mathbf{u}_2,\mathcal{K}_1(\mathbf{u}_1)-\mathcal{K}_1(\mathbf{u}_2)\rangle\equiv 0,\forall\mathbf{u}_1,\mathbf{u}_2\in\mathcal{M}. \tag{19}$$

由上述分析可知，集合 $\mathcal{M}^*$ 也是 (17) 和 (18) 的解集。因此，如果生成的序列能够表示为与 (17) 或 (18) 类似的不等式，并且多出一项趋于零的项，就可以证明 D-iPGM 的收敛性。基于这一观察，我们将在第III-A、III-B 和 III-C 节中展示 D-iPGM 生成序列的收敛性质。

A. 一般凸性下的收敛性分析

为简化收敛性分析中的记号，我们首先定义如下两个矩阵

$$\mathbf{Q}=\begin{pmatrix}\Gamma^{-1}&\mathbf{V}\\ \mathbf{0}&\frac{1}{\beta}\mathbf{I}\end{pmatrix},\ \mathbf{M}=\begin{pmatrix}\mathbf{I}&\Gamma\mathbf{V}\\ \mathbf{0}&\mathbf{I}\end{pmatrix}.$$

接下来，利用矩阵 $\mathbf{Q}$ 和$\mathbf{M}$，我们定义如下两个矩阵，它们对于建立 D-iPGM 的收敛性和收敛速率至关重要。

$$\mathbf{H}:=\mathbf{Q}\mathbf{M}^{-1},\ \mathbf{G}:=\mathbf{Q}^{\mathsf{T}}+\mathbf{Q}-\mathbf{M}^{\mathsf{T}}\mathbf{H}\mathbf{M}. \tag{20}$$

命题 1。如果 $\tau_i>0$、$i\in\mathcal{V}$ 和 $0<\beta\leq 0.5/$最大值$_i\,\tau_i$，则矩阵$\mathbf{H}$ 和$\mathbf{G}$ 是正定的。

证明：根据$\mathbf{Q}$ 和$\mathbf{M}$ 的定义可知

$$\mathbf{H}=\begin{pmatrix}\Gamma^{-1}&\mathbf{V}\\ \mathbf{0}&\frac{1}{\beta}\mathbf{I}\end{pmatrix}\begin{pmatrix}\mathbf{I}&-\Gamma\mathbf{V}\\ \mathbf{0}&\mathbf{I}\end{pmatrix}=\begin{pmatrix}\Gamma^{-1}&\mathbf{0}\\ \mathbf{0}&\frac{1}{\beta}\mathbf{I}\end{pmatrix}.$$

此外，有

$$\begin{aligned}\mathbf{G}&=\mathbf{Q}^{\mathsf{T}}+\mathbf{Q}-\mathbf{M}^{\mathsf{T}}\mathbf{H}\mathbf{M}=\mathbf{Q}^{\mathsf{T}}+\mathbf{Q}-\mathbf{M}^{\mathsf{T}}\mathbf{Q}\\ &=\begin{pmatrix}2\Gamma^{-1}&\mathbf{V}\\ \mathbf{V}&\frac{2}{\beta}\mathbf{I}\end{pmatrix}-\begin{pmatrix}\mathbf{I}&\mathbf{0}\\ \mathbf{V}\Gamma&\mathbf{I}\end{pmatrix}\begin{pmatrix}\Gamma^{-1}&\mathbf{V}\\ \mathbf{0}&\frac{1}{\beta}\mathbf{I}\end{pmatrix}\\ &=\begin{pmatrix}2\Gamma^{-1}&\mathbf{V}\\ \mathbf{V}&\frac{2}{\beta}\mathbf{I}\end{pmatrix}-\begin{pmatrix}\Gamma^{-1}&\mathbf{V}\\ \mathbf{V}&\frac{1}{\beta}\mathbf{I}+\mathbf{V}\Gamma\mathbf{V}\end{pmatrix}\\ &=\begin{pmatrix}\Gamma^{-1}&\mathbf{0}\\ \mathbf{0}&\frac{1}{\beta}\mathbf{I}-\mathbf{V}\Gamma\mathbf{V}\end{pmatrix}.\end{aligned}$$

由于$\mathbf{V}=\sqrt{\mathbf{I}-\mathbf{W}}$且 $\sigma_M(\mathbf{I}-\mathbf{W})\in[0,2)$，$0<\beta\leq 0.5/\max_i\tau_i$意味着 $\frac{1}{\beta}\mathbf{I}-\mathbf{V}\Gamma\mathbf{V}\succ\mathbf{0}$。命题得证。■

给定矩阵$\mathbf{G}$，定义另外两个矩阵如下：

$$\widehat{\mathbf{H}}_1=\mathbf{G}-\frac{1}{2}\mathrm{diag}\{\mathbf{L}_F,\mathbf{0}\},\ \text{and}\ \widehat{\mathbf{H}}_2=\mathbf{G}-\mathrm{diag}\{\mathbf{L}_F,\mathbf{0}\}.$$

$\widehat{\mathbf{H}}_1$ 的正定性对于建立 D-$\widehat{\text{iPGM}}$ 的收敛性和非遍历收敛率是必要的，而$\mathbf{H}_2$ 的正定性则用于建立 D-iPGM 的遍历收敛率。如果$\mathbf{H}_1\succ 0$，则对于任意$\mathbf{u}_1$,$\mathbf{u}_2\in\mathcal{M}$

$$\begin{aligned}c_2\|\mathbf{u}_1-\mathbf{u}_2\|&\leq\|\mathbf{u}_1-\mathbf{u}_2\|_{\widehat{\mathbf{H}}_1}\\ &\leq\|\mathbf{u}_1-\mathbf{u}_2\|_{\mathbf{H}}\leq c_1\|\mathbf{u}_1-\mathbf{u}_2\|,\end{aligned}$$

其中 $c_1=\sigma_M^{1/2}(\mathbf{H})$ 和 $c_2=\sigma_m^{1/2}(\widehat{\mathbf{H}}_1)$。令 $\tilde{\mathbf{u}}^k=$列向量$\{\tilde{\mathbf{x}}^k,\ \tilde{\boldsymbol{\alpha}}^k\}$可将算法 1 改写为

$$\tilde{\mathbf{x}}^k=\mathrm{prox}_G^{\Gamma^{-1}}(\mathbf{x}^k-\Gamma(\nabla F(\mathbf{x}^k)-\mathbf{V}\boldsymbol{\alpha}^k-\mathbf{d}^k)), \tag{21a}$$

$$\tilde{\boldsymbol{\alpha}}^k=\boldsymbol{\alpha}^k-\beta\mathbf{V}\tilde{\mathbf{x}}^k, \tag{21b}$$

$$\mathbf{u}^{k+1}=\mathbf{u}^k-\mathbf{M}(\mathbf{u}^k-\tilde{\mathbf{u}}^k), \tag{21c}$$

其中 $\mathbf{d}^k=$列向量$\{\mathrm{d}_1^k,\ \cdots,\ \mathrm{d}_N^k\}$。注意，若 $\boldsymbol{\alpha}^0=\boldsymbol{\lambda}^0=\mathbf{0}$，序列 $\{\mathbf{x}^k\}$ 由算法 1 和递推式（21）生成，则二者相同。在分析中，我们考虑序列$\{(\mathbf{x}^k,\ \boldsymbol{\alpha}^k)\}$，以说明算法 1的收敛性。

为证明 D-iPGM 的全局收敛性并推导其收敛速率，我们首先给出以下引理。

引理2。假设 *1* 和 *2* 成立。对于任意$\forall(\mathbf{x}\ \boldsymbol{\alpha})\in\mathcal{M}$，由 $\{\tilde{\mathbf{u}}^k\}$ 生成的序列满足(21)：

$$\begin{aligned}G(\mathbf{x})-G(\tilde{\mathbf{x}}^k)+\langle\mathbf{u}-\tilde{\mathbf{u}}^k,\mathcal{K}_1(\mathbf{u})\rangle&\geq\langle\mathbf{u}-\tilde{\mathbf{u}}^k,\mathbf{Q}(\mathbf{u}^k-\tilde{\mathbf{u}}^k)\rangle\\ &+\langle\mathbf{x}-\tilde{\mathbf{x}}^k,\mathbf{d}^k\rangle-\frac{1}{4}\|\mathbf{x}^k-\tilde{\mathbf{x}}^k\|_{\mathbf{L}_F}^2,\end{aligned} \tag{22}$$

and

$$\begin{aligned}\phi(\mathbf{x})-\phi(\tilde{\mathbf{x}}^k)+\langle\mathbf{u}-\tilde{\mathbf{u}}^k,\mathcal{K}_2(\mathbf{u})\rangle&\geq\langle\mathbf{u}-\tilde{\mathbf{u}}^k,\mathbf{Q}(\mathbf{u}^k-\tilde{\mathbf{u}}^k)\rangle\\ &+\langle\mathbf{x}-\tilde{\mathbf{x}}^k,\mathbf{d}^k\rangle-\frac{1}{2}\|\mathbf{x}^k-\tilde{\mathbf{x}}^k\|_{\mathbf{L}_F}^2.\end{aligned} \tag{23}$$

Proof: See Appendix A. ■

为确保$\mathbf{H}_1$, 的正定性，我们假设步长满足 $0 < \tau_i < 2/L_i,\ i \in \mathcal{V},\ 0 < \beta \leq 0.5/$最大值$_i\,\tau_i$。然后，我们给出以下定理，以证明由（21）生成的序列 $\{(\mathbf{x}^k,\ \boldsymbol{\alpha}^k)\}$ 收敛于问题（16）的原始-对偶最优解。

定理 1. 假设假设条件 *1* 和 *2* 成立。若 $0 < \tau_i < 2/L_i,\ i \in \mathcal{V}$, $0 < \beta \leq 0.5/\max_i \tau_i$，且 $\sum_{k=1}^{\infty} \varepsilon_k < \infty$，则由 $\{\mathbf{u}^k\}$生成的序列 (21) 满足:

1) 对任意 $\mathbf{u}^* \in \mathcal{M}^*$，有

$$\|\mathbf{u}^{k+1} - \mathbf{u}^*\|_{\mathbf{H}}^2 \leq \|\mathbf{u}^k - \mathbf{u}^*\|_{\mathbf{H}}^2 - \|\mathbf{u}^k - \tilde{\mathbf{u}}^k\|_{\widehat{\mathbf{H}}_1}^2 + 2\langle \tilde{\mathbf{x}}^k - \mathbf{x}^*, \mathbf{d}^k\rangle; \tag{24}$$

2) 序列 $\{\mathbf{u}^k\}$ 有界;

3) 存在 $\mathbf{u}^\infty \in \mathcal{M}^*$ 使得 $\lim_{k\to\infty}\mathbf{u}^k = \mathbf{u}^\infty$，即$\mathbf{x}^\infty = \mathbf{1}_N \otimes \mathbf{x}^\infty$, 且 $\mathbf{x}^\infty$ 是问题 (1)的解。

证明： 参见附录B。 ■

尽管（24）未必能保证序列 $\{\|\mathbf{u}^k - \mathbf{u}^\infty\|_{\mathbf{H}}^2\}$的单调性，但由于 $\{\mathbf{u}^k\}$ 有界且 $\|\mathbf{d}^k\|$可求和，序列 $\|\mathbf{u}^k - \mathbf{u}^\infty\|_{\mathbf{H}}^2$ 是一个准 Fej´er 单调序列（在 **H**-范数下），即

$$\|\mathbf{u}^{k+1} - \mathbf{u}^\infty\|_{\mathbf{H}} \leq \|\mathbf{u}^k - \mathbf{u}^\infty\|_{\mathbf{H}} + \bar{\mu}\varepsilon_k,$$

这保证了 Di-PGM 的收敛性。

当子问题（8）被精确求解，或者 g_i 的近端算子具有解析解时，我们有 $\|\mathbf{d}^k\| = 0$。在这种情况下，由（21）生成的序列 $\{\mathbf{u}^k\}$满足

$$\|\mathbf{u}^{k+1} - \mathbf{u}^*\|_{\mathbf{H}}^2 \leq \|\mathbf{u}^k - \mathbf{u}^*\|_{\mathbf{H}}^2 - \|\mathbf{u}^k - \tilde{\mathbf{u}}^k\|_{\widehat{\mathbf{H}}_1}^2, \forall \mathbf{u}^* \in \mathcal{M}^*,$$

这意味着 $\{\mathbf{u}^k\}$ 是关于 $\mathcal{M}^*$ 在 **H**-范数下的 Fej´er 单调序列。

B. 一般凸性下的次线性收敛率

在一般凸性条件下，本小节确立了 D-iPGM 的次线性收敛率。

我们给出以下定理，以展示 D-iPGM 的 $O(1/k)$ 非遍历和遍历收敛率。

定理 2. 假设 *1* 和 *2* 成立。若 $0 < \tau_i < 2/L_i,\ i \in \mathcal{V}$, $0 < \beta \leq 0.5/$最大值$_i\,\tau_i$，且 $\sum_{k=1}^{\infty} \varepsilon_k < \infty$，则有

1) 运行平均一阶最优性残差:

$$\frac{1}{K+1}\sum_{k=0}^{K} \mathrm{dist}^2(\mathbf{0}, \mathcal{J}(\tilde{\mathbf{u}}^k)) = O\left(\frac{1}{K+1}\right).$$

2) 运行最优一阶最优性残差:

$$\min_{0\leq k\leq K}\{\mathrm{dist}^2(\mathbf{0}, \mathcal{J}(\tilde{\mathbf{u}}^k))\} = o\left(\frac{1}{K+1}\right).$$

令 $\mathbf{X}^K = \frac{1}{K}\sum_{k=0}^{K-1} \tilde{}\mathbf{x}^k$ 且 $\mathbf{\Lambda}^K = \frac{1}{K}\sum_{k=0}^{K-1} \tilde{\boldsymbol{\alpha}}^k$。此外，若 $0 < \tau_i < 1/L_i, i \in \mathcal{V}$，则有

1) Primal-dual gap: $L(\mathbf{X}^K, \boldsymbol{\alpha}) - L(\mathbf{x}, \mathbf{\Lambda}^K) = O\left(\frac{1}{K}\right)$.

2) Primal optimality gap: $|\phi(\mathbf{X}^K) - \phi(\mathbf{x}^*)| = O\left(\frac{1}{K}\right)$.

3) Consensus error: $\|\sqrt{\mathbf{I} - \mathbf{W}}\mathbf{X}^K\| = O\left(\frac{1}{K}\right)$.

