arXiv:2304.02465v2 [math.OC] 3 Feb 2024

求解可分凸优化问题：更快速的预测-校正框架*

TAO ZHANG†，YONG XIA†，和 SHIRU LI†

摘要. He 和 Yuan 的预测-校正框架 [SIAM J. Numer. Anal. 50: 700-709,2012] 能够为求解可分凸优化问题提供收敛算法，速率为 $O(1/t)$（t 表示迭代次数），在遍历意义（迭代的平均值）和逐点意义下均成立。本文提出一种快速预测-校正框架，在非遍历意义（最后一次迭代）下速率为 $O(1/t)$，在逐点意义下为 $O(1/t^2)$。基于该快速预测-校正框架，我们给出三种更快算法，它们在原始-对偶间隙的非遍历意义下达到 $O(1/t)$，在逐点意义下达到 $O(1/t^2)$。第一种算法在求解带线性等式约束的两块可分凸优化时对对偶变量更新两次。第二种算法以 Gauss-Seidel 方式求解带线性等式约束的多块可分凸优化问题。第三种算法以更大的步长求解极小极大问题。

关键词。预测-校正、可分凸优化、非遍历意义、逐点意义。

AMS 学科分类。47H09、47H10、90C25、90C30

1. 引言。 He 和 Yuan 提出了预测-校正框架 [19]（或参见 [21]），用于分析交替方向乘子法（ADMM）的收敛率 [7, 8]。将 ADMM 表示为他们的预测-校正框架后，可立即得到原始-对偶间隙在遍历意义下的 $O(1/t)$ 收敛率（t 表示迭代次数）* 。在一系列后续工作中，He 等人 [11, 12, 13, 15, 20] 进一步证明，预测-校正框架在逐点意义下具有 $O(1/t)$ 收敛率。

预测-校正框架实际上为开发和分析求解以下两个可分凸优化问题的算法提供了一种统一方法。

EXAMPLE 1. 带等式约束的多块可分凸优化问题：

$$\text{(P1)} \qquad \min_{x_i}\left\{f(x)=\sum_{i=1}^{m} f_i(x_i):\ (Ax:=)\sum_{i=1}^{m} A_i x_i = b\right\},$$

其中 $m \geq 1$，f_i $\mathbb{R}^{n_i} \to \mathbb{R}$，$A_i \in \mathbb{R}^{l\times n_i}$，$\sum_{i=1}^{m} n_i =$ n，$b \in \mathbb{R}^l$，：是一个闭的适当凸函数，且。定义 (P) 的拉格朗日函数：

$$L(x,\lambda) := f(x) - \lambda^T(Ax-b),$$

*本研究得到中国国家自然科学基金（项目编号：12171021）和中央高校基本科研业务费的资助。

†中国北京市北航数学科学学院，邮编100191（(T. Zhang) shuxuekuangwu@buaa.edu.cn；(Y. Xia) yxia@buaa.edu.cn；(S. Li, Correspondingauthor) lishiru@buaa.edu.cn）。

*本文中，对于给定的迭代序列 $\{v^t\}$，遍历意义表示$\{v^t\}$的平均值；非遍历意义表示 $\{v^t\}$的最后一个迭代序列；逐点意义表示$\|v^t - v^{t-1}\|^2$。

其中 $\lambda \in \mathbb{R}^l$ 是相应的拉格朗日乘子。为方便起见，我们定义

$$\theta(u) := f(x) = \sum_{i=1}^{m} f_i(x_i),\ u := x = \begin{pmatrix} x_1 \\ \vdots \\ x_m \end{pmatrix},\ w := \begin{pmatrix} x_1 \\ \vdots \\ x_m \\ \lambda \end{pmatrix},\ F(w) := \begin{pmatrix} -A_1^T\lambda \\ \vdots \\ -A_m^T\lambda \\ Ax - b \end{pmatrix}.$$

然后根据最优性条件，(P1) 等价于找到 $w^* = (x_1^*, \ldots, x_m^*, \lambda^*)$ 使得

$$\text{(P1')} \qquad 0 \in T(w^*),\ \text{where } T(w) := \begin{pmatrix} \partial f_1(x_1) \\ \vdots \\ \partial f_m(x_m) \\ 0 \end{pmatrix} + \begin{pmatrix} -A_1^T\lambda \\ \vdots \\ -A_m^T\lambda \\ Ax - b \end{pmatrix}.$$

注意到，对于给定点 $\hat{w} := \begin{pmatrix} x \\ \hat{\lambda} \end{pmatrix} = \begin{pmatrix} \hat{u} \\ \hat{\lambda} \end{pmatrix}\hat{}$，成立

$$\theta(\hat{u}) - \theta(u) - (w - \hat{w})^T F(w) = L(\hat{x}, \lambda) - L(x, \hat{\lambda}),$$

这恰好是原始-对偶间隙。

EXAMPLE 2. 极小极大问题:

$$\text{(P2)} \qquad \min_x \max_y \left\{ \Phi(x, y) := f(x) - y^T Ax - g(y) \right\},$$

其中 $f\colon \mathbb{R}^n \to \mathbb{R}$ 和 $g\colon \mathbb{R}^m \to \mathbb{R}$ 是闭的真凸函数且 $A \in \mathbb{R}^{m\times n}$。为方便起见，我们定义

$$\theta(u) := f(x) + g(y),\ u = w := \begin{pmatrix} x \\ y \end{pmatrix},\ F(w) := \begin{pmatrix} -A^T y \\ Ax \end{pmatrix}.$$

那么，根据最优性条件，(P2) 等价于求解 $w^* = (x^*, y^*)$，使得

$$\text{(P2')} \qquad 0 \in T(w^*),\ \text{where } T(w) := \begin{pmatrix} \partial f(x) \\ \partial g(y) \end{pmatrix} + \begin{pmatrix} -A^T y \\ Ax \end{pmatrix}.$$

注意，对于给定点 $\hat{w} := \hat{u} = \begin{pmatrix} x \\ \hat{y} \end{pmatrix}\hat{}$，有

$$\theta(\hat{u}) - \theta(u) - (w - \hat{w})^T F(w) = \Phi(\hat{x}, y) - \Phi(x, \hat{y}),$$

这正是原始-对偶间隙。

本文假设 (P1) 或 (P2) 的最优解集非空且有界。

遵循这一框架的知名算法不在少数，包括增广拉格朗日法（ALM） [23, 35], 近端增广拉格朗日法 [36, 37], 交替方向乘子法、线性交替方向乘子法 [43], 以及严格压缩的 Peaceman-Rachford 分裂法[11, 13]。这些算法旨在求解带等式约束的两块可分凸优化问题。对于带等式约束的多块可分凸优化问题，雅可比增广拉格朗日法 [10, 42], 交替方向乘子法

TABLE 1.1
Algorithms for solving (P1)

	问题	原始的速率 对偶间隙 函数值或 可行性度量	在.....中的速率 逐点意义
ADMM	$(\mathrm{P1})_2$	$O(1/t)$（遍历） $O(1/\sqrt{t})$（非遍历）	$O(1/t)$
李和林 [26]	$(\mathrm{P1})_2$	$O(1/t)$（非遍历）	-
陈-丁 和朱 [39]	$(\mathrm{P1})_2$	$O(1/t)$（非遍历）	-
瓦尔科宁 [41]	$(\mathrm{P1})_2$	$O(1/t)$(非遍历)	-
Sabach 和 Teboulle [38]	$(\mathrm{P1})_2$	$O(1/t)$(非遍历)	-
Luo [30]	$(\mathrm{P1})_2$	$O(1/t)$(非遍历)	-
Attouch 等人 [1, 2]	$(\mathrm{P1})_2$	-	$O(1/t^2)$
本文	(P1)	$O(1/t)$（非遍历）	$O(1/t^2)$

$(\mathrm{P1})_2$ 表示带有 $m = 2$ 的 (P1)。

带代入的 [16, 17], ADMM 型算法 [18], 均基于预测-校正框架。值得注意的是，对多块 ADMM 发散性的分析 [6] 同样基于预测-校正框架。对于求解极小极大问题（P2），Chambolle-Pock（CP）算法及其变体 [4, 5, 12, 14] 也可以纳入预测-校正框架。

另一方面，针对无约束或简单约束凸优化问题，已有更快的算法。Nesterov [34] 率先提出了一种用于无约束凸优化的加速梯度方法，其收敛率为 $O(1/t^2)$。该方法已扩展至涉及简单近端算子 [3, 40] 的复合凸优化问题（关于二阶微分方程的解释，见 [32] ）。Attouch [1, 2] 通过连续动力系统方法，为求解单调包含问题提出了一种更快的算法，该算法在逐点意义下具有 $O(1/t^2)$ 的收敛率。

值得注意的是，He 和 Yuan [19] 通过预测-校正框架，给出了经典交替方向乘子法原始-对偶间隙的 $O(1/t)$ 遍历收敛率。至于非遍历收敛率，[28, 第 3.4.5 节 1]仅给出了经典交替方向乘子法的 $O(1/\sqrt{t})$ 非遍历收敛率。这比遍历情形慢。事实上，正如 [9], 所示，Golowich 等人给出了理论保证：求解鞍点问题时，非遍历收敛比遍历意义下的收敛更慢。实际上，非遍历收敛在理论和实践中都很重要。

为提高仅在凸性假设下求解线性约束优化问题时非遍历情形的收敛速率，Li 和 Lin [26] 以及 Tran-Dinh 和 Zhu [39] 、Valkonen [41] 提出了加速 ADMM，在求解带有 $m = 2$ 的 (P1) 时达到 $O(1/t)$ 非遍历收敛率。连续动力系统方法 [30]也可以得到相同的非遍历收敛率结果。通过引入所谓的 $nice\ primal\ algorithmic\ map$, Sabach 和 Teboulle [38]提出了一类基于拉格朗日函数的方法，在求解带有$m = 1$ 或 2 的 (P1) 时，对函数值和可行性度量均达到 $O(1/t)$ 非遍历收敛速率。请注意，所有论文 [26, 30, 38, 39, 41] 都未在逐点意义下给出 $O(1/t^2)$ 收敛率。更多详情请参阅表 1.1。

特别地，对于(P1)中带有 $m = 1$ 和等式约束的特殊情形，Luo

TABLE 1.2
Algorithms for solving (P2)

	条件	原始 对偶间隙或 函数值 ， 可行性度量	速率方面 逐点意义
Chambolle 和 Pock [4]	$rs > \rho(A^TA)$	$O(1/t)$（遍历）	$O(1/t)$
Jiang 等人 [25] Li 和 Yan [27]	$rs > 0.75\rho(A^TA)$	$O(1/t)$（遍历）	-
He 等人 [12]	$rs > 0.75\rho(A^TA)$	$O(1/t)$(遍历)	$O(1/t)$
Attouch 等人 [1, 2]	$rs > \rho(A^TA)$	-	$O(1/t^2)$
[26, 38, 39, 41]	$rs > \rho(A^TA)$	$O(1/t)$（非遍历）	-
本文	$rs > 0.75\rho(A^TA)$	$O(1/t)$（非遍历）	$O(1/t^2)$

[29, 31], He 等人 [22] 以及 Bot¸等人 [24] 给出了加速 ALM，具有函数值和可行性度量的 $O(1/t^2)$ 非遍历收敛。仅 Bot¸ 等人 [24] 在光滑目标函数下给出了迭代点的收敛。在求解(P1)中带有 $m = 2$ 和等式约束的特殊情形时，即使交替方向乘子法类型的算法具有易于求解的子问题， [22, 29, 31] 中给出的加速 ALM 的子问题也可能不易求解。未考虑多块结构是收敛率快于 $O(1/t)$ 的原因。带有 $m = 1$ 的(P1)特殊情形并非本文的主要工作。

考虑问题 (P2)。Chambolle 和 Pock [4] 给出了易于求解的子问题算法 (CP)，并在 $rs > \rho(A^TA)$†条件下证明了原始-对偶间隙的 $O(1/t)$遍历收敛。实际上，CP 可以看作求解 (P2) 的近端点算法 (PPA)。因此，可以得到迭代的逐点意义下的 $O(1/t)$。Jiang 等人 [25] 以及 Li 和 Yan [27] 扩展了 CP，在$rs > 0.75\rho(A^TA)$条件下实现了 $O(1/t)$遍历收敛,但没有逐点收敛率。从数值结果来看，当 $\rho(A^TA)$较大时，较小的 rs 有助于加快收敛。最近，He 等人 [12]提出了一种用于求解 (P2) 的 CP 型算法，在 $rs > 0.75\rho(A^TA)$条件下同时实现了遍历意义和逐点意义下的 $O(1/t)$。然而，我们指出， [1, 2, 26, 38, 39, 41] 中给出的用于求解 (P2) 的加速算法是在假设 $rs > \rho(A^TA)$成立的条件下提出的。更多详情请参见表 1.2。

考虑到 He 和 Yuan 的预测-校正框架仅在遍历意义和逐点意义上具有$O(1/t)$收敛率，且在求解 (P1) 和 (P2) 时没有具有非遍历收敛率的统一框架，因此有必要建立一个具有更快收敛率（非遍历意义和逐点意义）的框架。

贡献 我们在下文中列出本文的贡献：

1. 我们提出了一种用于求解 (P1) 和 (P2) 的更快速的预测-校正框架。不同于依赖于 $nice\ primal\ algorithmicmap$ in [38] 或 [1, 2, 30], 中连续动力系统方法的算法，我们的核心是 [19] 中的预测-校正框架，该框架在原始-对偶间隙的遍历意义下具有 $O(1/t)$，并在逐点意义下具有 $O(1/t)$。值得注意的是，我们的框架在原始 - 对偶间隙的非遍历意义下以 $O(1/t)$ 的速率运行算法，并在逐点意义下达到 $O(1/t^2)$。

†$\rho(A^TA)$ 表示 A^TA 的谱。 $1/r$ 和 $1/s$ 实际上作为迭代算法中的步长。

2. He 等人 [11, 13] 提出了一种 ADMM 型算法，在使用 $m = 2$ 和等式约束求解 (P1) 时，两次更新对偶变量，以获得更好的数值性能（或参见注 2）。基于所提出的更快速的预测-校正框架，我们在第 4.1 节给出了一种更快速的 ADMM 算法，通过两次更新对偶变量来求解该问题。这是首篇给出两次更新对偶变量且具有非遍历收敛算法的论文。

3. 基于所提出的更快速预测-校正框架，我们在第 4.2 节给出一种采用 Gauss-Seidel 方式求解 (P1) 的更快速 ADMM 型算法。现有具有非遍历收敛率的加速算法用于求解带有两块等式约束的可分凸优化问题，例如 [26, 30, 38, 39, 41]。鉴于如 [6] 所示，多块 ADMM 会发散，因此，要以 Gauss-Seidel 方式将 [26,30, 38, 39, 41] 中的加速算法扩展到多块情形似乎并不容易。据我们所知，本文首次给出一种采用 Gauss-Seidel 方式求解带等式约束的多块结构凸优化问题的更快速算法，并具有非遍历收敛率。有关求解 (P1) 的详细比较，请参见表 1.1。

4. 基于所提出的更快速预测-校正框架，我们在第 4.3 节给出一种用于求解 (P2) 的更快速 CP 型算法。与 [12, 25, 27] 中给出的非加速算法相比，该更快速 CP 型算法具有更快的收敛速度，例如在非遍历意义下达到 $O(1/t)$，并且在逐点意义下达到 $O(1/t^2)$。与采用条件 $rs > \rho(A^TA)$ 的加速算法 [1, 2, 26, 38, 39, 41] 相比，其收敛率是在条件$rs > 0.75\rho(A^TA)$ 下建立的。有关求解 (P2) 的详细比较，请参见表 1.2。

2. He 和 Yuan 的预测-校正框架。根据例 1 或 2 中定义的 $\theta(u)$、 u、 w、$F(w)$ 和 $T(w)$，求解 (P1') 和 (P2') （或 (P1) 和 (P2) ）的基本算法是近端点算法（ PPA ）。该算法最初由 Martinet [33], 提出，其形式如下:

PPA。给定 w^k，求满足以下条件的 w^{k+1} :

$$0 \in T(w^{k+1}) + Q(w^{k+1} - w^k), \tag{2.1}$$

其中， Q 是一个对称定义矩阵。

对于一般矩阵 Q，PPA 的子问题可能不易求解。因此，需要额外的算法来求解 PPA 的子问题。考虑到 (P1') 和 (P2') 的可分结构，我们可以采用 Jacobi 或 Gauss-Seidel 方式求解 (P1') 或 (P2') 中的每个块，以降低每次迭代的计算成本（例如交替方向乘子法）。

因此，我们考虑以下缩放矩阵经过特殊选取的预测-校正型近端点算法。需要指出的是，通过特殊选取缩放矩阵，许多著名算法都属于以下预测-校正框架，包括增广拉格朗日法（ALM） [23, 35], 近端增广拉格朗日法 [36, 37],交替方向乘子法、线性交替方向乘子法 [43], 以及严格压缩的 Peaceman-Rachford 分裂法 [11, 13], 用于求解（P1）多块问题的雅可比增广拉格朗日法 [10, 42], 带代入的交替方向乘子法 [16, 17] 以及 Chambolle-Pock（CP）算法及其变体

[4, 5, 12, 14]。

[预测步。] 给定 w^k，求出 $\widetilde{w}^k$，使得

$$0 \in T(\widetilde{w}^k) + L^T Q L(\widetilde{w}^k - w^k), \tag{PS}$$

其中 $Q^T + Q \succ 0$（注意，Q 不一定是对称的），L 是一个矩阵。[校正步。] 通过以下方式更新 $v^{k+1} = Lw^{k+1}$

$$v^{k+1} = v^k - M(v^k - \widetilde{v}^k), \tag{CS}$$

其中 $\widetilde{v}^k = L\widetilde{w}^k$。

REMARK 1. 满足 (PS)-(CS) 的三种算法见第 2.1-2.3 节。不同算法中 L、Q 和 M 的定义也一并给出。

那么存在 $\widetilde{g}^k \in (T - F)(\widetilde{w}^k)$ 使得

$$\begin{aligned} &(w - \widetilde{w}^k)^T(\widetilde{g}^k + F(\widetilde{w}^k) + L^T Q L(\widetilde{w}^k - w^k) = 0,\ \forall w \\ \Longrightarrow &\theta(u) - \theta(\widetilde{u}^k) + (w - \widetilde{w}^k)^T F(\widetilde{w}^k) \geq (w - \widetilde{w}^k)^T L^T Q L(w^k - \widetilde{w}^k),\ \forall w, \end{aligned} \tag{2.2}$$

其中，该不等式利用了 θ 的凸性。因此，He 和 Yuan[19] 提出的预测-校正框架 (PS)-(CS) 导出了以下框架（变分形式）。该框架对于构造求解 (P1) 和 (P2) 的收敛算法至关重要。有关详细说明，也可参阅 [21] 。

[预测步。] 给定 w^k，求 $\widetilde{w}^k$，使得

$$\theta(u) - \theta(\widetilde{u}^k) + (w - \widetilde{w}^k)^T F(\widetilde{w}^k) \geq (w - \widetilde{w}^k)^T L^T Q L(w^k - \widetilde{w}^k),\ \forall w, \tag{2.3}$$

其中 $Q^T + Q \succ 0$（注意，Q 不一定是对称的），且 L 是一个矩阵。[校正步。] 通过以下方式更新 $v^{k+1} = Lw^{k+1}$

$$v^{k+1} = v^k - M(v^k - \widetilde{v}^k), \tag{2.4}$$

其中 $\widetilde{v}^k = L\widetilde{w}^k$。

为确保满足上述框架的算法能够收敛，He 和 Yuan [21, 19] 对矩阵 Q 和 M 的选取增加了一些假设。以下是一个常用条件。

[收敛条件。] 对于给定的矩阵 Q 和非奇异矩阵M，令

$$H := QM^{-1} \succ 0, \tag{CC1}$$

$$G := Q^T + Q - M^T H M \succ 0. \tag{CC2}$$

预测-校正框架 (2.3)-(2.4) 的收敛速率如下所示。

THEOREM 2.1 ($O(1/t)$ 遍历收敛率). ([19, 21]) 设 $\{\widetilde{w}^k\}$ 由条件 (CC1)-(CC2) 下的预测-校正框架 (2.3)-(2.4) 生成，则有

$t = 1, 2, \ldots,$

$$\theta(\bar{u}^t) - \theta(u) + (\bar{w}^t - w)^T F(w) \leq \frac{1}{2(t+1)} \|v^0 - Lw\|_H^2, \ \forall w,$$

其中 $\bar{w}^t = \frac{1}{t+1} \sum_{k=0}^{t} \widetilde{w}^k$.