证明： 见附录C。 ■

在步长条件 $\tau_i < 2/L_i,\ i \in \mathcal{V}$ 下，我们建立了 D-iPGM 的非遍历迭代复杂度。在更强的收敛条件 $\tau_i < 1/L_i,\ i \in \mathcal{V}$ 下，证明了原始-对偶间隙、原始最优性间隙和共识误差的遍历 $O(1/k)$ 收敛速率。

下一个定理给出了当 $\varepsilon_k = 0, k \geq 0$ 时 D-iPGM 的 $o(1/k)$ 非遍历收敛率。

定理 3. 假设 *1* 和 *2* 成立。若 $0 < \tau_i < 2/L_i,\ i \in \mathcal{V}$, $0 < \beta \leq 0.5/$ 最大值$_i\,\tau_i$，且 $\varepsilon_k = 0,\ \forall k \geq 0$，对任意整数 $k \geq 0$，则有

$$\|\mathbf{M}(\mathbf{u}^{k+1} - \tilde{\mathbf{u}}^{k+1})\|_{\mathbf{H}}^2 \leq \|\mathbf{M}(\mathbf{u}^k - \tilde{\mathbf{u}}^k)\|_{\mathbf{H}}^2. \tag{25}$$

此外，对于任意$\mathbf{u}^* \in \mathcal{M}^*$，有

1) 连续差分:

$$\|\mathbf{u}^{k+1} - \mathbf{u}^k\|_{\mathbf{H}}^2 = o\left(\frac{1}{k+1}\right).$$

2) 一阶最优性残差:

$$\mathrm{dist}^2(\mathbf{0}, \mathcal{J}(\tilde{\mathbf{u}}^k)) = o\left(\frac{1}{k+1}\right).$$

证明： 参见附录 D。 ■

由（32）可得，即

$$G(\mathbf{x}) - G(\tilde{\mathbf{x}}^k) + \langle \mathbf{x} - \tilde{\mathbf{x}}^k, \nabla F(\mathbf{x}^k)\rangle + \langle \mathbf{u} - \tilde{\mathbf{u}}^k, \mathcal{K}_2(\tilde{\mathbf{u}}^k)\rangle \geq \langle \mathbf{u} - \tilde{\mathbf{u}}^k, \mathbf{Q}(\mathbf{u}^k - \tilde{\mathbf{u}}^k)\rangle + \langle \mathbf{x} - \tilde{\mathbf{x}}^k, \mathbf{d}^k\rangle, \forall \mathbf{u} \in \mathcal{M},$$

连续差分 $\|\mathbf{u}^{k+1} - \mathbf{u}^k\|_{\mathbf{H}}^2$ 可用于衡量迭代点 $\mathbf{u}^{k+1}$ 到 $\mathcal{M}^*$ 的精度，其中$\varepsilon_k = 0,\ k \geq 0$。更具体地，若 $\|\mathbf{u}^{k+1} - \mathbf{u}^k\|_{\mathbf{H}}^2 = 0$，由（21c）可得 $\mathbf{u}^k = \tilde{}\mathbf{u}^k = \mathbf{u}^{k+1}$。因此可推出

$$G(\mathbf{x}) - G(\tilde{\mathbf{x}}^k) + \langle \mathbf{u} - \tilde{\mathbf{u}}^k, \mathcal{K}_1(\tilde{\mathbf{u}}^k)\rangle \geq 0, \forall \mathbf{u} \in \mathcal{M}.$$

与（17）相比，$\mathbf{u}^{k+1} = \tilde{}\mathbf{u}^k \in \mathcal{M}^*$。因此，$\|\mathbf{u}^{k+1} - \mathbf{u}^k\|_{\mathbf{H}}^2$ 可以视为 D-iPGM 迭代 k 次后的误差度量。PG-EXTRA [20]和 NIDS 中也考虑了相同的误差度量 [21]。

C. 度量次正则性下的线性收敛率

借助一些变分分析技术，本小节证明了 D-iPGM 在度量次正则性下的线性收敛性。回顾度量次正则性的概念 [41], , 它对于建立局部线性收敛率 [42], [43] 至关重要。

定义 1。 集值映射 $\Psi : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ 在 $(\bar{u}, \bar{v}) \in \mathrm{gph}(\Psi)$ 处度量次正则，若存在某个 $\epsilon > 0$，使得存在 $\kappa \geq 0$，满足

$$\mathrm{dist}(u, \Psi^{-1}(\bar{v})) \leq \kappa\ \mathrm{dist}(\bar{v}, \Psi(u)),\ \forall u \in \mathcal{B}_\epsilon(\bar{u}),$$

其中 $\mathrm{gph}(\Psi) := \{(u, v) :\ v = \Phi(u)\}$, $\Psi^{-1}(v) := \{u \in \mathbb{R}^n : (u, v) \in \mathrm{gph}(\Psi)\}$。

为确保 D-iPGM 的线性收敛性，我们提出以下假设。

假设3。 存在一个整数 $k > 0$ 和一个序列$\{\varrho_k\}$，使得

$$\sup_{k\geq\bar{k}}\{\varrho_k\} < 1/\bar{\mu},\ \textit{and}\ \|\mathbf{d}^k\| \leq \varrho_k\|\mathbf{x}^k - \mathbf{x}^{k+1}\|, \forall k \geq \bar{k}, \tag{26}$$

其中 $\bar{\mu}=\sigma_M(\mathbf{H}^{\frac{1}{2}}\mathbf{M})\sigma_M(\Gamma)(\beta\sigma_M(\sqrt{\mathbf{I}-\mathbf{W}})+1)$。

注释 4 *1*, $\sum_{k=0}^{\infty}\|\mathbf{x}^{k+1}-\mathbf{x}^{k}\|<\infty$。由定理可得 $\mathbf{x}^{k}-\mathbf{x}^{k+1}=0$。这意味着极限 $\lim_{k\to\infty}(\|\mathbf{x}^{k-1}-\mathbf{x}^{k}\|-\|\mathbf{x}^{k}-\mathbf{x}^{k+1}\|)=0$。因此，为了便于实现，我们可以在实践中使用 $\|\mathbf{x}^{k}-\mathbf{x}^{k-1}\|$ 来代替 $\|\mathbf{x}^{k+1}-\mathbf{x}^{k}\|$。

随后给出了 D-iPGM 的线性收敛率。

定理 4. 假设假设 *1*、*2* 和 *3* 成立，$0<\tau_i<2/L_i, i\in\mathcal{V}$, $0<\beta\le 0.5/\max_i\tau_i$，且由 *(21)* 生成的序列$\{u^k\}$ 收敛到 u^∞。如果 *KKT* 映射$\mathcal{J}(u)$ 在 $(u^\infty, 0)$ 处度量次正则，则存在一个整数 $k>0$ 和一个常数 $\kappa>0$，使得对所有 $k\ge k$

$$\mathrm{dist}_{\mathbf{H}}(\mathbf{u}^{k+1},\mathcal{M}^*)\le\vartheta_k\ \mathrm{dist}_{\mathbf{H}}(\mathbf{u}^{k},\mathcal{M}^*),\tag{27}$$

with

$$\vartheta_k=\frac{\mu(1+\omega)\varrho_k+\bar{\vartheta}}{(1-\mu\varrho_k)},\ \mu=\bar{\mu}/c_2,\ \omega=\sqrt{c_1/c_2},$$
$$\bar{\vartheta}^2=1-\frac{c_2^2}{c_1^2(\kappa\kappa_2c_1+1)^2}<1,$$
$$\kappa_2^2=\frac{\max\{3\sigma_M(\mathbf{V}^2)+\frac{1}{\beta^2},3\max_i\{L_i^2\}+3\sigma_m(\Gamma^{-2})\}}{\sigma_m(\widehat{\mathbf{H}}_1)}.$$

此外，如果

$$\sup_{k\ge\bar{k}}\{\varrho_k\}<\frac{1-\bar{\vartheta}}{\mu(2+\omega)},\tag{28}$$

则有 $\vartheta_k<1$，因此收敛速率$\mathrm{dist}_{\mathbf{H}}(\mathbf{u}^k,\mathcal{M}^*)$ 是 *Q*-线性收敛。

证明： 见附录E。 ■

注意，如果 $\{\varrho_k\}$ 收敛到 0 当 $k\to\infty$时，条件(28)最终对 k 充分大时成立，因为 $\bar{\mu}$ 和 ϑ 均为常数。此外，如果 $\varrho_k=0,\ k\ge 0$，则有 $\vartheta_k\equiv\vartheta$，这意味着对所有 $k\ge k$

$$\mathrm{dist}^2_{\mathbf{H}}(\mathbf{u}^{k+1},\mathcal{M}^*)\le\left(1-\frac{c_2^2}{c_1^2(\kappa\kappa_2c_1+1)^2}\right)\mathrm{dist}^2_{\mathbf{H}}(\mathbf{u}^{k},\mathcal{M}^*).$$

因此，对于精确近端迭代，D-iPGM 的 Q-线性收敛成立。

注 5。 在 *[14]*, 中给出了问题 (1)与维度无关的下界，并指出，若每个智能体拥有不同的局部非光滑项，则最坏情况下无法实现精确线性收敛（f_i 是一个强凸光滑函数，而 g_i 是具有闭式近端映射的非光滑凸函数）。然而，在定理*4*中，可以建立 *D-iPGM* 依赖维度的线性收敛速率，这与已有结果并不矛盾。

接下来，利用 [42] 和 [43], 中的分析技术，我们给出使 $\mathcal{J}(\mathbf{u})$在$(\mathbf{u}^\infty,\mathbf{0})$处度量次正则的充分条件。与 [42] 和 [43], 类似，若凸函数 $F(\mathbf{x})$满足结构化假设，即 $F(\mathbf{x})=\sum_{i=i}^N f_i(\mathbf{x}_i)=\sum_{i=1}^N h_i(A_i\mathbf{x}_i)+\langle q_i,\mathbf{x}_i\rangle$，则称其满足结构化假设，其中$A_i$ 是一个 $m_i\times n$ 矩阵，q_i 是 $\mathbb{R}^n$ 中的向量，h_i 是光滑且本质局部强凸的，即对于任意紧凸子集 $\mathbb{K}$，h_i 在 $\mathbb{K}$上是强凸的。一些机器学习中常用的损失函数自动满足结构化假设。我们给出表格

表 III 满足结构化假设的一些常用凸损失函数。

	局部损失函数 h_i	f_i的凸性
线性回归	$\frac{1}{2}\|\mathbf{x}_i-\mathbf{b}_i^1\|^2$	凸
逻辑回归	$\sum_{j=1}^n\log(1+e^{\mathbf{x}_{ij}})-(\mathbf{b}_i^2)^\top\mathbf{x}_i$	凸
逻辑回归	$-\sum_{j=1}^n\log(\mathbf{x}_{ij})+(\mathbf{b}_i^3)^\top\mathbf{x}_i$	凸
似然估计	$\sum_{j=1}^n\log(1+e^{(\mathbf{b}_i^4)^\top\mathbf{x}^i})$	凸
泊松回归	$\sum_{j=1}^n e^{\mathbf{x}_{ij}}-(\mathbf{b}_i^5)^\top\mathbf{x}_i$	凸

III 以总结满足结构化假设的情形，其中 $\mathbf{b}_i^1\in\mathbb{R}^n$，$\mathbf{b}_i^2\in\{0,1\}^n$，$\mathbf{b}_i^3\in\mathbb{R}^n_+$，$\mathbf{b}_i^4\in\{-1,1\}^n$和$\mathbf{b}_i^5\in\{0,1,2,\cdots\}$。

根据 F的结构，给出以下命题，它提供了 $\mathcal{M}^*$的另一种刻画。

命题2。 假设 $F(\mathbf{x})$ 满足结构化假设。集合 $\mathcal{M}^*$ 可刻画为

$$\mathcal{M}^*=\{\mathbf{u}:\widetilde{\mathbf{A}}\mathbf{x}=\tilde{\mathbf{t}},\sqrt{\mathbf{I}-\mathbf{W}}\mathbf{x}=\mathbf{0},\ \mathbf{0}\in\partial G(\mathbf{x})+\tilde{\boldsymbol{\xi}}-\sqrt{\mathbf{I}-\mathbf{W}}\boldsymbol{\alpha}\},$$

其中某个$\tilde{\mathbf{t}}$ 满足 $\widetilde{\mathbf{A}}\mathbf{x}=\tilde{\mathbf{t}}$ 对所有$\mathbf{x}\in\mathcal{X}^*$成立，并且 $\tilde{\boldsymbol{\xi}}$ =列向量$\{A_1^\top\nabla h_1(\tilde{t}_1)+q_1,\cdots,A_N^\top\nabla h_N(\tilde{t}_N)+q_N\}$，其中 $\tilde{\mathbf{t}}$ =列向量$\{\tilde{t}_1,\cdots,\tilde{t}_N\}$，且 $\widetilde{\mathbf{A}}$是一个分块对角矩阵，其(i,i)块为 A_i，其余块为$\mathbf{0}$。

命题2的证明可由 [44,引理2直接得到，属于标准证明，1] 因此此处略去。为便于分析，引入一个带扰动的辅助集值映射，其扰动为$\mathrm{P}=(P_1,P_2,P_3)$，并与 $\mathcal{M}^*$相关： $\mathbf{\Gamma}(\mathrm{P})=\{\mathrm{u}:P_1=\widetilde{\mathrm{A}}\mathrm{x}-\tilde{\mathrm{t}}\ P_2=\mathrm{V}\mathrm{x}\ P_3\in\partial G(\mathrm{x})+\tilde{\boldsymbol{\xi}}-\mathrm{V}\boldsymbol{\alpha}\}$. 显然，$\Gamma(0)=\mathcal{M}^*$。类似于 [42, 命题 40] 和[43,命题4，1], 可得以下等价关系。

命题 3. 假设 F 满足结构化假设。 $\mathbf{\Gamma}^{-1}$ 和 $\mathcal{J}$的度量次正则性条件等价。具体而言，给定 $\bar{\mathbf{u}}\in\mathcal{M}^*$，以下两个陈述等价:

(i) 存在 $\tilde{\kappa}_1$、$\tilde{\epsilon}_1>0$，使得

$$\mathrm{dist}(\mathbf{u},\mathbf{\Gamma}(\mathbf{0}))\le\tilde{\kappa}_1\ \mathrm{dist}(\mathbf{0},\mathbf{\Gamma}^{-1}(\mathbf{u})),\forall\mathbf{u}\in\mathcal{B}_{\tilde{\epsilon}_1}(\bar{\mathbf{u}}).$$

(ii) 存在 $\tilde{\kappa}_2$、$\tilde{\epsilon}_2>0$，使得

$$\mathrm{dist}(\mathbf{u},\mathcal{J}^{-1}(\mathbf{0}))\le\tilde{\kappa}_2\ \mathrm{dist}(\mathbf{0},\mathcal{J}(\mathbf{u})),\forall\mathbf{u}\in\mathcal{B}_{\tilde{\epsilon}_2}(\bar{\mathbf{u}}).$$

假设 $F(\mathbf{x})$ 满足结构化假设，且∂G 是多面体集值映射，即其图是有限个多面体凸集的并。在这种情况下，$\mathbf{\Gamma}$ 是多面体集值映射，因此 $\mathbf{\Gamma}^{-1}$ 也是多面体集值映射。根据 [45,命题 1], , 对于任意点 $(\mathrm{u},\mathrm{P})\in\mathrm{gph}(\mathbf{\Gamma}^{-1})$，$\mathbf{\Gamma}^{-1}$ 在 (u,P) 处度量次正则。因此，根据命题 3，对于任意给定的 $\mathrm{u}^*\in\mathcal{M}^*$，$\mathcal{J}$在 $(\mathrm{u}^*,0)$ 处度量次正则。由 [43, 命题 7] 可知，若 g_i 是多面体凸函数（包括多面体集的指示函数和多面体凸正则项），或者 g_i是凸分段线性二次函数，则 ∂g_i 是多面体集值映射。这一情形涵盖了 g_i

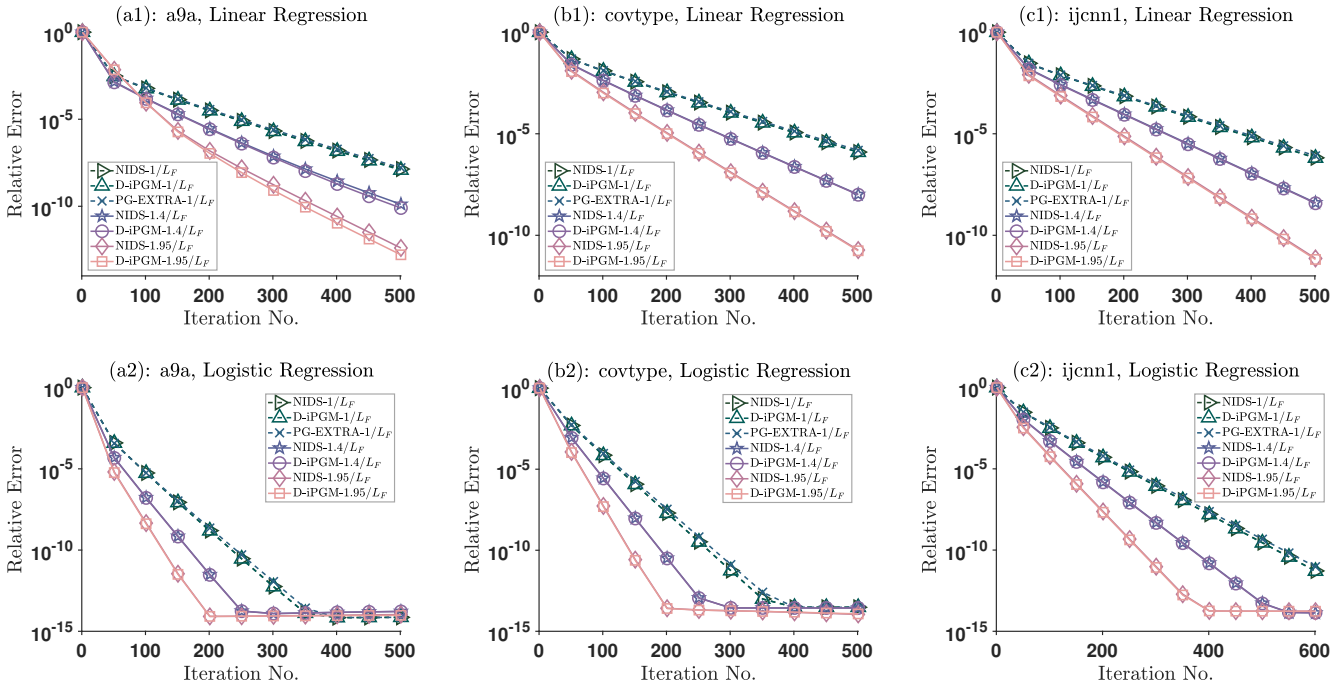


图 1. 当正则项为 $\|\mathbf{x}^k-\mathbf{x}^*\|/\|\mathbf{x}^*\|$ 时，相对于迭代次数的相对误差 $\nu_{1,i}\|\mathbf{x}\|+\frac{\nu_2}{2}{}_{,i}\|\mathbf{x}\|^2$，其中 $\nu_{1,i}=0.01$ 以及 $\nu_{2,i}=1$。这里考虑了 D-iPGM、PG-EXTRA 和 NIDS 的不同步长，并且网络中各智能体的步长相同。

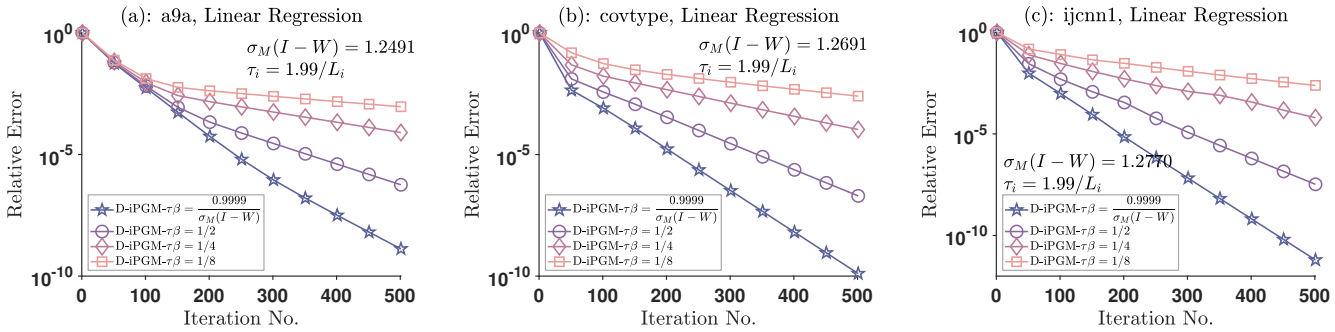


图 2. 相对于迭代次数的相对误差 $\|\mathbf{x}^k-\mathbf{x}^*\|/\|\mathbf{x}^*\|$ 当正则项为 $\nu_{1,i}\|\mathbf{x}\|_1+\frac{\nu_2}{2}{}_{,i}\|\mathbf{x}\|^2$ 时，其中 $\nu_{1,i}=0.01$ 以及 $\nu_{2,i}=1$。这里每个智能体都有自己的局部步长 τ_i，并令 $\tau\beta=0.9999/\sigma_M(\mathbf{I}-\mathbf{W})$、$1/2$、$1/4$ 和 $1/8$，其中 $\tau=\max_{i\in\mathcal{V}}\{\tau_i\}$

是 l_∞-范数正则项，l_1-范数正则项、弹性网络正则项 [46]，广义 LASSO 正则项 [23],以及融合 LASSO 正则项 [24]。

IV. 数值实验

本节针对去中心化线性回归和逻辑回归问题，在真实数据集上比较 D-iPGM 与 PG-EXTRA [20]（见（12）或（13））以及 NIDS [21]（见（14）或（15））在采用不同正则项时的性能。所有算法均在 Matlab R2020b 中实现，运行于一台配备 3.30 GHz AMD Ryzen 9 5900HS（带 Radeon Graphics）处理器和 16 GB 内存的计算机上。

对于所有实验，我们首先通过集中式方法计算方程（1）的解 $\mathbf{x}^*$，然后在一个随机生成的连通网络上运行，该网络包含 N 个智能体和 $\frac{\iota N}{2}{}_{(N-1)}$ 条无向边，其中 $N=50$ 和 $\iota=0.1$ 是连通率。混合矩阵 W 根据 Metropolis-Hastings 规则生成（参见 [8, 第 2.4] 节）。每次迭代的性能通过相对误差 $\|\mathbf{x}^k-\mathbf{x}^*\|/\|\mathbf{x}^*\|$进行评估。对于去中心化线性回归，损失函数为

$$\sum_{i=1}^{N}\Big\{\frac{1}{m_i}\sum_{j=1}^{m_i}\frac{1}{2}\|\mathcal{A}_{ij}^{\mathsf{T}}\mathbf{x}-\mathcal{B}_{ij}\|^2\Big\}.$$

对于去中心化逻辑回归，损失函数为

$$\sum_{i=1}^{N}\Big\{\frac{1}{m_i}\sum_{j=1}^{m_i}\ln(1+e^{-(\mathcal{A}_{ij}^{\mathsf{T}}\mathbf{x})\mathcal{B}_{ij}})\Big\}.$$

在此，任意智能体 i 都持有自己的训练数据（$\mathcal{A}_{ij}$，$\mathcal{B}_{ij}$）$\in\mathbb{R}^n\times\{-1,1\}$，$j=1,\cdots,m_i$，包括样本向量 $\mathcal{A}_{ij}$和相应的类别 $\mathcal{B}_{ij}$。我们使用了三个真实数据集，包括 a9a、covtype 和 ijcnn1[1]，，其属性分别为 $n=123$ 和$\sum_{i=1}^N m_i=32550$、$n=54$ 和 $\sum_{i=1}^N m_i=55500$，以及 $n=22$ 和 $\sum_{i=1}^N m_i=49950$。此外，

[1]https://www.csie.ntu.edu.tw/ cjlin/libsvmtools/datasets/

表 IV PG-EXTRA、NIDS 和 D-iPGM 求解广义套索去中心化线性回归的比较

算法	a9a：线性回归			covtype：线性回归			ijcnn1：线性回归		
	内迭代次数	外迭代次数	Time	内迭代次数	外迭代次数	Time	内迭代次数	外迭代次数	Time
PG-EXTRA	32687	146	14.1916	63843	186	6.8362	42901	139	1.7367
NIDS	31678	141	13.7203	61940	180	6.7148	41027	133	1.6229
D-iPGM-$\varepsilon_k = 10^{-10}$	31460	140	13.1901	61809	180	6.4375	40528	132	1.6263
D-iPGM-$\varepsilon_k = O(1/k^2)$	**18254**	140	**11.4739**	**35067**	185	**5.1945**	**27595**	132	**1.2733**
D-iPGM-$\varepsilon_k = O(\ln k/k^2)$	**16640**	146	**11.2607**	**33887**	189	**4.9323**	**24375**	132	**1.1896**
D-iPGM-$\varepsilon_k = O(\|\mathbf{x}_i^k - \mathbf{x}_i^{k-1}\|/k)$	**11836**	142	**10.6125**	**23520**	184	**4.4608**	**13388**	132	**0.8935**
D-iPGM-$\varepsilon_k = O(\|\mathbf{x}_i^k - \mathbf{x}_i^{k-1}\|/\ln k)$	**8568**	141	**9.9773**	**17539**	185	**3.9734**	**11529**	132	**0.8552**

表 V 使用 PG-EXTRA、NIDS 和 D-iPGM 求解去中心化逻辑回归与广义套索的比较

算法	a9a：逻辑回归			covtype：逻辑回归			ijcnn1：逻辑回归		
	内迭代次数	外迭代次数	Time	内迭代次数	外迭代次数	Time	内迭代次数	外迭代次数	Time
PG-EXTRA	32912	169	15.3620	52761	199	6.8032	54041	135	1.9368
NIDS	32910	169	15.3008	55084	208	7.0974	53794	130	1.8648
D-iPGM-$\varepsilon_k = 10^{-10}$	33794	169	15.7742	55281	209	7.0255	54050	130	1.9370
D-iPGM-$\varepsilon_k = O(1/k^2)$	**17701**	165	**13.6317**	**31920**	210	**5.8638**	**37284**	130	**1.4901**
D-iPGM-$\varepsilon_k = O(\ln k/k^2)$	**16055**	166	**13.0957**	**30864**	211	**5.6870**	**32520**	130	**1.4303**
D-iPGM-$\varepsilon_k = O(\|\mathbf{x}_i^k - \mathbf{x}_i^{k-1}\|/k)$	**13458**	166	**12.9017**	**19383**	160	**4.1614**	**17887**	123	**1.0219**
D-iPGM-$\varepsilon_k = O(\|\mathbf{x}_i^k - \mathbf{x}_i^{k-1}\|/\ln k)$	**8521**	156	**11.5524**	**18795**	173	**4.2117**	**13759**	117	**0.8875**

训练样本在所有 N 智能体之间随机且均匀地分配。

A. 场景 *1*：$\mathrm{prox}_{\tau_i g_i}(\cdot)$具有闭式解

我们首先考虑带有 $\nu_{1,i}\|\mathbf{x}\|_1 + \frac{\nu_{2,i}}{2}\|\mathbf{x}\|^2$ 正则项的去中心化线性回归和逻辑回归问题。对于这些问题，非光滑项为 $g_i(\mathbf{x}) = \nu_{1,i}\|x\|$，并且有 $\mathrm{prox}_{\tau_i g_i}(\mathbf{x}) = (1 - \frac{\tau_i\nu_{1,i}}{\max\{\|\mathbf{x}\|, \tau_i\nu_{1,i}\}})\mathbf{x}$。不难验证，这两个问题的光滑系数为 $L_i = \|\mathcal{A}_i^{\mathsf{T}}\mathcal{A}_i\| + 1$，其中 $\mathcal{A}_i = [\mathcal{A}_{i1}^{\mathsf{T}}; \cdots; \mathcal{A}_{im_i}^{\mathsf{T}}]$。在这种情况下，由于 g_i 的近端映射具有闭式解，我们采用精确版 D-iPGM，即 $\varepsilon_k = 0$，$k \geq 0$。因此，这三种算法的每次迭代复杂度相同。请注意，这些算法每次迭代只进行一轮通信。网络中的信息交换量与迭代次数成正比。因此，我们只记录迭代次数。

这些算法在每次迭代后的相对误差比较结果如图1所示，其中我们将 D-iPGM 的 $\tau\beta = 1$ 固定。对于 PG-EXTRA，我们仅展示 $\tau = 1/L_F$，因为当 $\tau = 1.4/L_F$ 且 $\tau = 1.95/L_F$时，PG-EXTRA 发散。当 $\tau = 1/L_F$时，这三种算法具有相似的收敛性能。当 $\tau = 1.4/L_F$ 且 $\tau = 1.95/L_F$ 时，D-iPGM 和 NIDS 几乎以相同的速度收敛。此外，步长越大，D-iPGM 和 NIDS 的收敛速度越快。

然后，我们考虑带有 $\nu_{1,i}\|\mathbf{x}\|_1 + \frac{\nu_{2,i}}{2}\|\mathbf{x}\|^2$ 正则项的分散式线性回归问题。因此，非光滑项为 $g_i(\mathbf{x}) = \nu_{1,i}\|\mathbf{x}\|_1$，它也有闭式解。我们采用非协调步长 $\tau_i = 1.99/L_i$ 并令 $\tau\beta = 0.9999/\sigma_M(\mathbf{I} - \mathbf{W})$，$1/2$，$1/4$，以及 $1/8$，其中τ = 最大值$_{i\in\mathcal{V}}\{\tau_i\}$。每次迭代后的相对误差如图 2所示。结果表明，D-iPGM 在较大的 $\tau\beta$下收敛更快。

B. 场景 *2*：$\mathrm{prox}_{\tau_i g_i}(\cdot)$没有闭式解

在本小节中，我们考虑带有 l_2+广义套索 [23]正则项的去中心化线性回归和逻辑回归问题，即 $\frac{\nu_{2,i}}{2}\|\mathbf{x}\|^2 + \nu_{1,i}\|\mathbf{D}_i\mathbf{x}\|_1$，其中 $\mathbf{D}_i \in \mathbb{R}^{10\times n}$取自正态分布 $N(\mathbf{0}, \mathbf{1})$，$\nu_{1,i} = 0.01$，且$\nu_{2,i} = 1$。在该场景中，非光滑项 $g_i(\mathbf{x}) = \|\mathbf{D}_i\mathbf{x}\|_1$ 的近端映射对于一般的 $\mathbf{D}_i$没有闭式解，因此我们将使用 C-V 算法 [36], [37]来计算 $\mathrm{prox}_{\tau_i g_i}(\cdot)$ 的数值解。