THEOREM 2.2 ($O(1/t)$ 的逐点意义). ([11, 12, 13, 15, 20, 21]) 设$\{\widetilde{w}^k\}$ 由条件 (CC1)-(CC2) 下的预测-校正框架 (2.3)-(2.4) 生成。则有

$$\|M(v^t - \widetilde{v}^t)\|_H^2 \leq O(1/t), \ t = 1, 2, \ldots.$$

下面给出一些满足 (PS)-(CS) 的算法，用于求解 (P1) 和 (P2)，以便更好地理解预测-校正框架 (PS)-(CS)。

2.1. 满足 (PS)-(CS) 的算法：求解带有 $m = 2$ 的 (P1)。考虑带有 $m = 2$ 的 (P1)，即,

$$\min_{x_1, x_2} \{f_1(x_1) + f_2(x_2) : \ A_1 x_1 + A_2 x_2 = b\}. \tag{2.5}$$

用于求解 (2.5) 的严格压缩 Peaceman-Rachford 分裂方法 [11, 13] 如下:

$$\begin{cases} x_1^{k+1} \in \arg\min_{x_1} \left\{ f_1(x_1) - x_1^T A_1^T \lambda^k + \frac{\beta}{2} \|A_1 x_1 + A_2 x_2^k - b\|^2 + \frac{1}{2} \|x_1 - x_1^k\|_P^2 \right\}, \\ \lambda^{k+\frac{1}{2}} = \lambda^k - r\beta(A_1 x_1^{k+1} + A_2 x_2^k - b), \\ x_2^{k+1} \in \arg\min_{x_2} \left\{ f_2(x_2) - x_2^T A_2^T \lambda^{k+\frac{1}{2}} + \frac{\beta}{2} \|A_1 x_1^{k+1} + A_2 x_2 - b\|^2 \right\}, \\ \lambda^{k+1} = \lambda^{k+\frac{1}{2}} - s\beta(A_1 x_1^{k+1} + A_2 x_2^{k+1} - b), \ P \succ 0, \ \beta > 0. \end{cases} \tag{2.6}$$

REMARK 2. 当 $r = 0$ 和 $P = 0$时，算法（2.6）退化为交替方向乘子法。与交替方向乘子法不同，算法（2.6）会两次更新 λ，并且在（2.8）的弱条件下即可证明其收敛性。如 *[11, 13],* 所示，与交替方向乘子法相比，算法（2.6）具有更好的数值性能。当 $r = s = 1$ 和$\alpha = 0$ 时，用于求解（2.5）对偶问题的算法（2.6）退化为 Peaceman-Rachford 分裂法（PRSM）。然而，PRSM 的收敛性是在强凸假设下证明的。为去除强凸假设，He 等人 *[11, 13]*对 r 和 s 提出了一些条件，以确保收敛（或参见（2.8））。

我们可以将 $P = \alpha I_{n_1} - \beta A_1^T A_1$ 设为 $\alpha > \|A_1\|^2$，以便子问题易于求解。

为方便起见，我们定义

$$L = \begin{pmatrix} I_{n_1} & 0 & 0 \\ 0 & I_{n_2} & 0 \\ 0 & 0 & I_l \end{pmatrix}, \ Q = \begin{pmatrix} P & 0 & 0 \\ 0 & \beta A_2^T A_2 & -r A_2^T \\ 0 & -A_2 & \frac{1}{\beta} I_l \end{pmatrix}, \ M = \begin{pmatrix} I_{n_1} & 0 & 0 \\ 0 & I_{n_2} & 0 \\ 0 & -s\beta A_2 & (r+s) I_l \end{pmatrix}. \tag{2.7}$$

THEOREM 2.3. 对于在（2.7）中定义的 L、 Q、 M，以及在例 1 中定义的 $\theta(u)$ u、w、 $F(w)$ $T(w)$，其中 $m = 2$，有:

(1) 算法（2.6）满足 (PS)-(CS)。(2) 若 A_2 满列秩，则满足 (CC1)-(CC2) 的 H 和 G 为正定,

$$r \in (-1, 1), \ s \in (0, 1) \text{ and } r + s > 0. \tag{2.8}$$

证明。令

$$\widetilde{x}_1^k := x_1^{k+1},\ \widetilde{x}_2^k := x_2^{k+1} \text{ and } \widetilde{\lambda}^k := \lambda^k - \beta(A_1\widetilde{x}_1^k + A_2x_2^k - b)$$

在算法（2.6）中。
证明（1）。 x_1-子问题的最优性条件如下:

$$\begin{aligned} 0 &\in \partial f_1(\widetilde{x}_1^k) - A_1^T[\lambda^k - \beta(A_1\widetilde{x}_1^k + A_2x_2^k - b)] + \alpha(\widetilde{x}_1^k - x_1^k) \\ &= \partial f_1(\widetilde{x}_1^k) - A_1^T\widetilde{\lambda}^k + P(\widetilde{x}_1^k - x_1^k). \end{aligned}$$

x_2-子问题的最优性条件如下:

$$\begin{aligned} 0 &\in \partial f_2(\widetilde{x}_2^k) - A_2^T[\lambda^k - r\beta(A_1\widetilde{x}_1^k + A_2x_2^k - b)] + \beta A_2^T[A_1\widetilde{x}_1^k + A_2\widetilde{x}_2^k - b] \\ &= \partial f_2(\widetilde{x}_2^k) - A_2^T\widetilde{\lambda}^k + \beta A_2^TA_2(\widetilde{x}_2^k - x_2^k) - rA_2^T(\widetilde{\lambda}^k - \lambda^k). \end{aligned}$$

$\widetilde{\lambda}^k$ 的定义可改写为:

$$0 = A_1\widetilde{x}_1^k + A_2\widetilde{x}_2^k - b - A_2(\widetilde{x}_2^k - x_2^k) + \frac{1}{\beta}(\widetilde{\lambda}^k - \lambda^k).$$

结合上述三个关系，我们得到

$$0 \in \begin{pmatrix} \partial f_1(\widetilde{x}_1^k) \\ \partial f_2(\widetilde{x}_2^k) \\ 0 \end{pmatrix} + \begin{pmatrix} -A_1^T\widetilde{\lambda}^k \\ -A_2^T\widetilde{\lambda}^k \\ A_1\widetilde{x}_1^k + A_2\widetilde{x}_2^k - b \end{pmatrix} + \begin{pmatrix} P(\widetilde{x}_1^k - x_1^k) \\ \beta A_2^TA_2(\widetilde{x}_2^k - x_2^k) - rA_2^T(\widetilde{\lambda}^k - \lambda^k) \\ -A_2(\widetilde{x}_2^k - x_2^k) + \frac{1}{\beta}(\widetilde{\lambda}^k - \lambda^k) \end{pmatrix},$$

这等价于

$$0 \in \begin{pmatrix} \partial f_1(\widetilde{x}_1^k) \\ \partial f_2(\widetilde{x}_2^k) \\ 0 \end{pmatrix} + \begin{pmatrix} -A_1^T\widetilde{\lambda}^k \\ -A_2^T\widetilde{\lambda}^k \\ A_1\widetilde{x}_1^k + A_2\widetilde{x}_2^k - b \end{pmatrix} + \begin{pmatrix} P & 0 & 0 \\ 0 & \beta A_2^TA_2 & -rA_2^T \\ 0 & -A_2 & \frac{1}{\beta}I_l \end{pmatrix} \begin{pmatrix} x_1^k - \widetilde{x}_1^k \\ x_2^k - \widetilde{x}_2^k \\ \lambda^k - \widetilde{\lambda}^k \end{pmatrix}.$$

于是，预测步成立。可得

$$\begin{aligned} \lambda^{k+1} =& \lambda^{k+\frac{1}{2}} - s\beta(A_1\widetilde{x}_1^k + A_2\widetilde{x}_2^k - b) \\ =& \lambda^k - r\beta(A_1\widetilde{x}_1^k + A_2x_2^k - b) - s\beta(A_1\widetilde{x}_1^k + A_2x_2^k - b) + s\beta A_2(x_2^k - \widetilde{x}_2^k) \\ =& \lambda^k - r(\lambda^k - \widetilde{\lambda}^k) - s(\lambda^k - \widetilde{\lambda}^k) + s\beta A_2(x_2^k - \widetilde{x}_2^k). \end{aligned} \tag{2.9}$$

因此，结合 $\widetilde{x}_1^k = x_1^{k+1}$和 $\widetilde{x}_2^k = x_2^{k+1}$，我们有

$$\begin{pmatrix} x_1^{k+1} \\ x_2^{k+1} \\ \lambda^{k+1} \end{pmatrix} = \begin{pmatrix} x_1^k \\ x_2^k \\ \lambda^k \end{pmatrix} - \begin{pmatrix} I_{n_1} & 0 & 0 \\ 0 & I_{n_2} & 0 \\ 0 & -s\beta A_2 & (r+s)I_l \end{pmatrix} \begin{pmatrix} x_1^k - \widetilde{x}_1^k \\ x_2^k - \widetilde{x}_2^k \\ \lambda^k - \widetilde{\lambda}^k \end{pmatrix}.$$

校正步成立。
式 (2) 的证明。关于满足 (CC1)-(CC2) 的 H 和 G 的正定性，可参见 [13, 引理 4.1] 和 [13, 第 1480] 页中的类似证明。 □

2.2. 满足 (PS)-(CS) 的求解 (P1) 的算法。以下求解 (P1) 的算法首先由 He 等人提出 [18]。

[预测步。] With 给定 $\beta > 0$ and $(x_1^k, x_2^k, \ldots, x_m^k, \lambda^k)$, find $(\widetilde{x}_1^k, \ldots, \widetilde{x}_m^k, \widetilde{\lambda}^k)$ 通过

(2.10)
$$\begin{cases} \widetilde{x}_1^k = \arg\min_{x_1} \left\{ f_1(x_1) - x_1^T A_1^T \lambda^k + \frac{\beta}{2} \|A_1(x_1 - x_1^k)\|^2 \right\}, \\ \widetilde{x}_2^k = \arg\min_{x_2} \left\{ f_2(x_2) - x_2^T A_2^T \lambda^k + \frac{\beta}{2} \|A_1(\widetilde{x}_1^k - x_1^k) + A_2(x_2 - x_2^k)\|^2 \right\}, \\ \vdots \\ \widetilde{x}_i^k = \text{最小自变量}_{x_i} \left\{ f_i(x_i) - x_i^T A_2^T \lambda^k + \frac{\beta}{2} \|\sum_{j=1}^{i-1} A_j(\widetilde{x}_j^k - x_j^k) + A_i(x_i - x_i^k)\|^2 \right\} \\ \vdots \\ \widetilde{x}_m^k = \arg\min_{x_m} \left\{ f_m(x_m) - x_m^T A_2^T \lambda^k + \frac{\beta}{2} \|\sum_{j=1}^{m-1} A_j(\widetilde{x}_j^k - x_j^k) + A_m(x_m - x_m^k)\|^2 \right\} \\ \widetilde{\lambda}^k = \text{最大化}_{\lambda} \left\{ -\lambda^T (\sum_{j=1}^{m} A_j \widetilde{x}_j^k - b) - \frac{1}{2\beta} \|\lambda - \lambda^k\|^2 \right\}. \end{cases}$$

[校正步。]更新 $(A_1 x_1^{k+1}, A_2 x_2^{k+1}, \ldots, A_m x_m^{k+1}, \lambda^{k+1})$ 通过 (2.11)

$$\begin{pmatrix} \sqrt{\beta} A_1 x_1^{k+1} \\ \sqrt{\beta} A_2 x_2^{k+1} \\ \vdots \\ \sqrt{\beta} A_m x_m^{k+1} \\ \frac{1}{\sqrt{\beta}} \lambda^{k+1} \end{pmatrix} = \begin{pmatrix} \sqrt{\beta} A_1 x_1^k \\ \sqrt{\beta} A_2 x_2^k \\ \vdots \\ \sqrt{\beta} A_m x_m^k \\ \frac{1}{\sqrt{\beta}} \lambda^k \end{pmatrix} - \begin{pmatrix} \alpha I_l & -\alpha I_l & 0 & \ldots & 0 \\ 0 & \alpha I_l & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & -\alpha I_l & 0 \\ 0 & \ldots & 0 & \alpha I_l & 0 \\ -\alpha I_l & 0 & \ldots & 0 & I_l \end{pmatrix} \begin{pmatrix} \sqrt{\beta}(A_1 x_1^k - A_1 \widetilde{x}_1^k) \\ \sqrt{\beta}(A_2 x_2^k - A_2 \widetilde{x}^k \\ 2) \ldots \sqrt{\beta}(\\ A_m x_m^k - A_m \widetilde{x}_m^k) \frac{1}{\sqrt{\beta}}(\\ \lambda^k - \widetilde{\lambda}^k) \end{pmatrix}.$$

为方便起见，我们定义

$$L = \text{Diag}\left(\sqrt{\beta} A_1, \quad \sqrt{\beta} A_2, \quad \ldots, \sqrt{\beta} A_m, \quad \frac{1}{\sqrt{\beta}} I_l \right),$$

(2.12)
$$Q = \begin{pmatrix} I_l & 0 & \ldots & 0 & I_l \\ I_l & I_l & \ddots & \vdots & I_l \\ \vdots & \vdots & \ddots & 0 & \vdots \\ I_l & I_l & \ldots & I_l & I_l \\ 0 & 0 & \ldots & 0 & I_l \end{pmatrix}, \; M = \begin{pmatrix} \alpha I_l & -\alpha I_l & 0 & \ldots & 0 \\ 0 & \alpha I_l & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & -\alpha I_l & 0 \\ 0 & \ldots & 0 & \alpha I_l & 0 \\ -\alpha I_l & 0 & \ldots & 0 & I_l \end{pmatrix}.$$

对于在 (2.12) 中定义的 L、Q、M 和 $\theta(u)$，以及在例 1 中定义的 u、w、$F(w)$ $T(w)$，有

(1) 算法 (2.10)-(2.11) 满足 (PS)-(CS)。(2) 满足 (CC1)-(CC2) 的 H 和 G 在 $\alpha \in$ (0,1) 时是正定的。

证明。 (1) 的证明。对于 $i = 1, 2, \ldots, m$, x_i- 子问题的最优性条件为

$$\begin{aligned} 0 &\in \partial f_i(\widetilde{x}_i^k) - A_i^T \lambda^k + \beta A_i^T \sum_{j=1}^{i} A_j(\widetilde{x}_j^k - x_j^k) \\ &= \partial f_i(\widetilde{x}_i^k) - A_i^T \widetilde{\lambda}^k + \beta A_i^T \sum_{j=1}^{i} A_j(\widetilde{x}_j^k - x_j^k) + A_i^T(\widetilde{\lambda}^k - \lambda^k). \end{aligned}$$

λ-子问题的最优性条件为

$$0 = (\sum_{j=1}^{m} A_j \widetilde{x}_j^k - b) + \frac{1}{\beta}(\widetilde{\lambda}^k - \lambda^k).$$

结合上述两个关系，我们得到

$$0 \in \begin{pmatrix} \partial f_1(\widetilde{x}_1^k) \\ \partial f_2(\widetilde{x}_2^k) \\ \vdots \\ \partial f_m(\widetilde{x}_m^k) \\ 0 \end{pmatrix} + \begin{pmatrix} -A_1^T\widetilde{\lambda}^k \\ -A_2^T\widetilde{\lambda}^k \\ \vdots \\ -A_m^T\widetilde{\lambda}^k \\ \sum_{j=1}^m A_j\widetilde{x}_j^k - b \end{pmatrix} + \begin{pmatrix} \beta A_1^T A_1(\widetilde{x}_1^k - x_1^k) + A_1^T(\widetilde{\lambda}^k - \lambda^k) \\ \beta A_2^T \sum_{j=1}^2 A_j(\widetilde{x}_j^k - x_j^k) + A_2^T(\widetilde{\lambda}^k - \lambda^k) \\ \vdots \\ \beta A_m^T \sum_{j=1}^m A_j(\widetilde{x}_j^k - x_j^k) + A_m^T(\widetilde{\lambda}^k - \lambda^k) \\ \frac{1}{\beta}(\widetilde{\lambda}^k - \lambda^k) \end{pmatrix},$$

这等价于

$$0 \in \begin{pmatrix} \partial f_1(\widetilde{x}_1^k) \\ \partial f_2(\widetilde{x}_2^k) \\ \vdots \\ \partial f_m(\widetilde{x}_m^k) \\ 0 \end{pmatrix} + \begin{pmatrix} -A_1^T\widetilde{\lambda}^k \\ -A_2^T\widetilde{\lambda}^k \\ \vdots \\ -A_m^T\widetilde{\lambda}^k \\ \sum_{j=1}^m A_j\widetilde{x}_j^k - b \end{pmatrix} + L^T \begin{pmatrix} I_l & 0 & \dots & 0 & I_l \\ I_l & I_l & \ddots & \vdots & I_l \\ \vdots & \vdots & \ddots & 0 & \vdots \\ I_l & I_l & \dots & I_l & I_l \\ 0 & 0 & \dots & 0 & I_l \end{pmatrix} L \begin{pmatrix} \widetilde{x}_1^k - x_1^k \\ \widetilde{x}_2^k - x_2^k \\ \vdots \\ \widetilde{x}_m^k - x_m^k \\ \widetilde{\lambda}^k - \lambda^k \end{pmatrix}.$$

随后，预测步成立。校正步很容易验证。(2) 的证明。我们可以参见 [18, 引理 7.1、7.2]。 □

2.3. 用于求解 (P2) 且满足 (PS)-(CS) 的算法。He 等人提出了以下用于求解 (P2) 的算法 [12]：

[预测步。] 给定 $\beta > 0$ 和 (x^k, y^k)，求 $(\widetilde{x}^k, \widetilde{y}^k)$，使其满足

$$\begin{cases} \widetilde{x}^k = \arg\min\limits_x \left\{ \Phi(x, y^k) + \frac{r}{2}\|x - x^k\|^2 \right\}, \\ \widetilde{y}^k = \arg\max\limits_y \left\{ \Phi([\widetilde{x}^k + \alpha(\widetilde{x}^k - x^k)], y) - \frac{s}{2}\|y - y^k\|^2 \right\}. \end{cases} \tag{2.13}$$

[校正步。] 通过以下方式更新 (x^{k+1}, y^{k+1})

$$\begin{pmatrix} x^{k+1} \\ y^{k+1} \end{pmatrix} = \begin{pmatrix} x^k \\ y^k \end{pmatrix} - \begin{pmatrix} I_n & 0 \\ -(1-\alpha)\frac{1}{s}A & I_m \end{pmatrix} \begin{pmatrix} x^k - \widetilde{x}^k \\ y^k - \widetilde{y}^k \end{pmatrix}. \tag{2.14}$$

为方便起见，我们定义

$$L = \begin{pmatrix} I_n & 0 \\ 0 & I_m \end{pmatrix}, \quad Q = \begin{pmatrix} rI_n & A^T \\ \alpha A & sI_m \end{pmatrix} \text{ and } M = \begin{pmatrix} I_n & 0 \\ -(1-\alpha)\frac{1}{s}A & I_m \end{pmatrix}. \tag{2.15}$$

THEOREM 2.5. 对于(2.15)中定义的 L、 Q、 M以及例 2中定义的 $\theta(u)$ u、 w、 $F(w)$ $T(w)$，成立

(1) 算法(2.13)-(2.14)满足(PS)-(CS)。(2) 满足(CC1)-(CC2)的 H 和 G 是正定的，如果

$$rs > (1 - \alpha + \alpha^2)\rho(A^T A), \ \alpha \in [0, 1].$$

证明。(1)的证明。 x-子问题的最优性条件如下：

$$\begin{aligned} 0 &\in \partial f(\widetilde{x}^k) - A^T y^k + r(\widetilde{x}^k - x^k) \\ &= \partial f(\widetilde{x}^k) - A^T\widetilde{y}^k + r(\widetilde{x}^k - x^k) + A^T(\widetilde{y}^k - y^k). \end{aligned}$$

y-子问题的最优性条件为：

$$0 \in \partial g(\widetilde{y}^k) + A[\widetilde{x}^k + \alpha(\widetilde{x}^k - x^k)] + s(\widetilde{y}^k - y^k).$$

将上述两个关系结合起来可得

$$0 \in \begin{pmatrix} \partial f(\widetilde{x}^k) \\ \partial g(\widetilde{y}^k) \end{pmatrix} + \begin{pmatrix} -A^T\widetilde{y}^k \\ A\widetilde{x}^k \end{pmatrix} + \begin{pmatrix} r(\widetilde{x}^k - x^k) + A^T(\widetilde{y}^k - y^k) \\ \alpha A(\widetilde{x}^k - x^k) + s(\widetilde{y}^k - y^k) \end{pmatrix}.$$

这等价于:

$$0 \in \begin{pmatrix} \partial f(\widetilde{x}^k) \\ \partial g(\widetilde{y}^k) \end{pmatrix} + \begin{pmatrix} -A^T\widetilde{y}^k \\ A\widetilde{x}^k \end{pmatrix} + \begin{pmatrix} rI_n & A^T \\ \alpha A & sI_m \end{pmatrix} \begin{pmatrix} \widetilde{x}^k - x^k \\ \widetilde{y}^k - y^k \end{pmatrix}.$$

然后我们得到预测步 (4.11) 满足 (FPS)，其中 L 和 Q 在 (2.15) 中定义。校正步很容易验证。
证明 (2)。我们可以参考 [12, 命题 4.1]。 □

REMARK 3. 当 $\alpha = 1$ 时，算法 (2.13)-(2.14) 简化为 *[4]* 中引入的 CP 算法。在条件 $rs > \rho(A^TA)$ 下建立了收敛性。通过在算法 (2.13)-(2.14) 中设置 $\alpha = \frac{1}{2}$，收敛条件放宽至 $rs > 0.75\rho(A^TA)$。与原始 CP 算法相比，这一扩展拓宽了允许的步长。

3. 快速预测-校正框架。我们首先给出以下新的预测-校正框架，其中 $\theta(u)$、 u、 w、 $F(w)$ 和 $T(w)$ 在例 1 或 2 中定义，然后建立收敛速率。

[预测步。] 给定 w^k 和 $\breve{w}^{k-1}$，求满足下式的 $\breve{w}^k$

(FPS) $$0 \in (T - F)(\breve{w}^k) + F(\widetilde{w}^k) + L^TQL(\widetilde{w}^k - w^k),$$

其中 $Q^T + Q \succ 0$ （注意， Q 不一定是对称的）， L 是一个矩阵，并且

(3.1) $$\widetilde{w}^k = \frac{1}{\tau^k}\breve{w}^k - \frac{1-\tau^k}{\tau^k}\breve{w}^{k-1},$$

且序列 $\{\tau^k\}$ 满足以下等式:

(Y) $$1/\tau^{k-1} = (1-\tau^k)/\tau^k,\ \tau^{-1} \in (0,1).$$

[校正步。] 按下式更新 $v^{k+1} = Lw^{k+1}$

(FCS) $$v^{k+1} = v^k - M(v^k - \widetilde{v}^k),$$

其中 $\widetilde{v}^k = L\widetilde{w}^k$。

根据 (FPS)，存在 $\breve{g}^k \in (T - F)(\breve{w}^k)$ 使得

(3.2) $$\begin{aligned} &(w - \breve{w}^k)^T(\breve{g}^k + F(\widetilde{w}^k) + L^TQL(\widetilde{w}^k - w^k) = 0,\ \forall w \\ \Longrightarrow &\theta(u) - \theta(\breve{u}^k) + (w - \breve{w}^k)^TF(\widetilde{w}^k) \geq (w - \breve{w}^k)^TL^TQL(w^k - \widetilde{w}^k),\ \forall w, \end{aligned}$$

其中，该不等式利用了 θ 的凸性。因此，预测-校正框架（FPS）-（FCS）推导出以下框架（变分形式）。

[预测步。] 给定 w^k 和 $\breve{w}^{k-1}$，求 $\breve{w}^k$，使得

$$\theta(u)-\theta(\breve{u}^k)+(w-\breve{w}^k)^TF(\widetilde{w}^k)\geq(w-\breve{w}^k)^TL^TQL(w^k-\widetilde{w}^k),\ \forall w, \tag{3.3}$$

其中 $Q^T+Q\succ 0$（注意，Q 不一定是对称的），L 是一个矩阵，$\widetilde{w}^k$ 在（3.1）中定义，且 τ^k 满足（Y）。

[校正步。] 按以下方式更新 $v^{k+1}=Lw^{k+1}$

$$v^{k+1}=v^k-M(v^k-\widetilde{v}^k), \tag{3.4}$$

其中 $\widetilde{v}^k=L\widetilde{w}^k$。

序列 $\{\tau^k\}$ 满足（Y）的以下性质很容易验证，因此省略证明。

LEMMA 3.1. 设 $\{\tau^k\}$ 满足（Y）。则 $\tau^k=\frac{1}{\tau^{-1}+k+1}$。

3.1. $O(1/t)$ 非遍历收敛率。我们建立 $O(1/t)$ 快速预测-校正框架（FPS）-（FCS）在条件 (CC1)-(CC2) 下原始-对偶间隙的非遍历收敛率。

LEMMA 3.2. 对于在 (CC1)-(CC2) 下的快速预测-校正框架（FPS）-（FCS），我们有

$$\begin{aligned}&\frac{1}{\tau^k}[\theta(u)-\theta(\breve{u}^k)]-\frac{1}{\tau^{k-1}}[\theta(u)-\theta(\breve{u}^{k-1})]+(w-\widetilde{w}^k)^TF(\widetilde{w}^k)\\ \geq&\frac{1}{2}\left(\|v^{k+1}-v\|_H^2-\|v^k-v\|_H^2+\|v^k-\widetilde{v}^k\|_G^2\right),\ \forall w,\end{aligned} \tag{3.5}$$

其中 $v=Lw$。

证明。预测步 (FPS) 意味着预测步 (3.3) 成立。将 (3.3) 两边同时乘以 $(1-\tau^k)/\tau^k$（其中 $u=\breve{u}^{k-1}$ 和 $w=\breve{w}^{k-1}$），然后将其加到 (3.3) 上，得到

$$\begin{aligned}&\frac{1}{\tau^k}[\theta(u)-\theta(\breve{u}^k)]-\frac{1-\tau^k}{\tau^k}[\theta(u)-\theta(\breve{u}^{k-1})]+(w-\widetilde{w}^k)^TF(\widetilde{w}^k)\\ \overset{(Y)}{=}&\frac{1}{\tau^k}[\theta(u)-\theta(\breve{u}^k)]-\frac{1}{\tau^{k-1}}[\theta(u)-\theta(\breve{u}^{k-1})]+(w-\widetilde{w}^k)^TF(\widetilde{w}^k)\\ \geq&(w-\widetilde{w}^k)^TL^TQL(w^k-\widetilde{w}^k)=(v-\widetilde{v}^k)^TQ(v^k-\widetilde{v}^k).\end{aligned} \tag{3.6}$$

另一方面，我们可以验证

$$\begin{aligned}&(v-\widetilde{v}^k)^TQ(v^k-\widetilde{v}^k)\overset{(FCS)}{=}(v-\widetilde{v}^k)^TH(v^k-v^{k+1})\\ =&\frac{1}{2}\left(\|v^{k+1}-v\|_H^2-\|v^k-v\|_H^2+\|v^k-\widetilde{v}^k\|_H^2-\|v^{k+1}-\widetilde{v}^k\|_H^2\right)\end{aligned} \tag{3.7}$$

and

$$\begin{aligned}&\|v^k-\widetilde{v}^k\|_H^2-\|v^{k+1}-\widetilde{v}^k\|_H^2\\ =&\|v^k-\widetilde{v}^k\|_H^2-\|(v^k-\widetilde{v}^k)-(v^k-v^{k+1})\|_H^2\\ \overset{(FCS)}{=}&\|v^k-\widetilde{v}^k\|_H^2-\|(v^k-\widetilde{v}^k)-M(v^k-\widetilde{v}^k)\|_H^2\\ =&(v^k-\widetilde{v}^k)^T(2HM-M^THM)(v^k-\widetilde{v}^k)\\ =&(v^k-\widetilde{v}^k)^T(Q^T+Q-M^THM)(v^k-\widetilde{v}^k)\\ =&\|v^k-\widetilde{v}^k\|_G^2.\end{aligned} \tag{3.8}$$

结合 (3.6)、(3.7) 和 (3.8)，证明即告完成。 □

LEMMA 3.3. 对于快速预测-校正框架（FPS）-（FCS）und er (CC1)-(CC2)，我们有

$$\begin{aligned}&\frac{1}{\tau^k}[\theta(u)-\theta(\breve{u}^k)+(w-\breve{w}^k)^TF(w)]\\&-\frac{1}{\tau^{k-1}}[\theta(u)-\theta(\breve{u}^{k-1})+(w-\breve{w}^{k-1})^TF(w)]\\\geq&\frac{1}{2}\left(\|v^{k+1}-v\|_H^2-\|v^k-v\|_H^2+\|v^k-\widetilde{v}^k\|_G^2\right),\ \forall w,\end{aligned}\tag{3.9}$$

其中 $v=Lw$。

证明。根据例 1 和例 2 中 $F(w)$ 的定义，我们有

$$(w-w')^T(F(w)-F(w'))=0,\ \forall w,w'.$$

于是得到

$$\begin{aligned}&(w-\widetilde{w}^k)^TF(\widetilde{w}^k)=(w-\widetilde{w}^k)^TF(w)\\=&\frac{1}{\tau^k}(w-\breve{w}^k)^TF(w)-\frac{1-\tau^k}{\tau^k}(w-\breve{w}^{k-1})^TF(w)\\\overset{(Y)}{=}&\frac{1}{\tau^k}(w-\breve{w}^k)^TF(w)-\frac{1}{\tau^{k-1}}(w-\breve{w}^{k-1})^TF(w).\end{aligned}\tag{3.10}$$

结合引理 3.2，我们得到结论。 □

然后我们立即得到以下结论。

THEOREM 3.4 ($O(1/t)$ 非遍历收敛率). 设 $\{\breve{w}^t\}$ 由 b 生成 y （FPS）-（FCS）在 (CC1)-(CC2) 下。然后我们有

$$\begin{aligned}&\theta(\breve{u}^t)-\theta(u)-(w-\breve{w}^t)^TF(w)\\\leq&\frac{1}{\tau^{-1}+t+1}\left\{\frac{1}{\tau^0}[\theta(\breve{u}^0)-\theta(u)-(w-\breve{w}^0)^TF(w)]+\frac{1}{2}\|v^1-v\|_H^2\right\},\ t=1,2,\ldots,\ \forall w,\end{aligned}$$

其中 $v=Lw$。

证明。将引理 3.3 从 $k=1$ 加到 $k=t$，然后

$$\begin{aligned}&\frac{1}{\tau^t}[\theta(\breve{u}^t)-\theta(u)-(w-\breve{w}^t)^TF(w)]+\frac{1}{2}\|v^{t+1}-v\|_H^2\\\leq&\frac{1}{\tau^0}[\theta(\breve{u}^0)-\theta(u)-(w-\breve{w}^0)^TF(w)]+\frac{1}{2}\|v^1-v\|_H^2,\ t=1,2,\ldots,\ \forall w.\end{aligned}$$

根据引理 3.1，我们得出结论。 □

3.2. $O(1/t^2)$逐点意义下的收敛率。我们在条件 (CC1)-(CC2) 下，建立了快速预测-校正框架（FPS）-（FCS）的 $O(1/t^2)$逐点意义下的收敛率。

THEOREM 3.5 ($O(1/t^2)$ 在逐点意义下). 对于满足（CC1）-（CC2）的快速预测-校正框架（FPS）-（FCS），我们有

$$\|M(\breve{v}^t-\breve{v}^{t-1})\|_H^2\leq O(1/t^2),\ t=1,2,\ldots.$$

证明。设 w^* 是（P1）或（P2）的鞍点，且 $v^* = Lw^*$。根据校正步（FCS），我们有

$$\begin{aligned}\|v^{k+1}-v^*\|_H^2 =& \|(I-M)(v^k-v^*)+M(\widetilde{v}^k-v^*)\|_H^2\\ =& \underbrace{\|(I-M)(v^k-v^*)\|_H^2}_{:=A^k}+\underbrace{\|M(\widetilde{v}^k-v^*)\|_H^2}_{:=B^k}\\ &+2\underbrace{(v^k-v^*)^T(I-M)^THM(\widetilde{v}^k-v^*)}_{:=C^k}.\end{aligned} \tag{3.11}$$

根据 $\widetilde{v}^k$ 的定义，我们得到

$$\begin{aligned}&\|M(\widetilde{v}^k-v^*)\|_H^2+\frac{1-\tau^k}{\tau^k}\underbrace{\|M(\breve{v}^{k-1}-v^*)\|_H^2}_{:=D^{k-1}}\\ =&\frac{1-\tau^k}{(\tau^k)^2}\underbrace{\|M(\breve{v}^k-\breve{v}^{k-1})\|_H^2}_{:=E^k}+\frac{1}{\tau^k}\underbrace{\|M(\breve{v}^k-v^*)\|_H^2}_{D^k}.\end{aligned} \tag{3.12}$$

在引理 3.3 中令 $u=u^*$、 $w=w^*$ 和 $v=v^*$，由 (3.11) 和 (3.12) 可得

$$\begin{aligned}&\|v^{k+1}-v^*\|_H^2-\|v^k-v^*\|_H^2 = A^k-A^{k-1}+B^k-B^{k-1}+2(C^k-C^{k-1})\\ =&A^k-A^{k-1}+2(C^k-C^{k-1})+\frac{1-\tau^k}{(\tau^k)^2}E^k-\frac{1-\tau^{k-1}}{(\tau^{k-1})^2}E^{k-1}\\ &+\frac{1}{\tau^k}(D^k-D^{k-1})-\frac{1}{\tau^{k-1}}(D^{k-1}-D^{k-2})+D^{k-1}-D^{k-2}\\ \le&-\frac{2}{\tau^t}S^t+\frac{2}{\tau^{k-1}}S^{k-1},\end{aligned} \tag{3.13}$$