回顾 D-iPGM 的原始变量更新，

$$\tilde{\mathbf{x}}_i^k \approx \tilde{\mathbf{y}}_i^k := \mathrm{prox}_{\tau_i g_i}(\mathbf{x}_i^k - \tau_i(\nabla f_i(\mathbf{x}_i^k) - \lambda_i^k)).$$

令 $s_i^k = \mathbf{x}_i^k - \tau_i(\nabla f_i(\mathbf{x}_i^k) - \lambda_i^k)$。根据 g_i 的近端映射定义，有

$$\tilde{\mathbf{x}}_i^k \approx \tilde{\mathbf{y}}_i^k = \arg\min_{\mathbf{x}\in\mathbb{R}^n}\{\tau_i\nu_{1,i}\|\mathbf{D}_i\mathbf{x}\|_1 + \frac{1}{2}\|\mathbf{x} - \mathbf{s}_i^k\|^2\}. \quad (29)$$

将 C-V 算法应用于问题 (29)，得到

$$\begin{aligned}
\mathbf{x}_i^{k,l+1} &= \mathbf{x}_i^{k,l} - t_{1,i}(\mathbf{x}_i^{k,l} - s_i^k + \mathbf{D}_i^{\mathsf{T}}\mathbf{w}_i^l),\\
\mathbf{v}_i^{l+1} &= \mathbf{w}_i^l + t_{2,i}\mathbf{D}_i(2\mathbf{x}_i^{k,l+1} - \mathbf{x}_i^{k,l}),\\
\mathbf{w}_i^{l+1} &= \mathbf{v}_i^{l+1} - t_{2,i}\mathrm{prox}_{\frac{\tau_i\nu_{1,i}}{t_{2,i}}\|\cdot\|_1}(\mathbf{v}_i^{l+1}/t_{2,i}),
\end{aligned}$$

其中 $\mathrm{x}_i^{k\,0}=$、x_i^k、$\mathrm{w}_i^0=\mathbf{0}$，以及 t_1、$_i$ 和 t_2、$_i$ 是满足以下条件的步长：

$$\frac{t_1{}_i}{2} + t_1{}_i t_2{}_i\| \quad {}_i^{\mathrm{T}} \quad {}_i\| < 1 \qquad {}_i^{k\,l+1} \in 、 、 、$$

$$\partial g_i \quad {}_i^{k\,l+1} + \nabla f_i \quad {}_i^k - \lambda_i^k + \frac{1}{\tau_i} {}_i^{k\,l+1} - \quad {}_i^k$$

D D。令 d , (x ,) (x) (x , x)。根据 [33, 第 7] 节和 [47, 引理 3]，存在常数 ε_0，使得 $\|\mathrm{d}_i^{k\,l+1}\| \le \varepsilon_0\|, \mathrm{x}_i^{k\,l+1} -, \mathrm{x}_i^{k\,l}\|,$。因此，若 $\|\mathrm{x}$ ${}_i^{k\,l+1} -, \mathrm{x}_i^{k\,l}\| < \varepsilon_k/\varepsilon_0,,$ 则停止内层过程，并令 $\tilde{}\mathrm{x}_i^k = \mathrm{x}_i^{k\,l+1},,$ 即 D-iPGM 的内层停止准则可选为

$$\text{[D-iPGM]: } \|\mathrm{x}_i^{k,l+1} - \mathrm{x}_i^{k,l}\| \le \varepsilon_k.$$

在这一场景中，我们采用 PG-EXTRA（12）和 NIDS（14）的原始-对偶形式来求解所考虑的问题。需要注意的是，这三种算法的原始变量更新相同，而 PG-EXTRA 和 NIDS 是精确迭代算法。我们采用 C-V 算法，结合内层停止准则来计算（12a）和（14a）

$$\text{[PG-EXTRA/NIDS]: } \|\mathrm{x}_i^{k,l+1} - \mathrm{x}_i^{k,l}\| \le 10^{-10}.$$

为了比较非精确迭代的影响，并确保所有所考虑算法的外层迭代性能相近，根据场景 1 的实验结果，我们为外层迭代设置 $\tau = 1/L_F, \tau\beta = 1/2$。此外，我们对所有所考虑的方法采用停止准则 $\|\mathbf{x}^k - \mathbf{x}^*\|/\|\mathbf{x}^*\| \le 10^{-5}$。

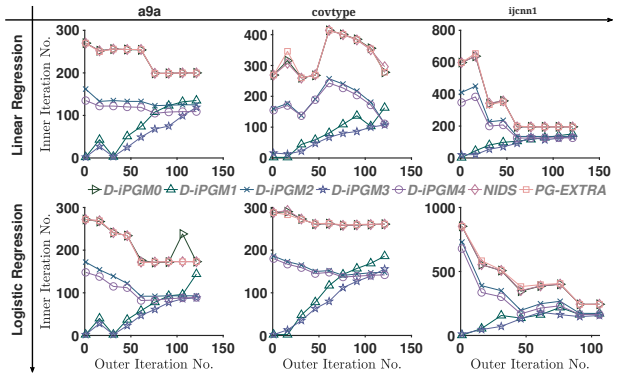


图 3. 内层迭代次数与外层迭代次数的关系。D-iPGM0、D-iPGM1、D-iPGM2、D-iPGM3 和 $\varepsilon_k =$ D-iPGM4 分别表示具有 10^{-10} O O O k O ,(‖x ${}_i^k - \mathrm{x}_i^{k-1}\|/k)$ $(1/k^2)$ (‖$\mathrm{x}_i^k - \mathrm{x}_i^{k-1}\|/\ln$) (ln k/k^2) 的 D-iPGM。

注意，不同算法每次迭代的复杂度各不相同。在表 IV和V中，我们报告了各方法达到外层迭代停止准则时所需的内迭代次数（内迭代次数）和外迭代次数（外层迭代），以及计算时间（秒）（时间）。从表IV 和V可以看出，D-iPGM 的计算时间和内迭代次数显著少于包括其精确版本在内的精确算法，而 D-iPGM 的外迭代次数与精确算法非常接近。因此，显然，就计算时间而言，D-iPGM 的表现优于 D-iPGM 的精确版本、PG-EXTRA 和 NIDS。此外，通过比较采用不同内层停止准则的 D-iPGM 的表现，结果表明，最好选择 $\varepsilon_k = O(\|\mathrm{x}_i^k - \mathrm{x}_i^{k-1}\|/\ln k)$ 作为内层停止准则。为了进一步直观展示数值结果，我们给出图 3 以展示每次迭代所需的内迭代次数，这表明 D-iPGM 每次迭代所需的内迭代次数少于精确算法。如果选择 $\varepsilon_k = O(\|\mathrm{x}_i^k - \mathrm{x}_i^{k-1}\|/\ln k)$ 或 $\varepsilon_k = O(\|\mathrm{x}_i^k - \mathrm{x}_i^{k-1}\|/k)$ 作为内层停止准则，每次迭代所需的内迭代次数大致呈上升趋势。如果选择$\varepsilon_k = O(1/k^2)$ 或 $\varepsilon_k = O(\ln k/k^2)$，每次迭代所需内部迭代次数的变化趋势与精确算法相似。

V. 结论

本文提出了一种基于 CTA 的去中心化算法 D-iPGM，用于网络上的凸复合优化，并采用非协调的网络无关常数步长。D-iPGM 在每次迭代中不要求精确求解 g_i 的近端映射，因此适用于近端映射可能难以计算的去中心化复合优化。我们证明了在一般凸性条件下，D-iPGM 的收敛速率为 $O(1/k)$；若近端映射得到精确求解，则可将其改进为 $o(1/k)$。在度量次正则性条件下，我们进一步证明了其具有线性收敛率。数值实验验证了 D-iPGM 的有效性。D-iPGM 的加速、异步和随机设置留待未来研究。

附录 A 引理证明引理 2

证明：由 (21a) 可得 $G(\mathbf{x}) - G(\tilde{}\mathbf{x}^k) + \langle \mathbf{x} - \tilde{}\mathbf{x}^k, \nabla F(\mathbf{x}^k) - \mathbf{V}\boldsymbol{\alpha}^k + \Gamma^{-1}(\tilde{}\mathbf{x}^k - \mathbf{x}^k) - \mathbf{d}^k\rangle \ge 0, \forall \mathbf{x} \in \mathbb{R}^{Nn}$，这意味着对于 $\forall \mathbf{x} \in \mathbb{R}^{Nn}$，

$$G(\mathbf{x}) - G(\tilde{\mathbf{x}}^k) + \langle \mathbf{x} - \tilde{\mathbf{x}}^k, \nabla F(\mathbf{x}^k) - \mathbf{V}\tilde{\boldsymbol{\alpha}}^k + \Gamma^{-1}(\tilde{\mathbf{x}}^k - \mathbf{x}^k) + \mathbf{V}(\tilde{\boldsymbol{\alpha}}^k - \boldsymbol{\alpha}^k) - \mathbf{d}^k\rangle \ge 0. \tag{30}$$

由 (21b) 可知，对于 $\forall \boldsymbol{\alpha} \in \mathbb{R}^{Nn}$

$$\langle \boldsymbol{\alpha} - \tilde{\boldsymbol{\alpha}}^k, \mathbf{V}\tilde{\mathbf{x}}^k\rangle = \langle \boldsymbol{\alpha} - \tilde{\boldsymbol{\alpha}}^k, \frac{1}{\beta}(\boldsymbol{\alpha}^k - \tilde{\boldsymbol{\alpha}}^k)\rangle. \tag{31}$$

结合 (30) 和 (31)，可得

$$G(\mathbf{x}) - G(\tilde{\mathbf{x}}^k) + \langle \mathbf{x} - \tilde{\mathbf{x}}^k, \nabla F(\mathbf{x}^k)\rangle + \langle \mathbf{u} - \tilde{\mathbf{u}}^k, \mathcal{K}_2(\tilde{\mathbf{u}}^k)\rangle \ge \langle \mathbf{u} - \tilde{\mathbf{u}}^k, \mathbf{Q}(\mathbf{u}^k - \tilde{\mathbf{u}}^k)\rangle + \langle \mathbf{x} - \tilde{\mathbf{x}}^k, \mathbf{d}^k\rangle, \forall \mathbf{u} \in \mathcal{M}. \tag{32}$$

利用 (19)，(32) 可改写为

$$G(\mathbf{x}) - G(\tilde{\mathbf{x}}^k) + \langle \mathbf{x} - \tilde{\mathbf{x}}^k, \nabla F(\mathbf{x}^k)\rangle + \langle \mathbf{u} - \tilde{\mathbf{u}}^k, \mathcal{K}_2(\mathbf{u})\rangle \ge \langle \mathbf{u} - \tilde{\mathbf{u}}^k, \mathbf{Q}(\mathbf{u}^k - \tilde{\mathbf{u}}^k)\rangle + \langle \mathbf{x} - \tilde{\mathbf{x}}^k, \mathbf{d}^k\rangle, \forall \mathbf{u} \in \mathcal{M}. \tag{33}$$

注意到

$$\langle \mathbf{u} - \tilde{\mathbf{u}}^k, \mathcal{K}_1(\mathbf{u})\rangle = \langle \mathbf{x} - \tilde{\mathbf{x}}^k, \nabla F(\mathbf{x})\rangle + \langle \mathbf{u} - \tilde{\mathbf{u}}^k, \mathcal{K}_2(\mathbf{u})\rangle.$$

重新整理（33）中的一些项，可得对于 $\forall \mathbf{u} \in \mathcal{M}$

$$G(\mathbf{x}) - G(\tilde{\mathbf{x}}^k) + \langle \mathbf{u} - \tilde{\mathbf{u}}^k, \mathcal{K}_1(\mathbf{u})\rangle \ge -\langle \mathbf{x} - \tilde{\mathbf{x}}^k, \nabla F(\mathbf{x}^k) - \nabla F(\mathbf{x})\rangle + \langle \mathbf{u} - \tilde{\mathbf{u}}^k, \mathbf{Q}(\mathbf{u}^k - \tilde{\mathbf{u}}^k)\rangle + \langle \mathbf{x} - \tilde{\mathbf{x}}^k, \mathbf{d}^k\rangle.$$

由（5），断言（22）成立。结合（4）和（33），断言（23）成立。 ■

附录 B 定理1的证明

证明：1）根据（21c）和 $\mathbf{Q}=\mathbf{HM}$，可得

$$\langle \mathbf{u}-\tilde{\mathbf{u}}^k, \mathbf{Q}(\mathbf{u}^k-\tilde{\mathbf{u}}^k)\rangle = \langle \mathbf{u}-\tilde{\mathbf{u}}^k, \mathbf{H}(\mathbf{u}^k-\mathbf{u}^{k+1})\rangle.$$

将恒等式 $\langle a-b,\ c-d\rangle = \frac{1}{2}\{\|a-d\|_{\mathbf{H}}^2 - \|a-c\|_{\mathbf{H}}^2 + \|b-c\|_{\mathbf{H}}^2 - \|b-d\|_{\mathbf{H}}^2\}$ 应用于具有指定条件 $a=\tilde{\mathbf{u}}^k,\ b=\mathbf{u},\ c=\mathbf{u}^{k+1}$ and $d=\mathbf{u}^k$ 的 uuH(uu)，可得

$$\begin{aligned}\langle \tilde{\mathbf{u}}^k-\mathbf{u}, \mathbf{H}(\mathbf{u}^{k+1}-\mathbf{u}^k)\rangle &= \frac{1}{2}\{\|\tilde{\mathbf{u}}^k-\mathbf{u}^k\|_{\mathbf{H}}^2 \\ &- \|\mathbf{u}^{k+1}-\tilde{\mathbf{u}}^k\|_{\mathbf{H}}^2 + \|\mathbf{u}^{k+1}-\mathbf{u}\|_{\mathbf{H}}^2 - \|\mathbf{u}^k-\mathbf{u}\|_{\mathbf{H}}^2\}. \quad (34)\end{aligned}$$

注意到 $\mathbf{Q}=\mathbf{HM}$ 以及 $\mathbf{G}=\mathbf{Q}^{\mathsf{T}}+\mathbf{Q}-\mathbf{M}^{\mathsf{T}}\mathbf{HM}$，可得

$$\begin{aligned}&\|\mathbf{u}^k-\tilde{\mathbf{u}}^k\|_{\mathbf{H}}^2 - \|\mathbf{u}^{k+1}-\tilde{\mathbf{u}}^k\|_{\mathbf{H}}^2 \\ &= \|\mathbf{u}^k-\tilde{\mathbf{u}}^k\|_{\mathbf{H}}^2 - \|(\mathbf{u}^k-\tilde{\mathbf{u}}^k)-\mathbf{M}(\mathbf{u}^k-\tilde{\mathbf{u}}^k)\|_{\mathbf{H}}^2 \\ &= 2\langle \mathbf{u}^k-\tilde{\mathbf{u}}^k, \mathbf{HM}(\mathbf{u}^k-\tilde{\mathbf{u}}^k)\rangle \\ &\quad - \langle \mathbf{u}^k-\tilde{\mathbf{u}}^k, \mathbf{M}^{\mathsf{T}}\mathbf{HM}(\mathbf{u}^k-\tilde{\mathbf{u}}^k)\rangle \\ &= \langle \mathbf{u}^k-\tilde{\mathbf{u}}^k, (\mathbf{Q}^{\mathsf{T}}+\mathbf{Q}-\mathbf{M}^{\mathsf{T}}\mathbf{HM})(\mathbf{u}^k-\tilde{\mathbf{u}}^k)\rangle \\ &= \|\mathbf{u}^k-\tilde{\mathbf{u}}^k\|_{\mathbf{G}}^2. \quad (35)\end{aligned}$$

因此，将（34）和（35）代入（22），可得

$$\begin{aligned}&G(\mathbf{x})-G(\tilde{\mathbf{x}}^k)+\langle \mathbf{u}-\tilde{\mathbf{u}}^k, \mathcal{K}_1(\mathbf{u})\rangle \geq \langle \mathbf{x}-\tilde{\mathbf{x}}^k, \mathbf{d}^k\rangle \\ &+\frac{1}{2}(\|\mathbf{u}-\mathbf{u}^{k+1}\|_{\mathbf{H}}^2 - \|\mathbf{u}-\mathbf{u}^k\|_{\mathbf{H}}^2) + \frac{1}{2}\|\mathbf{u}^k-\tilde{\mathbf{u}}^k\|_{\widehat{\mathbf{H}}_1}^2. \quad (36)\end{aligned}$$

令 $\mathbf{u}$ 在（36）中取任意 $\mathbf{u}^*\in\mathcal{M}^*$，可得

$$\begin{aligned}&\frac{1}{2}\|\mathbf{u}^{k+1}-\mathbf{u}^*\|_{\mathbf{H}}^2 \leq \frac{1}{2}\|\mathbf{u}^k-\mathbf{u}^*\|_{\mathbf{H}}^2 - \frac{1}{2}\|\mathbf{u}^k-\tilde{\mathbf{u}}^k\|_{\widehat{\mathbf{H}}_1}^2 \\ &+\langle \tilde{\mathbf{x}}^k-\mathbf{x}^*, \mathbf{d}^k\rangle + G(\mathbf{x}^*)-G(\tilde{\mathbf{x}}^k)+\langle \mathbf{u}^*-\tilde{\mathbf{u}}^k, \mathcal{K}_1(\mathbf{u}^*)\rangle.\end{aligned}$$

那么，由（17），命题（24）成立。

2) 我们引入以下记号。

$$\tilde{\mathbf{y}}^k = \mathrm{prox}_G^{\Gamma^{-1}}(\mathbf{x}^k-\Gamma(\nabla F(\mathbf{x}^k)-\mathbf{V}\boldsymbol{\alpha}^k)), \quad (37a)$$

$$\tilde{\mathbf{z}}^k = \boldsymbol{\alpha}^k-\beta\mathbf{V}\tilde{\mathbf{y}}^k, \quad (37b)$$

$$\mathbf{v}^{k+1} = \mathbf{u}^k-\mathbf{M}(\mathbf{u}^k-\tilde{\mathbf{v}}^k), \quad (37c)$$

其中 $\tilde{\mathbf{v}}^k=$ 列向量$\{\tilde{\mathbf{y}}^k, \tilde{\mathbf{z}}^k\}$。由 (24) 可知

$$\|\mathbf{v}^{k+1}-\mathbf{u}^*\|_{\mathbf{H}}^2 \leq \|\mathbf{u}^k-\mathbf{u}^*\|_{\mathbf{H}}^2 - \|\mathbf{u}^k-\tilde{\mathbf{v}}^k\|_{\widehat{\mathbf{H}}_1}^2, \forall k\geq 0,$$

这意味着 $\|\mathbf{v}^{k+1}-\mathbf{u}^*\|_{\mathbf{H}} \leq \|\mathbf{u}^k-\mathbf{u}^*\|_{\mathbf{H}}, k\geq 0$。因此，对于 $\forall \mathbf{u}^*\in\mathcal{M}^*$，有

$$\begin{aligned}\|\mathbf{u}^{k+1}-\mathbf{u}^*\|_{\mathbf{H}} &\leq \|\mathbf{v}^{k+1}-\mathbf{u}^*\|_{\mathbf{H}} + \|\mathbf{u}^{k+1}-\mathbf{v}^{k+1}\|_{\mathbf{H}} \\ &\leq \|\mathbf{u}^k-\mathbf{u}^*\|_{\mathbf{H}} + \|\mathbf{u}^{k+1}-\mathbf{v}^{k+1}\|_{\mathbf{H}}. \quad (38)\end{aligned}$$

根据近端映射的非扩张性，有

$$\|\tilde{\mathbf{x}}^k-\tilde{\mathbf{y}}^k\| \leq \sigma_M(\Gamma)\|\mathbf{d}^k\| \leq \sqrt{N}\sigma_M(\Gamma)\varepsilon_k, \forall k\geq 0.$$

因此，对于 $\forall k\geq 0$，可得

$$\begin{aligned}&\|\mathbf{u}^{k+1}-\mathbf{v}^{k+1}\|_{\mathbf{H}} = \|\mathbf{M}(\tilde{\mathbf{u}}^k-\tilde{\mathbf{v}}^k)\|_{\mathbf{H}} \\ &\leq \sigma_M(\mathbf{H}^{\frac{1}{2}}\mathbf{M})(\|\tilde{\mathbf{x}}^k-\tilde{\mathbf{y}}^k\| + \|\beta\mathbf{V}(\tilde{\mathbf{x}}^k-\tilde{\mathbf{y}}^k)\|) \\ &\leq \underbrace{\sqrt{N}\sigma_M(\mathbf{H}^{\frac{1}{2}}\mathbf{M})\sigma_M(\Gamma)(\beta\sigma_M(\mathbf{V})+1)}_{:=\bar{\mu}>0}\varepsilon_k. \quad (39)\end{aligned}$$

结合式（38）和式（39），可得

$$\|\mathbf{u}^{k+1}-\mathbf{u}^*\|_{\mathbf{H}} \leq \|\mathbf{u}^k-\mathbf{u}^*\|_{\mathbf{H}} + \bar{\mu}\varepsilon_k, \forall \mathbf{u}^*\in\mathcal{M}^*. \quad (40)$$

将不等式 (40) 对 $k=0、1,\cdots、\mathrm{K}-1$求和，可得对任意 $\mathrm{K}\geq 1$,

$$\sum_{k=0}^{\mathrm{K}-1}\{\|\mathbf{u}^{k+1}-\mathbf{u}^*\|_{\mathbf{H}} - \|\mathbf{u}^k-\mathbf{u}^*\|_{\mathbf{H}}\} \leq \sum_{k=0}^{\mathrm{K}-1}\bar{\mu}\varepsilon_k,$$

这意味着

$$\|\mathbf{u}^{\mathrm{K}}-\mathbf{u}^*\|_{\mathbf{H}} \leq \|\mathbf{u}^0-\mathbf{u}^*\|_{\mathbf{H}} + \sum_{k=0}^{\mathrm{K}-1}\bar{\mu}\varepsilon_k, \forall \mathrm{K}\geq 1.$$

由于$\mathbf{H}\succ 0$ 和 $\{\varepsilon_k\}$ 可求和，因此可推出，对于任意$\mathbf{u}^*\in\mathcal{M}^*$, $\{\|\mathbf{u}^k-\mathbf{u}^*\|\}$ 都有界。

3）对 24 在 $k=0,1,\cdots,\infty$上求和，可得

$$\sum_{k=0}^{\infty}\|\mathbf{u}^k-\tilde{\mathbf{u}}^k\|_{\widehat{\mathbf{H}}_1}^2 \leq \|\mathbf{u}^0-\mathbf{u}^*\|_{\mathbf{H}}^2 + \sum_{k=0}^{\infty}2\sqrt{N}\|\tilde{\mathbf{x}}^k-\mathbf{x}^*\|\varepsilon_k.$$

由于 $\{\varepsilon_k\}$ 可求和，且 $\{\mathbf{u}^k\}$ 有界，因此有

$$\sum_{k=0}^{\infty}\|\mathbf{u}^k-\tilde{\mathbf{u}}^k\|_{\widehat{\mathbf{H}}_1}^2 < \infty, \quad (41)$$

这意味着 $\|\tilde{\mathbf{u}}^k-\mathbf{u}^k\|^2\to 0, k\to\infty$. 设$\mathbf{u}^\infty$ 是 $\{\tilde{\mathbf{u}}^k\}$ 和 $\{\tilde{\mathbf{u}}^{k_j}\}$ 的一个聚簇点，并且 $\mathbf{u}^\infty$是一个收敛至该点的子序列。于是，由（32）可得

$$\begin{aligned}&G(\mathbf{x})-G(\tilde{\mathbf{x}}^{k_j})+\langle \mathbf{x}-\tilde{\mathbf{x}}^{k_j}, \nabla F(\mathbf{x}^{k_j})\rangle + \langle \mathbf{u}-\tilde{\mathbf{u}}^{k_j}, \mathcal{K}_2(\tilde{\mathbf{u}}^{k_j})\rangle \\ &\geq \langle \mathbf{u}-\tilde{\mathbf{u}}^{k_j}, \mathbf{Q}(\mathbf{u}^{k_j}-\tilde{\mathbf{u}}^{k_j})\rangle + \langle \mathbf{x}-\tilde{\mathbf{x}}^{k_j}, \mathbf{d}^{k_j}\rangle, \forall \mathbf{u}\in\mathcal{M}.\end{aligned}$$

由于矩阵$\mathbf{Q}$ 是非奇异的，且 G、 ∇F 和 $\mathcal{K}_2$ 均连续，在上述不等式中令 $k_j\to\infty$，可得

$$G(\mathbf{x})-G(\mathbf{x}^\infty)+\langle \mathbf{u}-\mathbf{u}^\infty, \mathcal{K}_1(\mathbf{u}^\infty)\rangle \geq 0, \forall \mathbf{u}\in\mathcal{M}.$$

与（17）相比，我们可以得出 $\mathbf{u}^\infty\in\mathcal{M}^*$。因此，由（40），有

$$\|\mathbf{u}^{k+1}-\mathbf{u}^\infty\|_{\mathbf{H}} \leq \|\mathbf{u}^k-\mathbf{u}^\infty\|_{\mathbf{H}} + \bar{\mu}\varepsilon_k.$$

由于$\sum_{k=0}^{\infty}\varepsilon_k<\infty$，根据 [27, 引理 3.2]，准 Fej´er 单调序列 $\{\|\mathbf{u}^k-\mathbf{u}^\infty\|_{\mathbf{H}}\}$ 收敛到唯一的极限点。于是，由于 $\mathbf{u}^\infty$ 是{$\mathbf{u}^k$}的一个聚点，可得 $\lim_{k\to\infty}\mathbf{u}^k=\mathbf{u}^\infty$。 ■

Appendix C
Proof of Theorem 2

Proof: By (21a) and (21b), we have

$$\begin{aligned}&\mathbf{0}\in\partial G(\tilde{\mathbf{x}}^k)+\nabla F(\mathbf{x}^k)-\mathbf{V}\boldsymbol{\alpha}^k+\Gamma^{-1}(\tilde{\mathbf{x}}^k-\mathbf{x}^k)-\mathbf{d}^k, \\ &\mathbf{0}=\beta\mathbf{V}\tilde{\mathbf{x}}^k+(\tilde{\boldsymbol{\alpha}}^k-\boldsymbol{\alpha}^k),\end{aligned}$$

这意味着

$$\begin{aligned}&\nabla F(\tilde{\mathbf{x}}^k)-\nabla F(\mathbf{x}^k)-\mathbf{V}(\tilde{\boldsymbol{\alpha}}^k-\boldsymbol{\alpha}^k)-\Gamma^{-1}(\tilde{\mathbf{x}}^k-\mathbf{x}^k)+\mathbf{d}^k \\ &\in\partial G(\tilde{\mathbf{x}}^k)+\nabla F(\tilde{\mathbf{x}}^k)-\mathbf{V}\tilde{\boldsymbol{\alpha}}^k,\end{aligned}$$

并且 $-\frac{1}{\beta}(\tilde{\boldsymbol{\alpha}}^k-\boldsymbol{\alpha}^k)=\mathbf{V}\tilde{\mathbf{x}}^k$。因此，有

$$\mathrm{dist}^2(\mathbf{0}, \mathcal{J}(\tilde{\mathbf{u}}^k)) \leq \|\nabla F(\tilde{\mathbf{x}}^k)-\nabla F(\mathbf{x}^k)-\mathbf{V}(\tilde{\boldsymbol{\alpha}}^k-\boldsymbol{\alpha}^k)$$

$$
\begin{aligned}
&-\Gamma^{-1}(\tilde{\mathbf{x}}^k-\mathbf{x}^k)+\mathbf{d}^k\|^2+\|\frac{1}{\beta}(\tilde{\boldsymbol{\alpha}}^k-\boldsymbol{\alpha}^k)\|^2\\
&\le(4\max_i\{L_i^2\}+4\sigma_m(\Gamma^2))\|\tilde{\mathbf{x}}^k-\mathbf{x}^k\|^2\\
&\quad+(4\sigma_M(\mathbf{V}^2)+\frac{1}{\beta^2})\|\tilde{\boldsymbol{\alpha}}^k-\boldsymbol{\alpha}^k\|^2+4\|\mathbf{d}^k\|^2\\
&\le\kappa_1^2\|\tilde{\mathbf{u}}^k-\mathbf{u}^k\|^2_{\hat{\mathbf{H}}_1}+4N(\varepsilon_k)^2, \qquad (42)
\end{aligned}
$$

其中 $\kappa_1^2=\max\{4\sigma_M(\mathbf{V}^2)+\frac{1}{\beta^2},4\max_i\{L_i^2\}+4\sigma_m(\Gamma^{-2})\}$。

因此，由（41）和 [20,命题1]，$O(1/(K+1))$运行平均最优性残差和 $o(1/(K+1))$ 运行最优最优性残差成立。

将（34）和（35）代入（23）可得

$$
\begin{aligned}
&\phi(\mathbf{x})-\phi(\tilde{\mathbf{x}}^k)+\langle\mathbf{u}-\tilde{\mathbf{u}}^k,\mathcal{K}_2(\mathbf{u})\rangle\ge\langle\mathbf{x}-\tilde{\mathbf{x}}^k,\mathbf{d}^k\rangle\\
&+\frac{1}{2}(\|\mathbf{u}-\mathbf{u}^{k+1}\|^2_{\mathbf{H}}-\|\mathbf{u}-\mathbf{u}^k\|^2_{\mathbf{H}})+\frac{1}{2}\|\mathbf{u}^k-\tilde{\mathbf{u}}^k\|^2_{\hat{\mathbf{H}}_2}. \qquad (43)
\end{aligned}
$$

注意到

$$
L(\tilde{\mathbf{x}}^k,\boldsymbol{\alpha})-L(\mathbf{x},\tilde{\boldsymbol{\alpha}}^k)=\phi(\tilde{\mathbf{x}}^k)-\phi(\mathbf{x})+\langle\tilde{\mathbf{u}}^k-\mathbf{u},\mathcal{K}_2(\mathbf{u})\rangle.
$$