其中 $S^t=\theta(\breve{u}^t)-\theta(u^*)+(\breve{w}^t-w^*)^TF(w^*)\ge 0$. 对于每个 $t=1,2,\ldots$，将 (3.13) 两边从 $k=1$ 到 t 求和可得

$$\begin{aligned}&\sum_{k=1}^{t}\left(\|v^{k+1}-v^*\|_H^2-\|v^k-v^*\|_H^2\right)\\ =\;& A^t-A^0+2(C^t-C^0)+\frac{1-\tau^t}{(\tau^t)^2}E^t-\frac{1-\tau^0}{(\tau^0)^2}E^0\\ &+\frac{1}{\tau^t}(D^t-D^{t-1})-\frac{1}{\tau^0}(D^0-D^{-1})+D^{t-1}-D^{-1}\\ \le\;& -\frac{2}{\tau^t}S^t+\frac{2}{\tau^0}S^0,\end{aligned} \tag{3.14}$$

根据 (3.13)， v^k 有界。因此，校正步 (FCS) 表明 $\widetilde{v}^k$ 也有界。接着，根据其定义， C^t 有界。由 (3.14) 可知，存在一个正的上界 $N_1<\infty$，使得

$$\frac{2}{\tau^t}S^t+A^t+\frac{1-\tau^t}{(\tau^t)^2}E^t+\frac{1}{\tau^t}(D^t-D^{t-1})+D^{t-1}<N_1. \tag{3.15}$$

由于 S^t、 A^t、 N^t 和 E^t 均为非负数，由 (3.15) 可知

$$\frac{1}{\tau^t}(D^t-D^{t-1})+D^{t-1}\overset{(Y)}{=}\frac{1}{\tau^t}D^t-\frac{1}{\tau^{t-1}}D^{t-1}<N_1.$$

因此有

$$\frac{1}{\tau^t}D^t < tN_1 + \frac{1}{\tau^0}D^0.$$

因此，我们有

$$D^t \le \tau^t t N_1 + \frac{\tau^t}{\tau^0}D^0 < \infty,\ t \to \infty,$$

也就是说， D^t 有界，即，对于某个 $0 < N_2 < \infty.$ ， $D^t \le N_2$ 。根据柯西-施瓦茨不等式，我们有

$$\begin{aligned} D^{t-1} - D^t =& -(M(\breve{v}^t - v^*) + M(\breve{v}^{t-1} - v^*))^T H M(\breve{v}^t - \breve{v}^{t-1}) \\ \le& \|M(\breve{v}^t - v^*) + M(\breve{v}^{t-1} - v^*)\|_H \|M(\breve{v}^t - \breve{v}^{t-1})\|_H \\ \le& 2\sqrt{N_2}\sqrt{E^t}. \end{aligned} \tag{3.16}$$

结合 (3.15) 和 (3.16) 可得

$$\frac{1-\tau^0}{(\tau^t)^2}E^t \le \frac{1-\tau^t}{(\tau^t)^2}E^t \le N_1 + \frac{2}{\tau^t}\sqrt{N_2}\sqrt{E^t}. \tag{3.17}$$

令 $h^t := \sup_{t\ge 0}\sqrt{E^t}/\tau^t$。由 (3.17) 可得

$$(1-\tau^0)h^t \le \frac{N_1}{h^t} + 2\sqrt{N_2} < \infty,$$

即， $E^t \le O((\tau^t)^2)$，其中 $t \to \infty$。根据引理 3.1，证毕。 □

4. 应用。设计满足我们快速预测-校正框架（FPS）-（FCS）的算法的一种简单方法是将满足 (PS)-(CS) 的算法转换为满足（FPS）-（FCS）的更快版本。特别地，我们可以基于满足 (PS)-(CS) 的算法，如交替方向乘子法、线性交替方向乘子法 [43], 严格收缩的 Peaceman- Rachford 分裂方法 [11, 13], 雅可比增广拉格朗日法 [10, 42], 带替换的交替方向乘子法 [16, 17], ADMM 型算法 [18], CP型算法 [12] 具有更大步长等等，来构造更快算法以求解 (P1) 和 (P2) 的特殊情况。在本节中，我们举例说明三种这样的更快算法。

4.1. 满足（FPS）-（FCS）的更快算法，用于求解带有 $m = 2$和等式约束的 (P1)。我们设计一种满足（FPS）-（FCS）的更快算法来求解 (P1) $m = 2$。

我们考虑以下用于求解 (2.5) 的算法：

[**Prediction step.**] For given $P \succ 0$, $\beta > 0$, $(\breve{x}_1^{k-1}, \breve{x}_2^{k-1}, \breve{\lambda}^{k-1})$ and $(x_1^k, x_2^k, \lambda^k)$, find $(\breve{x}_1^k, \breve{x}_2^k, \breve{\lambda}^k)$ by

(4.1)

$$\begin{cases} \breve{x}_1^k \in \arg\min\limits_{x_1} \Big\{ f_1(x_1) - x_1^T A_1^T \lambda^k + \frac{\beta\tau^k}{2}\|A_1(\frac{1}{\tau^k}x_1 - \frac{1-\tau^k}{\tau^k}\breve{x}_1^{k-1}) + A_2 x_2^k - b\|^2 \\ \qquad\qquad + \frac{\tau^k}{2}\|\frac{1}{\tau^k}x_1 - \frac{1-\tau^k}{\tau^k}\breve{x}_1^{k-1} - x_1^k\|_P^2 \Big\}, \\ \breve{x}_2^k \in \arg\min\limits_{x_2} \big\{ f_2(x_2) - x_2^T A_2^T[\lambda^k - r\beta(A_1\widetilde{x}_1^k + A_2 x_2^k - b)] \\ \qquad\qquad + \frac{\beta\tau^k}{2}\|A_1\widetilde{x}_1^k + A_2(\frac{1}{\tau^k}x_2 - \frac{1-\tau^k}{\tau^k}\breve{x}_2^{k-1}) - b\|^2 \Big\}, \\ \breve{\lambda}^k = \arg\max\limits_{\lambda} \Big\{ -\frac{\tau^k}{2}\|(\frac{1}{\tau^k}\lambda - \frac{1-\tau^k}{\tau^k}\breve{\lambda}^{k-1}) - [\lambda^k - \beta(A_1\widetilde{x}_1^k + A_2 x_2^k - b)]\|^2 \Big\}, \end{cases}$$

其中 τ^k 满足 (Y)。
[校正步。] 按以下方式更新 $(x_1^{k+1}, x_2^{k+1}, \lambda^{k+1})$

$$\begin{pmatrix} x_1^{k+1} \\ x_2^{k+1} \\ \lambda^{k+1} \end{pmatrix} = \begin{pmatrix} x_1^k \\ x_2^k \\ \lambda^k \end{pmatrix} - \begin{pmatrix} I_{n_1} & 0 & 0 \\ 0 & I_{n_2} & 0 \\ 0 & -s\beta A_2 & (r+s)I_l \end{pmatrix} \begin{pmatrix} x_1^k - \widetilde{x}_1^k \\ x_2^k - \widetilde{x}_2^k \\ \lambda^k - \widetilde{\lambda}^k \end{pmatrix}, \tag{4.2}$$

其中

$$\begin{pmatrix} \widetilde{x}_1^k \\ \widetilde{x}_2^k \\ \widetilde{\lambda}^k \end{pmatrix} = \begin{pmatrix} \frac{1}{\tau^k}\breve{x}_1^k - \frac{1-\tau^k}{\tau^k}\breve{x}_1^{k-1} \\ \frac{1}{\tau^k}\breve{x}_2^k - \frac{1-\tau^k}{\tau^k}\breve{x}_2^{k-1} \\ \frac{1}{\tau^k}\breve{\lambda}^k - \frac{1-\tau^k}{\tau^k}\breve{\lambda}^{k-1} \end{pmatrix}.$$

下述定理说明，算法 (4.1)-(4.2) 满足（FPS）-（FCS）及收敛条件 (CC1)-(CC2)。因此，定理 3.4 和 3.5 可分别用于得到 $O(1/t)$ 在原始-对偶间隙的非遍历意义下的收敛速率，以及 $O(1/t^2)$ 在逐点意义下的收敛速率。

THEOREM 4.1. 对于由 (2.7) 定义的 L, Q, M，以及例 1 中在 $m = 2$ 条件下定义的 $\theta(u)$, u, w, $F(w)$ $T(w)$，有

(1) 算法 (4.1)-(4.2) 满足（FPS）-（FCS）。(2) 若 A_2 满列秩，则满足 (CC1)-(CC2) 的 H 和 G 为正定，

$$r \in (-1, 1),\ s \in (0, 1) \text{ and } r + s > 0.$$

证明。式（1）的证明。 λ-子问题的最优性条件为:

$$0 = A_1\widetilde{x}_1^k + A_2\widetilde{x}_2^k - b - A_2(\widetilde{x}_2^k - x_2^k) + \frac{1}{\beta}(\widetilde{\lambda}^k - \lambda^k). \tag{4.3}$$

或者等价地,

$$\widetilde{\lambda}^k = \lambda^k - \beta(A_1\widetilde{x}_1^k + A_2x_2^k - b). \tag{4.4}$$

x_1-子问题的最优性条件为

$$\begin{aligned} 0 &\in \partial f_1(\breve{x}_1^k) - A_1^T[\lambda^k - \beta(A_1\widetilde{x}_1^k + A_2x_2^k - b)] + P(\widetilde{x}_1^k - x_1^k) \\ &= \partial f_1(\breve{x}_1^k) - A_1^T\widetilde{\lambda}^k + P(\widetilde{x}_1^k - x_1^k). \end{aligned} \tag{4.5}$$

x_2-子问题的最优性条件为:

$$\begin{aligned} 0 &\in \partial f_2(\breve{x}_2^k) - A_2^T[\lambda^k - r\beta(A_1\widetilde{x}_1^k + A_2x_2^k - b)] + \beta A_2^T[A_1\widetilde{x}_1^k + A_2\widetilde{x}_2^k - b] \\ &= \partial f_2(\breve{x}_2^k) - A_2^T\widetilde{\lambda}^k + \beta A_2^TA_2(\widetilde{x}_2^k - x_2^k) - rA_2^T(\widetilde{\lambda}^k - \lambda^k). \end{aligned} \tag{4.6}$$

结合 (4.5)、(4.6) 和 (4.3)，可得

$$0 \in \begin{pmatrix} \partial f_1(\breve{x}_1^k) \\ \partial f_2(\breve{x}_2^k) \\ 0 \end{pmatrix} + \begin{pmatrix} -A_1^T\widetilde{\lambda}^k \\ -A_2^T\widetilde{\lambda}^k \\ A_1\widetilde{x}_1^k + A_2\widetilde{x}_2^k - b \end{pmatrix} + \begin{pmatrix} P(\widetilde{x}_1^k - x_1^k) \\ \beta A_2^TA_2(\widetilde{x}_2^k - x_2^k) - rA_2^T(\widetilde{\lambda}^k - \lambda^k) \\ -A_2(\widetilde{x}_2^k - x_2^k) + \frac{1}{\beta}(\widetilde{\lambda}^k - \lambda^k) \end{pmatrix},$$

这等价于

$$0 \in \begin{pmatrix} \partial f_1(\breve{x}_1^k) \\ \partial f_2(\breve{x}_2^k) \\ 0 \end{pmatrix} + \begin{pmatrix} -A_1^T\widetilde{\lambda}^k \\ -A_2^T\widetilde{\lambda}^k \\ A_1\widetilde{x}_1^k + A_2\widetilde{x}_2^k - b \end{pmatrix} + \begin{pmatrix} P & 0 & 0 \\ 0 & \beta A_2^T A_2 & -rA_2^T \\ 0 & -A_2 & \frac{1}{\beta} I_l \end{pmatrix} \begin{pmatrix} x_1^k - \widetilde{x}_1^k \\ x_2^k - \widetilde{x}_2^k \\ \lambda^k - \widetilde{\lambda}^k \end{pmatrix}.$$

于是我们得到满足 (FPS) 的预测步 (4.1)，其中 L 和 Q 定义见 (2.7)。校正步很容易验证。

(2) 的证明。我们可以参考定理 2.3 (2)。 □

REMARK 4. 根据校正步 (4.2)，有

$$x_1^{k+1} = \widetilde{x}_1^k, \quad x_2^{k+1} = \widetilde{x}_2^k.$$

and

$$\begin{aligned} \lambda^{k+1} &= \lambda^k + s\beta A_2(x_2^k - \widetilde{x}_2^k) - (r+s)(\lambda^k - \widetilde{\lambda}^k) \\ &\overset{(4.4)}{=} \underbrace{\lambda^k - r\beta(A_1\widetilde{x}_1^k + A_2 x_2^k - b)}_{:\lambda^{k+\frac{1}{2}}} - s\beta(A_1\widetilde{x}_1^k + A_2\widetilde{x}_2^k - b) \\ &= \lambda^{k+\frac{1}{2}} - s\beta(A_1\widetilde{x}_1^k + A_2\widetilde{x}_2^k - b) \\ &= \lambda^{k+\frac{1}{2}} - s\beta(A_1 x_1^{k+1} + A_2 x_2^{k+1} - b) \end{aligned}$$

于是算法 (4.1)-(4.2) 可以改写为:

$$\begin{cases} \breve{x}_1^k & \in \arg\min\limits_{x_1} \Big\{ f_1(x_1) - x_1^T A_1^T \lambda^k \\ & \quad + \frac{\beta\tau^k}{2} \|A_1(\frac{1}{\tau^k}x_1 - \frac{1-\tau^k}{\tau^k}\breve{x}_1^{k-1}) + A_2 x_2^k - b\|^2 + \frac{\tau^k}{2}\|\frac{1}{\tau^k}x_1 - \frac{1-\tau^k}{\tau^k}\breve{x}_1^{k-1} - x_1^k\|_P^2 \Big\}, \\ x_1^{k+1} & = \frac{1}{\tau^k}\breve{x}_1^k - \frac{1-\tau^k}{\tau^k}\breve{x}_1^{k-1}, \\ \lambda^{k+\frac{1}{2}} & = \lambda^k - r\beta(A_1\widetilde{x}_1^k + A_2 x_2^k - b), \\ \breve{x}_2^k & \in \arg\min\limits_{x_2} \Big\{ f_2(x_2) - x_2^T A_2^T \lambda^{k+\frac{1}{2}} \\ & \quad + \frac{\beta\tau^k}{2} \|A_1\widetilde{x}_1^k + A_2(\frac{1}{\tau^k}x_2 - \frac{1-\tau^k}{\tau^k}\breve{x}_2^{k-1}) - b\|^2 \Big\}, \\ x_2^{k+1} & = \frac{1}{\tau^k}\breve{x}_2^k - \frac{1-\tau^k}{\tau^k}\breve{x}_2^{k-1}, \\ \lambda^{k+1} & = \lambda^{k+\frac{1}{2}} - s\beta(A_1\widetilde{x}_1^k + A_2\widetilde{x}_2^k - b). \end{cases}$$

我们可以发现，该算法在一次迭代中将 λ 更新两次。从数值结果来看，将 λ 更新两次比更新 λ 一次性能更好。这是第一篇以非遍历收敛方式将 λ 更新两次的论文。

4.2. 满足（FPS）-（FCS）的更快算法用于求解 (P1)。我们设计了一种满足（FPS）-（FCS）的更快算法来求解 (P1)。

我们考虑以下用于求解 (P1) 的算法:

[预测步。] With 给定 $\beta > 0$, $(\breve{x}_1^{k-1}, \breve{x}_2^{k-1}, \ldots, \breve{x}_m^{k-1}, \breve{\lambda}^{k-1})$ 和

$(x_1^k, x_2^k, \ldots, x_m^k, \lambda^k)$, find $(\breve{x}_1^k, \breve{x}_2^k, \ldots, \breve{x}_m^k, \breve{\lambda}^k)$ by

$$
(4.7)\quad \begin{cases}
\breve{x}_1^k = \arg\min\limits_{x_1} \left\{ f_1(x_1) - x_1^T A_1^T \lambda^k + \frac{\beta\tau^k}{2} \| A_1([\frac{1}{\tau^k} x_1 - \frac{1-\tau^k}{\tau^k} \breve{x}_1^{k-1}] - x_1^k) \|^2 \right\}, \\
\breve{x}_2^k = \arg\min\limits_{x_2} \left\{ f_2(x_2) - x_2^T A_2^T \lambda^k \right. \\
\qquad \left. + \frac{\beta\tau^k}{2} \| A_1(\widetilde{x}_1^k - x_1^k) + A_2([\frac{1}{\tau^k} x_2 - \frac{1-\tau^k}{\tau^k} \breve{x}_2^{k-1}] - x_2^k) \|^2 \right\}, \\
\qquad \vdots \\
\breve{x}_i^k = \arg\min\limits_{x_i} \left\{ f_i(x_i) - x_i^T A_2^T \lambda^k \right. \\
\qquad \left. + \frac{\beta\tau^k}{2} \| \sum_{j=1}^{i-1} A_j(\widetilde{x}_j^k - x_j^k) + A_i([\frac{1}{\tau^k} x_i - \frac{1-\tau^k}{\tau^k} \breve{x}_i^{k-1}] - x_i^k) \|^2 \right\}, \\
\qquad \vdots \\
\breve{x}_m^k = \arg\min\limits_{x_m} \left\{ f_m(x_m) - x_m^T A_2^T \lambda^k \right. \\
\qquad \left. + \frac{\beta\tau^k}{2} \| \sum_{j=1}^{m-1} A_j(\widetilde{x}_j^k - x_j^k) + A_m([\frac{1}{\tau^k} x_m - \frac{1-\tau^k}{\tau^k} \breve{x}_m^{k-1}] - x_m^k) \|^2 \right\}, \\
\breve{\lambda}^k = \arg\max\limits_{\lambda} \left\{ -\lambda^T (\sum_{j=1}^m A_j \widetilde{x}_j^k - b) - \frac{\tau^k}{2\beta} \| [\frac{1}{\tau^k} \lambda - \frac{1-\tau^k}{\tau^k} \breve{\lambda}^{k-1}] - \lambda^k \|^2 \right\},
\end{cases}
$$

其中 τ^k 满足 (Y)。

[校正步。] 按以下方式更新 $(A_1 x_1^{k+1}, A_2 x_2^{k+1}, \ldots, A_m x_m^{k+1}, \lambda^{k+1})$:

(4.8)

$$
\begin{pmatrix} \sqrt{\beta} A_1 x_1^{k+1} \\ \sqrt{\beta} A_2 x_2^{k+1} \\ \vdots \\ \sqrt{\beta} A_m x_m^{k+1} \\ \frac{1}{\sqrt{\beta}} \lambda^{k+1} \end{pmatrix} = \begin{pmatrix} \sqrt{\beta} A_1 x_1^k \\ \sqrt{\beta} A_2 x_2^k \\ \vdots \\ \sqrt{\beta} A_m x_m^k \\ \frac{1}{\sqrt{\beta}} \lambda^k \end{pmatrix} - \begin{pmatrix} \alpha I_l & -\alpha I_l & 0 & \ldots & 0 \\ 0 & \alpha I_l & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & -\alpha I_l & 0 \\ 0 & \ldots & 0 & \alpha I_l & 0 \\ -\alpha I_l & 0 & \ldots & 0 & I_l \end{pmatrix} \begin{pmatrix} \sqrt{\beta}(A_1 x_1^k - A_1 \widetilde{x}_1^k) \\ \sqrt{\beta}(A_2 x_2^k - A_2 \widetilde{x}_2^k) \\ \vdots \\ \sqrt{\beta}(A_m x_m^k - A_m \widetilde{x}_m^k) \\ \frac{1}{\sqrt{\beta}} (\lambda^k - \widetilde{\lambda}^k) \end{pmatrix},
$$

其中

$$
\begin{pmatrix} \widetilde{x}_1^k \\ \vdots \\ \widetilde{x}_m^k \\ \widetilde{\lambda}^k \end{pmatrix} = \begin{pmatrix} \frac{1}{\tau^k} \breve{x}_1^k - \frac{1-\tau^k}{\tau^k} \breve{x}_1^{k-1} \\ \vdots \\ \frac{1}{\tau^k} \breve{x}_m^k - \frac{1-\tau^k}{\tau^k} \breve{x}_m^{k-1} \\ \frac{1}{\tau^k} \breve{\lambda}^k - \frac{1-\tau^k}{\tau^k} \breve{\lambda}^{k-1} \end{pmatrix}.
$$

下述定理阐明，算法 (4.7)-(4.8) 满足（FPS）-（FCS）及收敛条件 (CC1)-(CC2)。因此，定理 3.4 和 3.5 可分别用于得到原始-对偶间隙在非遍历意义下的 $O(1/t)$ 速率，以及在逐点意义下的 $O(1/t^2)$ 速率。

THEOREM 4.2. 对于在 (2.12) 中定义的 L, Q, M，以及在例 1 中定义的 $\theta(u)$、u, w, $F(w)$ $T(w)$，有

(1) 算法 (4.7)-(4.8) 满足（FPS）-（FCS）。(2) 若 $\alpha \in (0,1)$，则满足 (CC1)-(CC2) 的 H 和 G 为正定。

证明。式 (1) 的证明。对于 $i = 1, 2, \ldots, m$, x_i-子问题的最优性条件为

子问题为

$$\begin{aligned}(4.9)\qquad & 0 \in \partial f_i(\breve{x}_i^k) - A_i^T\lambda^k + \beta A_i^T\sum_{j=1}^{i} A_j(\widetilde{x}_j^k - x_j^k) \\ & = \partial f_i(\breve{x}_i^k) - A_i^T\widetilde{\lambda}^k + \beta A_i^T\sum_{j=1}^{i} A_j(\widetilde{x}_j^k - x_j^k) + A_i^T(\widetilde{\lambda}^k - \lambda^k)\end{aligned}$$