对（43）关于 $k=0,1,\cdots,K-1$求和，可得

$$
\begin{aligned}
&2\sum_{k=0}^{K-1}(L(\tilde{\mathbf{x}}^k,\boldsymbol{\alpha})-L(\mathbf{x},\tilde{\boldsymbol{\alpha}}^k))\\
&\le\|\mathbf{u}^0-\mathbf{u}\|^2_{\mathbf{H}}+2\sqrt{N}\sum_{k=0}^{K-1}\|\tilde{\mathbf{x}}^k-\mathbf{x}\|\varepsilon_k.
\end{aligned}
$$

由 F、G 的凸性以及（$\mathbf{X}^K$，$\boldsymbol{\Lambda}^K$）的定义可知

$$
K(L(\mathbf{X}^K,\boldsymbol{\alpha})-L(\mathbf{x},\boldsymbol{\Lambda}^K))\le\sum_{k=0}^{K-1}(L(\tilde{\mathbf{x}}^k,\boldsymbol{\alpha})-L(\mathbf{x},\tilde{\boldsymbol{\alpha}}^k)).
$$

因此，原始–对偶间隙

$$
\begin{aligned}
&L(\mathbf{X}^K,\boldsymbol{\alpha})-L(\mathbf{x},\boldsymbol{\Lambda}^K)\\
&\le\frac{\|\mathbf{u}^0-\mathbf{u}\|^2_{\mathbf{H}}+2\sqrt{N}\sum_{k=0}^{K-1}\|\tilde{\mathbf{x}}^k-\mathbf{x}\|\varepsilon_k}{2K} \qquad (44)
\end{aligned}
$$

成立。由于 $\{\varepsilon_k\}$ 可求和且 $\{\mathbf{u}^k\}$ 有界，原始-对偶间隙为 $O(1/K)$。

注意，不等式（44）对所有（$\mathbf{x}$ $\boldsymbol{\alpha}$）$\in\mathcal{M}$均成立，因此对任意最优解$\mathbf{x}^*$ 和 $\mathcal{B}_\rho=\{\boldsymbol{\alpha}:\|\boldsymbol{\alpha}\|\le\rho\}$，其中 $\rho>0$ 是任意给定的正数。令$\mathbf{u}=$列向量$\{\mathbf{x}^*,\boldsymbol{\alpha}\}$ 以及$\mathbf{U}^K=$列向量$\{\mathbf{X}^K,\boldsymbol{\Lambda}^K\}$可得

$$
\begin{aligned}
&\sup_{\boldsymbol{\alpha}\in\mathcal{B}_\rho}\{\mathcal{L}(\mathbf{X}^K,\boldsymbol{\alpha})-\mathcal{L}(\mathbf{x}^*,\boldsymbol{\Lambda}^K)\}\\
&=\sup_{\boldsymbol{\alpha}\in\mathcal{B}_\rho}\{\phi(\mathbf{X}^K)-\phi(\mathbf{x}^*)+\langle\mathbf{U}^K-\mathbf{u},\mathcal{K}_2(\mathbf{u})\rangle\}\\
&\overset{(19)}{=}\sup_{\boldsymbol{\alpha}\in\mathcal{B}_\rho}\{\phi(\mathbf{X}^K)-\phi(\mathbf{x}^*)+\langle\mathbf{U}^K-\mathbf{u},\mathcal{K}_2(\mathbf{U}^K)\rangle\}\\
&=\sup_{\boldsymbol{\alpha}\in\mathcal{B}_\rho}\{\phi(\mathbf{X}^K)-\phi(\mathbf{x}^*)+\langle\mathbf{X}^K-\mathbf{x}^*,-\mathbf{V}\boldsymbol{\Lambda}^K\rangle\\
&\qquad+\langle\boldsymbol{\Lambda}^k-\boldsymbol{\alpha},\mathbf{V}\mathbf{X}^K\rangle\}\\
&=\sup_{\boldsymbol{\alpha}\in\mathcal{B}_\rho}\{\phi(\mathbf{X}^K)-\phi(\mathbf{x}^*)-\langle\boldsymbol{\alpha},\mathbf{V}\mathbf{X}^K\rangle\}\\
&=\phi(\mathbf{X}^K)-\phi(\mathbf{x}^*)+\rho\|\mathbf{V}\mathbf{X}^K\|. \qquad (45)
\end{aligned}
$$

注意 $\boldsymbol{\alpha}^0=\mathbf{0}$。结合（44）和（45），可得

$$
\begin{aligned}
&\phi(\mathbf{X}^K)-\phi(\mathbf{x}^*)+\rho\|\mathbf{V}\mathbf{X}^K\|\\
&\le\sup_{\boldsymbol{\alpha}\in\mathcal{B}_\rho}\{\frac{\|\mathbf{u}^0-\mathbf{u}\|^2_{\mathbf{H}}+2\sqrt{N}\sum_{k=0}^{K-1}\|\tilde{\mathbf{x}}^k-\mathbf{x}^*\|\varepsilon_k}{2K}\}\\
&=\frac{\|\mathbf{x}^0-\mathbf{x}^*\|^2_{\Gamma^{-1}}+\frac{1}{\beta}\rho^2+2\sqrt{N}\sum_{k=0}^{K-1}\|\tilde{\mathbf{x}}^k-\mathbf{x}^*\|\varepsilon_k}{2K}.
\end{aligned}
$$

由于 $\{\mathbf{u}^k\}$ 有界且 $\sum_{k=0}^{\infty}\varepsilon_k<\infty$，因此存在 $0<\psi<\infty$ 使得

$$
\phi(\mathbf{X}^K)-\phi(\mathbf{x}^*)+\rho\|\mathbf{V}\mathbf{X}^K\|\le\frac{\psi}{K}. \qquad (46)
$$

对于任意 $\rho>\|\boldsymbol{\alpha}^*\|$，应用 [34, 引理2。3] 在（46）中，可得

$$
\begin{aligned}
&\|\mathbf{V}\mathbf{X}^K\|\le\frac{\psi}{K(\rho-\|\boldsymbol{\alpha}^*\|)},\\
&-\frac{\psi\|\boldsymbol{\alpha}^*\|}{K(\rho-\|\boldsymbol{\alpha}^*\|)}\le\phi(\mathbf{X}^K)-\phi(\mathbf{x}^*)\le\frac{\psi}{K}.
\end{aligned}
$$

令 $\rho=\max\{1+\|\boldsymbol{\alpha}^*\|,2\|\boldsymbol{\alpha}^*\|\}$，则有

$$
|\phi(\mathbf{X}^K)-\phi(\mathbf{x}^*)|\le\frac{\psi}{K},\|\sqrt{\mathbf{I}-\mathbf{W}}\mathbf{X}^K\|\le\frac{\psi}{K}.
$$

因此，原始间隙和共识误差为 $O(1/K)$。 ■

附录 D 定理的证明 3

证明：由于 $\varepsilon_k=0$ 对所有 $k\ge0$，由（32）可得

$$
\begin{aligned}
&G(\mathbf{x})-G(\tilde{\mathbf{x}}^k)+\langle\mathbf{x}-\tilde{\mathbf{x}}^k,\nabla F(\mathbf{x}^k)\rangle\\
&\quad+\langle\mathbf{u}-\tilde{\mathbf{u}}^k,\mathcal{K}_2(\tilde{\mathbf{u}}^k)\rangle\\
&\ge\langle\mathbf{u}-\tilde{\mathbf{u}}^k,\mathbf{Q}(\mathbf{u}^k-\tilde{\mathbf{u}}^k)\rangle,\forall\mathbf{u}\in\mathcal{M}. \qquad (47)
\end{aligned}
$$

令 $\mathbf{u}=\tilde{\mathbf{u}}^{k+1}$ 代入（47），可得

$$
\begin{aligned}
&G(\tilde{\mathbf{x}}^{k+1})-G(\tilde{\mathbf{x}}^k)+\langle\tilde{\mathbf{x}}^{k+1}-\tilde{\mathbf{x}}^k,\nabla F(\mathbf{x}^k)\rangle\\
&\quad+\langle\tilde{\mathbf{u}}^{k+1}-\tilde{\mathbf{u}}^k,\mathcal{K}_2(\tilde{\mathbf{u}}^k)\rangle\\
&\ge\langle\tilde{\mathbf{u}}^{k+1}-\tilde{\mathbf{u}}^k,\mathbf{Q}(\mathbf{u}^k-\tilde{\mathbf{u}}^k)\rangle. \qquad (48)
\end{aligned}
$$

将不等式（47）改写为第（$k+1$）次迭代的形式，可得

$$
\begin{aligned}
&G(\mathbf{x})-G(\tilde{\mathbf{x}}^{k+1})+\langle\mathbf{x}-\tilde{\mathbf{x}}^{k+1},\nabla F(\mathbf{x}^{k+1})\rangle\\
&\quad+\langle\mathbf{u}-\tilde{\mathbf{u}}^{k+1},\mathcal{K}_2(\tilde{\mathbf{u}}^{k+1})\rangle\\
&\ge\langle\mathbf{u}-\tilde{\mathbf{u}}^{k+1},\mathbf{Q}(\mathbf{u}^{k+1}-\tilde{\mathbf{u}}^{k+1})\rangle,\forall\mathbf{u}\in\mathcal{M}. \qquad (49)
\end{aligned}
$$

在（**49**）中令 $\mathrm{u}=\tilde{\mathbf{u}}^k$，可得

$$
\begin{aligned}
&G(\tilde{\mathbf{x}}^k)-G(\tilde{\mathbf{x}}^{k+1})+\langle\tilde{\mathbf{x}}^k-\tilde{\mathbf{x}}^{k+1},\nabla F(\mathbf{x}^{k+1})\rangle\\
&\quad+\langle\tilde{\mathbf{u}}^k-\tilde{\mathbf{u}}^{k+1},\mathcal{K}_2(\tilde{\mathbf{u}}^{k+1})\rangle\\
&\ge\langle\tilde{\mathbf{u}}^k-\tilde{\mathbf{u}}^{k+1},\mathbf{Q}(\mathbf{u}^{k+1}-\tilde{\mathbf{u}}^{k+1})\rangle,\forall\mathbf{u}\in\mathcal{M}. \qquad (50)
\end{aligned}
$$

注意 $\langle\tilde{\mathbf{u}}^k-\tilde{\mathbf{u}}^{k+1},\mathcal{K}_2(\tilde{\mathbf{u}}^k)-\mathcal{K}_2(\tilde{\mathbf{u}}^{k+1})\rangle\equiv0$。将（48）与（50）相加，得到

$$
\begin{aligned}
&\langle\tilde{\mathbf{u}}^k-\tilde{\mathbf{u}}^{k+1},\mathbf{Q}((\mathbf{u}^k-\tilde{\mathbf{u}}^k)-(\mathbf{u}^{k+1}-\tilde{\mathbf{u}}^{k+1}))\rangle\\
&\ge\langle\tilde{\mathbf{x}}^k-\tilde{\mathbf{x}}^{k+1},\nabla F(\mathbf{x}^k)-\nabla F(\mathbf{x}^{k+1})\rangle. \qquad (51)
\end{aligned}
$$

然后，将该项

$$
\langle(\mathbf{u}^k-\tilde{\mathbf{u}}^k)-(\mathbf{u}^{k+1}-\tilde{\mathbf{u}}^{k+1}),\mathbf{Q}((\mathbf{u}^k-\tilde{\mathbf{u}}^k)-(\mathbf{u}^{k+1}-\tilde{\mathbf{u}}^{k+1}))\rangle
$$

添加到等式两边（51），并利用恒等式

$$\mathbf{u}^\top\mathbf{Q}\mathbf{u}=\frac{1}{2}\mathbf{u}^\top(\mathbf{Q}^\top+\mathbf{Q})\mathbf{u},\forall\mathbf{u}\in\mathcal{M},$$

可得

$$\begin{aligned}&\underbrace{\langle\mathbf{u}^k-\mathbf{u}^{k+1},\mathbf{Q}((\mathbf{u}^k-\tilde{\mathbf{u}}^k)-(\mathbf{u}^{k+1}-\tilde{\mathbf{u}}^{k+1}))\rangle}_{\text{(I)}}\\&\geq\underbrace{\langle\tilde{\mathbf{x}}^k-\tilde{\mathbf{x}}^{k+1},\nabla F(\mathbf{x}^k)-\nabla F(\mathbf{x}^{k+1})\rangle}_{\text{(II)}}\\&\quad+\frac{1}{2}\|(\mathbf{u}^k-\tilde{\mathbf{u}}^k)-(\mathbf{u}^{k+1}-\tilde{\mathbf{u}}^{k+1})\|^2_{\mathbf{Q}^\top+\mathbf{Q}}.\end{aligned}\tag{52}$$

一方面，我们有

$$\begin{aligned}\text{(I)}=&\langle\mathbf{u}^k-\mathbf{u}^{k+1},\mathbf{Q}((\mathbf{u}^k-\tilde{\mathbf{u}}^k)-(\mathbf{u}^{k+1}-\tilde{\mathbf{u}}^{k+1}))\rangle\\\overset{(21c)}{=}&\langle\mathbf{M}(\mathbf{u}^k-\tilde{\mathbf{u}}^k),\mathbf{Q}((\mathbf{u}^k-\tilde{\mathbf{u}}^k)-(\mathbf{u}^{k+1}-\tilde{\mathbf{u}}^{k+1}))\rangle\\\overset{(20)}{=}&\langle\mathbf{u}^k-\tilde{\mathbf{u}}^k,\mathbf{M}^\top\mathbf{H}\mathbf{M}((\mathbf{u}^k-\tilde{\mathbf{u}}^k)-(\mathbf{u}^{k+1}-\tilde{\mathbf{u}}^{k+1}))\rangle.\end{aligned}$$

另一方面，我们有

$$\begin{aligned}-\text{(II)}=&\langle\nabla F(\mathbf{x}^{k+1})-\nabla F(\mathbf{x}^k),(\tilde{\mathbf{x}}^k-\mathbf{x}^k)-(\tilde{\mathbf{x}}^{k+1}-\mathbf{x}^{k+1})\rangle\\&+\langle\nabla F(\mathbf{x}^{k+1})-\nabla F(\mathbf{x}^k),\mathbf{x}^k-\mathbf{x}^{k+1}\rangle\\\leq&\frac{\epsilon}{2}\|(\tilde{\mathbf{x}}^k-\mathbf{x}^k)-(\tilde{\mathbf{x}}^{k+1}-\mathbf{x}^{k+1})\|^2_{\mathbf{L}_F}\\&+\frac{1}{2\epsilon}\|\nabla F(\mathbf{x}^{k+1})-\nabla F(\mathbf{x}^k)\|^2_{\mathbf{L}_F^{-1}}\\&+\langle\nabla F(\mathbf{x}^{k+1})-\nabla F(\mathbf{x}^k),\mathbf{x}^k-\mathbf{x}^{k+1}\rangle\\\overset{(3)}{\leq}&\frac{\epsilon}{2}\|(\tilde{\mathbf{x}}^k-\mathbf{x}^k)-(\tilde{\mathbf{x}}^{k+1}-\mathbf{x}^{k+1})\|^2_{\mathbf{L}_F}\\&+(1-\frac{1}{2\epsilon})\langle\nabla F(\mathbf{x}^{k+1})-\nabla F(\mathbf{x}^k),\mathbf{x}^k-\mathbf{x}^{k+1}\rangle.\end{aligned}$$

令 $\epsilon=\frac{1}{2}$，则有

$$\text{(II)}\geq-\frac{1}{4}\|(\tilde{\mathbf{x}}^k-\mathbf{x}^k)-(\tilde{\mathbf{x}}^{k+1}-\mathbf{x}^{k+1})\|^2_{\mathbf{L}_F}.$$

因此，由（52），可得

$$\begin{aligned}&2\langle\mathbf{u}^k-\tilde{\mathbf{u}}^k,\mathbf{M}^\top\mathbf{H}\mathbf{M}((\mathbf{u}^k-\tilde{\mathbf{u}}^k)-(\mathbf{u}^{k+1}-\tilde{\mathbf{u}}^{k+1}))\rangle\\&\geq\|(\mathbf{u}^k-\tilde{\mathbf{u}}^k)-(\mathbf{u}^{k+1}-\tilde{\mathbf{u}}^{k+1})\|^2_{\mathbf{Q}^\top+\mathbf{Q}}\\&\quad-\frac{1}{2}\|(\tilde{\mathbf{x}}^k-\mathbf{x}^k)-(\tilde{\mathbf{x}}^{k+1}-\mathbf{x}^{k+1})\|^2_{\mathbf{L}_F}.\end{aligned}\tag{53}$$

由恒等式 $\|a\|^2_{\mathbf{H}}-\|b\|^2_{\mathbf{H}}=2\langle a,\mathbf{H}(a-b)\rangle-\|a-b\|^2_{\mathbf{H}}$ 其中 $a=\mathbf{M}(\mathbf{u}^k-\tilde{}\mathbf{u}^k)$ 且 $b=\mathbf{M}(\mathbf{u}^{k+1}-\tilde{}\mathbf{u}^{k+1})$，以及 $\widehat{\ }\mathbf{H}_1=\mathbf{Q}^\top+\mathbf{Q}-\mathbf{M}^\top\mathbf{H}\mathbf{M}-\frac{1}{2}\text{điag}\{\mathbf{L}_F,\mathbf{0}\}$，我们有

$$\begin{aligned}&\|\mathbf{M}(\mathbf{u}^k-\tilde{\mathbf{u}}^k)\|^2_{\mathbf{H}}-\|\mathbf{M}(\mathbf{u}^{k+1}-\tilde{\mathbf{u}}^{k+1})\|^2_{\mathbf{H}}\\&=2\langle\mathbf{u}^k-\tilde{\mathbf{u}}^k,\mathbf{M}^\top\mathbf{H}\mathbf{M}((\mathbf{u}^k-\tilde{\mathbf{u}}^k)-(\mathbf{u}^{k+1}-\tilde{\mathbf{u}}^{k+1}))\rangle\\&\quad+\|\mathbf{M}((\mathbf{u}^k-\tilde{\mathbf{u}}^k)-(\mathbf{u}^{k+1}-\tilde{\mathbf{u}}^{k+1}))\|^2_{\mathbf{H}}\\&\overset{(53)}{\geq}\|(\mathbf{u}^k-\tilde{\mathbf{u}}^k)-(\mathbf{u}^{k+1}-\tilde{\mathbf{u}}^{k+1})\|^2_{\mathbf{Q}^\top+\mathbf{Q}}\\&\quad+\|\mathbf{M}((\mathbf{u}^k-\tilde{\mathbf{u}}^k)-(\mathbf{u}^{k+1}-\tilde{\mathbf{u}}^{k+1}))\|^2_{\mathbf{H}}\\&\quad-\frac{1}{2}\|(\tilde{\mathbf{x}}^k-\mathbf{x}^k)-(\tilde{\mathbf{x}}^{k+1}-\mathbf{x}^{k+1})\|^2_{\mathbf{L}_F}\\&=\|(\mathbf{u}^k-\tilde{\mathbf{u}}^k)-(\mathbf{u}^{k+1}-\tilde{\mathbf{u}}^{k+1})\|^2_{\widehat{\mathbf{H}}_1}\geq 0.\end{aligned}$$

因此，(25)成立。

根据(24)以及不同范数之间的等价性，存在一个常数 c_0 使得对于任意 $\mathbf{u}^*\in\mathcal{M}^*$

$$\|\mathbf{u}^{k+1}-\mathbf{u}^*\|^2_{\mathbf{H}}\leq\|\mathbf{u}^k-\mathbf{u}^*\|^2_{\mathbf{H}}-c_0\|\mathbf{M}(\mathbf{u}^k-\tilde{\mathbf{u}}^k)\|^2_{\mathbf{H}}.\tag{54}$$

对（54）在 $k=0$、$\cdots$、K 上求和，可得 $\forall\mathbf{u}^*\in\mathcal{M}^*$

$$\sum_{k=0}^{K}c_0\|\mathbf{M}(\mathbf{u}^k-\tilde{\mathbf{u}}^k)\|^2_{\mathbf{H}}\leq\|\mathbf{u}^0-\mathbf{u}^*\|^2_{\mathbf{H}}<\infty.$$

由（25）可知，序列 $\{\|\mathbf{M}(\mathbf{u}^k-\tilde{}\mathbf{u}^k)\|^2_{\mathbf{H}}\}$ 单调不增。因此，我们有

$$\|\mathbf{M}(\mathbf{u}^k-\tilde{\mathbf{u}}^k)\|^2_{\mathbf{H}}\leq\frac{\|\mathbf{u}^0-\mathbf{u}^*\|^2_{\mathbf{H}}}{c_0(k+1)}.$$

由于 $\mathbf{u}^{k+1}-\mathbf{u}^k=-\mathbf{M}(\mathbf{u}^k-\tilde{}\mathbf{u}^k)$，$o(1/k)$ 收敛速率 $\|\mathbf{u}^{k+1}-\mathbf{u}^k\|^2_{\mathbf{H}}$ 可由 [20, 命题 1]得出。由（42），$o(1/k)$一阶最优性残差的收敛速率成立。 ■

附录 E 定理4的证明

证明：由于 $\{\mathbf{u}^k\}$ 和 $\{\tilde{}\mathbf{u}^k\}$ 收敛到$\mathbf{u}^\infty$，且 $\{\|\mathbf{d}^k\|\}$收敛到 0 当 $k\to\infty$时，由（39）可知 $\{\mathbf{v}^k\}$ 和 $\{\tilde{}\mathbf{v}^k\}$也收敛到 $\mathbf{u}^\infty$ 当 $k\to\infty$时。由（37），可得

$$\begin{aligned}&\mathbf{0}\in\partial G(\tilde{\mathbf{y}}^k)+\nabla F(\mathbf{x}^k)-\mathbf{V}\boldsymbol{\alpha}^k+\Gamma^{-1}(\tilde{\mathbf{y}}^k-\mathbf{x}^k),\\&\mathbf{0}=\beta\mathbf{V}\tilde{\mathbf{y}}^k+(\tilde{\mathbf{z}}^k-\boldsymbol{\alpha}^k),\end{aligned}$$

这意味着

$$\begin{aligned}&\nabla F(\tilde{\mathbf{y}}^k)-\nabla F(\mathbf{x}^k)-\mathbf{V}(\tilde{\mathbf{z}}^k-\boldsymbol{\alpha}^k)-\Gamma^{-1}(\tilde{\mathbf{y}}^k-\mathbf{x}^k)\\&\in\partial G(\tilde{\mathbf{y}}^k)+\nabla F(\tilde{\mathbf{x}}^k)-\mathbf{V}\mathbf{z}^k,\end{aligned}$$

并且 $-\frac{1}{\beta}(\tilde{}\mathbf{z}^k-\boldsymbol{\alpha}^k)=\mathbf{V}\tilde{}\mathbf{y}^k$。因此，有

$$\begin{aligned}&\text{dist}^2(\mathbf{0},\mathcal{J}(\tilde{\mathbf{v}}^k))\leq\|\nabla F(\tilde{\mathbf{y}}^k)-\nabla F(\mathbf{x}^k)-\mathbf{V}(\tilde{\mathbf{z}}^k-\boldsymbol{\alpha}^k)\\&\quad-\Gamma^{-1}(\tilde{\mathbf{y}}^k-\mathbf{x}^k)\|^2+\|\frac{1}{\beta}(\tilde{\mathbf{z}}^k-\boldsymbol{\alpha}^k)\|^2\\&\leq(3\max_i\{L_i^2\}+3\sigma_m(\Gamma^2))\|\tilde{\mathbf{y}}^k-\mathbf{x}^k\|^2\\&\quad+(3\sigma_M(\mathbf{V}^2)+\frac{1}{\beta^2})\|\tilde{\mathbf{z}}^k-\boldsymbol{\alpha}^k\|^2\\&\leq\kappa_2^2\|\tilde{\mathbf{v}}^k-\mathbf{u}^k\|^2_{\widehat{\mathbf{H}}_1},\end{aligned}$$

其中 $\kappa_2^2=\underline{\text{最大值}\{3\sigma_M(\mathbf{V}^2)+\frac{1}{\beta^2},3\max_i\{L_i^2\}+3\sigma_m\Gamma^{-2}\}\text{最大值}()(}$ $\widehat{\mathbf{H}})$。由于 KKT 映射 $\mathcal{J}$在 $(\mathbf{u}^\infty,\mathbf{0})$ 处度量次正则，因此存在 $\kappa>0$ 和 $\epsilon>0$，使得当 $\tilde{}\mathbf{v}^k\in\mathcal{B}_\epsilon(\mathbf{u}^\infty)$

$$\begin{aligned}&\text{dist}(\tilde{\mathbf{v}}^k,\mathcal{M}^*)=\text{dist}(\tilde{\mathbf{v}}^k,\mathcal{T}^{-1}(0))\\&\leq\kappa\text{dist}(0,\mathcal{T}(\tilde{\mathbf{v}}^k))\leq\kappa\kappa_2\|\tilde{\mathbf{v}}^k-\mathbf{u}^k\|_{\widehat{\mathbf{H}}_1}.\end{aligned}$$

由于 $\{\mathbf{u}^k\}$ 和 $\{\tilde{}\mathbf{v}^k\}$ 收敛到 $\mathbf{u}^\infty$，因此存在$k\geq 0$ 使得

$$\text{dist}(\tilde{\mathbf{v}}^k,\mathcal{M}^*)\leq\kappa\kappa_2\|\tilde{\mathbf{v}}^k-\mathbf{u}^k\|_{\widehat{\mathbf{H}}_1},\forall k\geq\bar{k}.\tag{55}$$

$\widehat{\ }$由于注意到 $0\prec\widehat{\mathbf{H}}_1\prec\mathbf{H}$，因此有

$$\begin{aligned}\text{dist}(\tilde{\mathbf{v}}^k,\mathcal{M}^*)\geq&\frac{1}{c_1}\text{dist}_{\widehat{\mathbf{H}}_1}(\tilde{\mathbf{v}}^k,\mathcal{M}^*)\\\geq&\frac{1}{c_1}(\text{dist}_{\widehat{\mathbf{H}}_1}(\mathbf{u}^k,\mathcal{M}^*)-\|\tilde{\mathbf{v}}^k-\mathbf{u}^k\|_{\widehat{\mathbf{H}}_1}),\end{aligned}\tag{56}$$

其中 $c_1 = \sigma_M^{1/2}(\mathbf{H})$。结合式 (55) 和式 (56)，可得

$$\begin{aligned}\mathrm{dist}_{\mathbf{H}}(\mathbf{u}^k, \mathcal{M}^*) &\le \frac{c_1}{c_2}\mathrm{dist}_{\widehat{\mathbf{H}}_1}(\mathbf{u}^k, \mathcal{M}^*) \\ &\le \frac{c_1(\kappa\kappa_2 c_1 + 1)}{c_2}\|\tilde{\mathbf{v}}^k - \mathbf{u}^k\|_{\widehat{\mathbf{H}}_1}, \forall k \ge \bar{k}, \end{aligned} \tag{57}$$

其中 $c_2 = \sigma_m^{1/2}(\widehat{\mathbf{H}}_1)$。由（24）可得，对于 $\forall k \ge 0$

$$\|\mathbf{v}^{k+1} - \mathbf{u}^\infty\|_{\mathbf{H}}^2 \le \|\mathbf{u}^k - \mathbf{u}^\infty\|_{\mathbf{H}}^2 - \|\tilde{\mathbf{v}}^k - \mathbf{u}^k\|_{\widehat{\mathbf{H}}_1}^2. \tag{58}$$

结合（57），可得

$$\begin{aligned}&\mathrm{dist}_{\mathbf{H}}^2(\mathbf{u}^k, \mathcal{M}^*) - \mathrm{dist}_{\mathbf{H}}^2(\mathbf{v}^{k+1}, \mathcal{M}^*) \ge \|\tilde{\mathbf{v}}^k - \mathbf{u}^k\|_{\widehat{\mathbf{H}}_1}^2 \\ &\ge \frac{c_2^2}{c_1^2(\kappa\kappa_2 c_1 + 1)^2}\mathrm{dist}_{\mathbf{H}}^2(\mathbf{u}^k, \mathcal{M}^*), \forall k \ge \bar{k},\end{aligned}$$

这意味着

$$\mathrm{dist}_{\mathbf{H}}(\mathbf{v}^{k+1}, \mathcal{M}^*) \le \bar{\vartheta}\ \mathrm{dist}_{\mathbf{H}}(\mathbf{u}^k, \mathcal{M}^*), \forall k \ge \bar{k}, \tag{59}$$

其中 $\bar{\vartheta} = \sqrt{1 - \frac{c_2^2}{c_1^2(\kappa\kappa_2 c_1+1)_2}} < 1$。由（58）以及事实 $\succ 0 \prec \mathbf{H}_1 \prec \mathbf{H}$，可得

$$\frac{c_2}{c_1}\|\tilde{\mathbf{v}}^k - \mathbf{u}^k\|_{\mathbf{H}}^2 \le \|\tilde{\mathbf{v}}^k - \mathbf{u}^k\|_{\widehat{\mathbf{H}}_1}^2 \le \mathbf{dist}_{\mathbf{H}}^2(\mathbf{u}^k, \mathcal{M}^*)\ \forall k \ge \bar{k}.$$

由此可得

$$\|\tilde{\mathbf{v}}^k - \mathbf{u}^k\|_{\mathbf{H}} \le \omega\ \mathrm{dist}_{\mathbf{H}}(\mathbf{u}^k, \mathcal{M}^*), \forall k \ge \bar{k}, \tag{60}$$

其中 $\omega = \sqrt{c_1/c_2}$。根据三角不等式，有

$$\begin{aligned}&\|\mathbf{u}^k - \mathcal{P}_{\mathcal{M}^*}^{\mathbf{H}}(\mathbf{v}^{k+1})\|_{\mathbf{H}} \le \mathrm{dist}_{\mathbf{H}}(\mathbf{u}^k, \mathcal{M}^*) \\ &\quad + \|\mathcal{P}_{\mathcal{M}^*}^{\mathbf{H}}(\mathbf{u}^k) - \mathcal{P}_{\mathcal{M}^*}^{\mathbf{H}}(\mathbf{v}^{k+1})\|_{\mathrm{H}}, \forall k \ge \bar{k}.\end{aligned} \tag{61}$$

随后，根据 $\mathcal{P}_{\mathcal{M}^*}^{\mathbf{H}}$ 的 Lipschitz 连续性，有

$$\|\mathcal{P}_{\mathcal{M}^*}^{\mathbf{H}}(\mathbf{u}^k) - \mathcal{P}_{\mathcal{M}^*}^{\mathbf{H}}(\mathbf{v}^{k+1})\|_{\mathbf{H}} \le \|\mathbf{u}^k - \mathbf{v}^{k+1}\|_{\mathbf{H}},$$

结合（60）和（61），可得

$$\|\mathbf{u}^k - \mathcal{P}_{\mathcal{M}^*}^{\mathbf{H}}(\mathbf{v}^{k+1})\|_{\mathbf{H}} \le (1+\omega)\mathrm{dist}_{\mathbf{H}}(\mathbf{u}^k, \mathcal{M}^*). \tag{62}$$

根据（26）、（39）以及三角不等式，有

$$\begin{aligned}&\|\mathbf{u}^{k+1} - \mathcal{P}_{\mathcal{M}^*}^{\mathbf{H}}(\mathbf{v}^{k+1})\|_{\mathbf{H}} \\ &\le \|\mathbf{v}^{k+1} - \mathbf{u}^{k+1}\|_{\mathbf{H}} + \mathrm{dist}_{\mathbf{H}}(\mathbf{v}^{k+1}, \mathcal{M}^*) \\ &\le \bar{\mu}\|\mathbf{d}^k\| + \mathrm{dist}_{\mathbf{H}}(\mathbf{v}^{k+1}, \mathcal{M}^*) \\ &\le \mu\varrho_k\|\mathbf{u}^k - \mathbf{u}^{k+1}\|_{\mathbf{H}} + \mathrm{dist}_{\mathbf{H}}(\mathbf{v}^{k+1}, \mathcal{M}^*) \\ &\le \mu\varrho_k(\|\mathbf{u}^{k+1} - \mathcal{P}_{\mathcal{M}^*}^{\mathbf{H}}(\mathbf{v}^{k+1})\| \\ &\quad + \|\mathbf{u}^k - \mathcal{P}_{\mathcal{M}^*}^{\mathbf{H}}(\mathbf{v}^{k+1})\|) + \mathrm{dist}_{\mathbf{H}}(\mathbf{v}^{k+1}, \mathcal{M}^*),\end{aligned}$$

其中 $\mu = \bar{\mu}/c_2$。然后，根据

$$\|\mathbf{u}^{k+1} - \mathcal{P}_{\mathcal{M}^*}^{\mathbf{H}}(\mathbf{v}^{k+1})\|_{\mathbf{H}} \ge \mathrm{dist}_{\mathbf{H}}(\mathbf{u}^{k+1}, \mathcal{M}^*)$$

¯以及 (62)，可推得当 $k \ge \bar{k}$

$$\begin{aligned}&(1 - \mu\varrho_k)\mathrm{dist}_{\mathbf{H}}(\mathbf{u}^{k+1}, \mathcal{M}^*) \\ &\le \mu\varrho_k(1+\omega)\mathrm{dist}_{\mathbf{H}}(\mathbf{u}^k, \mathcal{M}^*) + \mathrm{dist}_{\mathbf{H}}(\mathbf{v}^{k+1}, \mathcal{M}^*),\end{aligned}$$

¯这与（59）结合可知，当 $k \ge \bar{k}$

$$\mathrm{dist}_{\mathbf{H}}(\mathbf{u}^{k+1}, \mathcal{M}^*) \le \underbrace{\frac{\mu(1+\omega)\varrho_k + \bar{\vartheta}}{(1 - \mu\varrho_k)}}_{:=\vartheta_k}\mathrm{dist}_{\mathbf{H}}(\mathbf{u}^k, \mathcal{M}^*).$$

由此可知，如果

$$\sup_{k \ge \bar{k}}\{\varrho_k\} < \frac{1 - \bar{\vartheta}}{\mu(2+\omega)},$$

此时有 $\sup_{k \ge \bar{k}}\{\varrho_k\} < 1$。证明完毕。 ■

参考文献

[1] A. Nedi´c，《分布式平均动态与网络中的优化的收敛速率》，Found. Trends Syst. Control，第2卷，第1期，第1–100页，2015年。[2] A. Nedi´c 和 A. Ozdaglar，《多智能体优化的分布式次梯度方法》，IEEE Trans. Autom. Control，第54卷，第1期，第48–61页，2009年1月。[3] K. Yuan、Q. Ling 和 W. Yin，《关于去中心化梯度下降的收敛性》，SIAM J. Optim.，第26卷，第3期，第1835–1854页，2016年。[4] D. Jakoveti´c、J. Xavier 和 J. M. F. Moura，《快速分布式梯度方法》，IEEE Trans. Autom. Control，第59卷，第5期，第1131–1146页，2014年5月。[5] J. Zhang、K. You 和 T. Bas.ar，《多智能体网络中仅利用相对状态符号的分布式离散时间优化》，IEEE Trans. Autom. Control，第64卷，第6期，第2352–2367页，2019年6月。[6] W. Shi、Q. Ling、K. Yuan、G. Wu 和 W. Yin，《去中心化共识优化中 ADMM 的线性收敛性》，IEEE Trans. Signal Process.，第62卷，第7期，第1750–1761页，2014年4月。[7] Q. Ling、W. Shi、G. Wu 和 A. Ribeiro，《DLM：去中心化线性化交替方向乘子法》，IEEE Trans. Signal Process.，第63卷，第15期，第4051–4064页，2015年8月。[8] W. Shi、Q. Ling、G. Wu 和 W. Yin，《EXTRA：用于去中心化共识优化的精确一阶算法》，SIAM J. Optim.，第25卷，第2期，第944–966页，2015年。[9] A. Nedi´c、A. Olshevsky 和 W. Shi，《在时变图上实现分布式优化的几何收敛》，SIAM J. Optim.，第27卷，第4期，第2597–2633页，2017年。[10] G. Qu 和 N. Li，《利用光滑性加速分布式优化》，IEEE Trans. Control Netw. Syst.，第5卷，第3期，第1245–1260页，2018年9月。[11] J. Xu、S. Zhu、Y. C. Soh 和 L. Xie，《非协调常数步长下多智能体优化的增广分布式梯度方法》，载于第54届 IEEE 决策与控制会议论文集，2015年，第2055–2060页。[12] A. Nedi´c、A. Olshevsky、W. Shi 和 C. A. Uribe，《采用非协调步长的几何收敛分布式优化》，载于美国控制会议论文集，2017年，第3950–3955页。[13] K. Yuan、B. Ying、X. Zhao 和 A. H. Sayed，《用于分布式优化与学习的 Exact Diffusion——第一部分：算法开发》，IEEE Trans. Signal Process.，第67卷，第3期，第708–723页，2019年2月。[14] S. A. Alghunaim、E. Ryu、K. Yuan 和 A. H. Sayed，《具有线性收敛速率的去中心化近端梯度算法》，IEEE Trans. Autom. Control，第66卷，第6期，第2787–2794页，2020年6月。[15] J. Xu、Y. Tian、Y. Sun 和 G. Scutari，《复合优化的分布式算法：统一框架与收敛性分析》，IEEE Trans. Signal Process.，第69卷，第3555–3570页，2021年6月。[16] P. D. Lorenzo 和 G. Scutari，《NEXT：网络内非凸优化》，IEEE Trans. Signal Inf. Process. Netw.，第2卷，第2期，第120–136页，2016年6月。[17] Y. Sun、G. Scutari 和 A. Daneshmand，《重新审视基于梯度跟踪的分布式优化：通过替代函数构造提升收敛速率》，SIAM J. Optim.，第32卷，第2期，第345–385页，2022年。[18] T.-H. Chang、M. Hong 和 X. Wang，《通过非精确共识交替方向乘子法进行多智能体分布式优化》，IEEE Trans. Signal Process.，第63卷，第2期，第482–497页，2015年1月。[19] N. S. Aybat、Z. Wang、T. Lin 和 S. Ma，《用于复合凸共识优化的分布式线性化交替方向乘子法》，IEEE Trans. Autom. Control，第63卷，第1期，第5–20页，2017年1月。[20] W. Shi、Q. Ling、G. Wu 和 W. Yin，《用于去中心化复合优化的近端梯度算法》，IEEE Trans. Signal Process.，第63卷，第22期，第6013–6023页，2015年11月。[21] Z. Li、W. Shi 和 M. Yan，《一种具有网络无关步长和分离收敛速率的去中心化近端梯度方法》，IEEE Trans. Signal Process.，第67卷，第17期，第4494–4506页，2019年9月。[22] J. Zhang、H. Liu、A. M.-C. So 和 Q. Ling，《用于去中心化网络上凸复合优化的惩罚交替方向乘子法》，IEEE Trans. Signal Process.，第69卷，第4282–4295页

，2021年6月。

[23] R. J. Tibshirani 和 J. Taylor，《广义套索的解路径》，统计年刊，第39卷，第3期，第1335–1371页，2011年6月。[24] R. Tibshirani、M. Saunders、S. Rosset、J. Zhu 和 K. Knight，《通过融合套索实现稀疏性和平滑性》，英国皇家统计学会杂志B辑：统计方法论，第67卷，第1期，第91–108页，2005年2月。[25] H. D. Bondell 和 B. J. Reich，《利用 OSCAR 同时进行回归收缩、变量选择和预测变量的监督聚类》，生物统计学，第64卷，第1期，第115–123页，2008年3月。[26] L. Jacob、G. Obozinski、J.-P. Vert，《带重叠的组套索与图套索》，载于第26届国际机器学习会议论文集，2009年，第433–440页。[27] L. Chen、X. Li、D. Sun 和 K.-C. Toh，《关于非精确近端增广拉格朗日法与交替方向乘子法在一类凸复合规划中的等价性》，《数学规划》，第185卷，第111–161页，2021年。[28] F. Jiang、X. Cai、Z. Wu 和 D. Han，《用于鞍点问题的近似一阶原始-对偶算法》，《数学计算》，第90卷，第1227–1262页，2021年。[29] J. Eckstein 和 W. Yao，《由拉格朗日分裂导出的近似交替方向乘子法算法》，《计算与优化应用》，第68卷，2017年，第405–?页。[30] Y. Arjevani、J. Bruna、B. Can、M. Gurbuzbalaban、S. Jegelka 和 H. Lin，《IDEAL：非精确分布式加速增广拉格朗日法》，载于神经信息处理系统进展会议论文集，2020年，第20648–20659页。[31] M. Hong，《通过近端原始-对偶方法分解线性约束非凸问题：算法、收敛性与应用》，2016年，arXiv:1604.00543。[32] A. Beck，《优化中的一阶方法》，第25卷。工业与应用数学学会，费城，2017年。[33] L. Condat、D. Kitahara、A. Contreras 和 A. Hirabayashi，《凸优化的近端分裂算法：近期进展综述及新见解》，2021年，arXiv:1912.00137v7。[34] Y. Xu，《用于线性约束复合凸规划的加速一阶原始-对偶近端方法》，《工业与应用数学学会优化杂志》，第27卷，第3期，第1459–1484页，2017年。[35] E. K. Ryu 和 W. Yin，《大规模凸优化：基于单调算子的算法与分析》，剑桥，2022年。[36] L. Condat，《一种用于涉及利普希茨连续、可近端化函数及线性复合项的凸优化的原始-对偶分裂方法》，《优化理论与应用杂志》，第158卷，第2期，第460–479页，2013年。[37] B. C. V˜u，《一种用于涉及余强制算子的对偶单调包含的分裂算法》，《计算数学进展》，第38卷，2013年，第681–?页。[38] M. Yan，《一种用于最小化含线性算子的三个函数之和的新型原始-对偶算法》，《科学计算杂志》，第76卷，第1698–1717页，2018年。[39] P. Latafat 和 P. Patrinos，《用于求解涉及三个算子的单调包含的非对称前向-后向-伴随分裂方法》，《计算与优化应用》，第68卷，第1期，第57–93页，2017年。[40] B. S. He 和 X. M. Yuan，《预测-校正框架下可分凸优化分裂收缩算法的构造》，2022年，arXiv:2204.11522。[41] T. R. Rockafellar 和 J. B. Wets，《变分分析》，施普林格出版社，柏林，海德堡，2004年。[42] X. Yuan、S. Zeng 和 J. Zhang，《透过变分分析的视角辨析结构化凸优化中交替方向乘子法的线性收敛性》，《机器学习研究杂志》，第21卷，第1–75页，2020年。[43] J. J. Ye、X. Yuan、S. Zeng 和 J. Zhang，《从变分分析视角看若干非光滑凸优化问题一阶方法的线性收敛性》，《集合值变分分析》，第29卷，2021年，第803–837页。[44] Z. Luo 和 P. Tseng，《凸本质光滑最小化中下降方法的线性收敛性》，《工业与应用数学学会控制与优化杂志》，第30卷，第408–425页，1992年。[45] S. M. Robinson，《多面体多值函数的一些连续性性质》，《数学规划》，第14卷，第206–214页，1981年。[46] H. Zou 和 T. Hastie，《通过弹性网进行正则化与变量选择》，英国皇家统计学会杂志B辑：统计方法论，第67卷，第2期，第301–320页，2005年。[47] F. Jiang、Z. Wu、X. Cai 和 H. Zhao，《鞍点问题一阶原始-对偶算法的统一线性收敛性》，《优化快报》，第16卷，第1675–1700页，2022年。



郭路遥 于2020年获得中国太原山西大学信息与计算科学理学学士学位。目前，他正在中国南京东南大学数学学院江苏省网络群体智能重点实验室攻读应用数学博士学位。

他目前的研究方向包括分布式优化与学习。



施新立（IEEE高级会员）分别于2013年、2016年和2019年获得中国南京东南大学软件工程理学学士学位、应用数学理学硕士学位和控制科学与工程博士学位。

2018年，他获得国家留学基金管理委员会奖学金，在澳大利亚维多利亚州墨尔本的皇家墨尔本理工大学学习一年。目前，他是网络科学学院的副教授，并且
东南大学工程学院。他目前的研究兴趣包括分布式优化、非光滑分析和网络控制系统。

施博士曾获中国江苏省优秀博士学位论文奖。



曹进德（IEEE会士）于1986年获中国芜湖安徽师范大学数学学士学位，1989年获中国昆明云南大学数学/应用数学理学硕士学位，1998年获中国成都四川大学数学/应用数学博士学位。

他是讲席教授，担任中国东南大学数学学院院长，以及江苏省网络群体智能重点实验室主任，并负责复杂系统与网络科学研究
中心（中国南京东南大学）。

曹教授曾获中国国家创新奖、俄罗斯自然科学院金奖、奥巴达奖，以及汤森路透/科睿唯安颁发的工程、计算机科学与数学领域高被引研究者奖。他当选为俄罗斯科学院外籍院士、欧洲科学院院士、俄罗斯工程院外籍院士、欧洲科学院与艺术院院士、俄罗斯自然科学院外籍院士、巴基斯坦科学院外籍院士、非洲科学院院士、立陶宛科学院外籍院士，以及国际系统与控制科学院院士。



王子豪 于2020年获得中国南京南京晓庄学院网络工程专业工学学士学位。目前，他正在南京东南大学网络空间科学与工程学院攻读网络空间安全专业工学硕士学位。

他目前的研究方向是网络化多智能体系统的隐私保护分布式优化。