λ子问题的最优性条件为:

$$(4.10)\qquad 0 = (\sum_{j=1}^{m} A_j\widetilde{x}_j^k - b) + \frac{1}{\beta}(\widetilde{\lambda}^k - \lambda^k).$$

结合 (4.9) 和 (4.10)，得到

$$0 \in \begin{pmatrix} \partial f_1(\breve{x}_1^k) \\ \partial f_2(\breve{x}_2^k) \\ \vdots \\ \partial f_m(\breve{x}_m) \\ 0 \end{pmatrix} + \begin{pmatrix} -A_1^T\widetilde{\lambda}^k \\ -A_2^T\widetilde{\lambda}^k \\ \vdots \\ -A_m^T\widetilde{\lambda}^k \\ \sum_{j=1}^m A_j\widetilde{x}_j^k - b \end{pmatrix} + \begin{pmatrix} \beta A_1^TA_1(\widetilde{x}_1^k - x_1^k) + A_1^T(\widetilde{\lambda}^k - \lambda^k) \\ \beta A_2^T\sum_{j=1}^2 A_j(\widetilde{x}_j^k - x_j^k) + A_2^T(\widetilde{\lambda}^k - \lambda^k) \\ \vdots \\ \beta A_m^T\sum_{j=1}^m A_j(\widetilde{x}_j^k - x_j^k) + A_m^T(\widetilde{\lambda}^k - \lambda^k) \\ \frac{1}{\beta}(\widetilde{\lambda}^k - \lambda^k) \end{pmatrix},$$

这等价于:

$$0 \in \begin{pmatrix} \partial f_1(\breve{x}_1^k) \\ \partial f_2(\breve{x}_2^k) \\ \vdots \\ \partial f_m(\breve{x}_m) \\ 0 \end{pmatrix} + \begin{pmatrix} -A_1^T\widetilde{\lambda}^k \\ -A_2^T\widetilde{\lambda}^k \\ \vdots \\ -A_m^T\widetilde{\lambda}^k \\ \sum_{j=1}^m A_j\widetilde{x}_j^k - b \end{pmatrix} + L^T\begin{pmatrix} I_l & 0 & \dots & 0 & I_l \\ I_l & I_l & \ddots & \vdots & I_l \\ \vdots & \vdots & \ddots & 0 & \vdots \\ I_l & I_l & \dots & I_l & I_l \\ 0 & 0 & \dots & 0 & I_l \end{pmatrix} L\begin{pmatrix} \widetilde{x}_1^k - x_1^k \\ \widetilde{x}_2^k - x_2^k \\ \vdots \\ \widetilde{x}_m^k - x_m^k \\ \widetilde{\lambda}^k - \lambda^k \end{pmatrix}.$$

然后，预测步 (4.8) 满足 (FPS)，其中 L 和 Q 由 (2.12) 定义。校正步很容易验证。

(2) 的证明。可参见定理 2.4 (2)。 □

4.3. 满足（FPS）-（FCS）的用于求解（P2）的更快算法。本小节介绍一种满足（FPS）-（FCS）的用于求解（P2）的算法。我们考虑以下用于求解（P2）的算法:

[预测步。] 给定 r、$s>0$、$(\breve{x}^{k-1},\ \breve{y}^{k-1})$ 和 $(x^k,\ y^k)$，通过以下方式求得 $(\breve{x}^k,\ \breve{y}^k)$:

$$(4.11)\qquad \begin{cases} \breve{x}^k = \arg\min\limits_x \left\{\Phi(x, y^k) + \frac{r\tau^k}{2}\|(\frac{1}{\tau^k}x - \frac{1-\tau^k}{\tau^k}\breve{x}^{k-1}) - x^k\|^2\right\}, \\ \breve{y}^k = \arg\max\limits_y \left\{\Phi([\widetilde{x}^k + \alpha(\widetilde{x}^k - x^k)], y) - \frac{s\tau^k}{2}\|(\frac{1}{\tau^k}y - \frac{1-\tau^k}{\tau^k}\breve{y}^{k-1}) - y^k\|^2\right\}, \end{cases}$$

其中，τ^k 满足（Y）。

[校正步。] 通过更新 (x^{k+1}, y^{k+1})

$$(4.12)\qquad \begin{pmatrix} x^{k+1} \\ y^{k+1} \end{pmatrix} = \begin{pmatrix} x^k \\ y^k \end{pmatrix} - \begin{pmatrix} I_n & 0 \\ -(1-\alpha)\frac{1}{s}A & I_m \end{pmatrix}\begin{pmatrix} x^k - \widetilde{x}^k \\ y^k - \widetilde{y}^k \end{pmatrix},$$

其中

$$\begin{pmatrix}\widetilde{x}^k\\ \widetilde{y}^k\end{pmatrix}=\begin{pmatrix}\frac{1}{\tau^k}\breve{x}^k-\frac{1-\tau^k}{\tau^k}\breve{x}^{k-1}\\ \frac{1}{\tau^k}\breve{y}^k-\frac{1-\tau^k}{\tau^k}\breve{y}^{k-1}\end{pmatrix}.$$

下面的定理说明，算法（4.11）-（4.12）满足（FPS）-（FCS）以及收敛条件（CC1）-（CC2）。因此，可由定理3.4和3.5得到原始-对偶间隙的非遍历意义下的 O（$1/t$）收敛率，以及逐点意义下的 O（$1/t^2$）收敛率。

THEOREM 4.3. 对于在（2.12）中定义的 L、 Q、 M以及如例 2中所定义的 θ（u） u、 w、F（w） T（w），有

（1）算法（4.11）-（4.12）满足（FPS）-（FCS）。（2）满足（CC1）-（CC2）的 H 和 G是正定的，如果

$$rs>(1-\alpha+\alpha^2)\rho(A^TA),\ \alpha\in[0,1].$$

证明。（1）的证明。 x-子问题的最优性条件为：

$$\begin{aligned}0 &\in \partial f(\breve{x}^k)-A^Ty^k+r(\widetilde{x}^k-x^k)\\ &= \partial f(\breve{x}^k)-A^T\widetilde{y}^k+r(\widetilde{x}^k-x^k)+A^T(\widetilde{y}^k-y^k).\end{aligned}$$

y-子问题的最优性条件为：

$$0\in\partial g(\breve{y}^k)+A[\widetilde{x}^k+\alpha(\widetilde{x}^k-x^k)]+s(\widetilde{y}^k-y^k).$$

将上述两个关系结合起来可得

$$0\in\begin{pmatrix}\partial f(\breve{x}^k)\\ \partial g(\breve{y}^k)\end{pmatrix}+\begin{pmatrix}-A^T\widetilde{y}^k\\ A\widetilde{x}^k\end{pmatrix}+\begin{pmatrix}r(\widetilde{x}^k-x^k)+A^T(\widetilde{y}^k-y^k)\\ \alpha A(\widetilde{x}^k-x^k)+s(\widetilde{y}^k-y^k)\end{pmatrix}.$$

这等价于：

$$0\in\begin{pmatrix}\partial f(\breve{x}^k)\\ \partial g(\breve{y}^k)\end{pmatrix}+\begin{pmatrix}-A^T\widetilde{y}^k\\ A\widetilde{x}^k\end{pmatrix}+\begin{pmatrix}rI_n & A^T\\ \alpha A & sI_m\end{pmatrix}\begin{pmatrix}\widetilde{x}^k-x^k\\ \widetilde{y}^k-y^k\end{pmatrix}.$$

于是我们得到，预测步(4.11)满足(FPS)，其中 L 和 Q 由(2.15)定义。校正步很容易验证。

(2)的证明。我们可以参见定理2.5中的(2)。 □

5. 结论。我们提出了一种更快速的预测-校正框架，无需任何附加假设，即可在非遍历意义下获得$O(1/t)$收敛率，并在逐点意义下获得 $O(1/t^2)$收敛率。相比之下，He 和 Yuan 的框架在遍历意义和逐点意义下均达到 $O(1/t)$收敛率。我们的框架可为求解一般凸优化问题提供更快算法。具体而言，我们提出了三种更快算法：用于求解带线性等式约束的两块可分凸优化问题的双重更新对偶变量的 ADMM 型算法；用于求解带线性等式约束的多块可分凸优化问题的多块ADMM型算法；以及采用较大步长的求解极小极大问题的 CP型算法（$rs>0.75\rho(A^TA)$）。未来的工作包括深入理解我们的框架，例如从二阶微分方程的角度进行研究，并证明迭代序列的弱收敛性及KKT度量的收敛性。

参考文献

[1] H. Attouch and S. László, *Newton-like inertial dynamics and proximal algorithms governed by maximally monotone operators* 30 2020 3252 3283 , SIAM 优化杂志, ( ), 第 – 页。

[2] H. Attouch and J. Peypouquet, *Convergence of inertial dynamics and proximal algorithms governed by maximal monotone operators* 2017 1 42 , 数学规划, ( ), 第 – 页。

[3] A. Beck and M. Teboulle, *A fast iterative shrinkage-thresholding algorithm for linear inverse problems* 2 2009 183 202 , SIAM 成像科学杂志, ( ), 第 – 页。

[4] A. Chambolle and T. Pock, *A first-order primal-dual algorithm for convex problems with applications to imaging* 40 2011 120 145 , 数学成像与视觉杂志, ( ), 第 – 页。

[5] A. Chambolle and T. Pock, *On the ergodic convergence rates of a first-order primal-dual algorithm* 159 2016 253 287 , 数学规划, ( ), 第 – 页。

[6] C. Chen, B. He, Y. Ye, and X. Yuan, *The direct extension of ADMM for multi-block convex minimization problems is not necessarily convergent* 155 2016 57 , 数学规划, ( ), 第 – 79 页。

[7] D. Gabay and B. Mercier, *A dual algorithm for the solution of nonlinear variational problems via finite element approximation* 2 1976 17 40 , 计算数学与应用, ( ), 第 – 页。

[8] R. Glowinski and A. Marroco, *Sur l'approximation, par éléments finis d'ordre un, et larésolution, par pénalisation-dualité d'une classe de problèmes de Dirichlet non linéaires*, Esaim-Math. Model. Num., 9 (1975), 第 41–76 页。

[9] N. Golowich, S. Pattathil, C. Daskalakis, and A. Ozdaglar, *Last iterate is slower than averaged iterate in smooth convex-concave saddle point problems*, 学习理论会议（PMLR）, (2020), 第 1758–1784 页。

[10] B. He, L. Hou, and X. Yuan, *On full Jacobian decomposition of the augmented Lagrangian method for separable convex programming* 25 2015 2274 2312 , SIAM 优化杂志, ( ), 第 – 页。

[11] B. He, H. Liu, Z. Wang, and X. Yuan, *A strictly contractive Peaceman-Rachford splitting method for convex programming* 24 2014 1011 1040 , SIAM 优化杂志, ( ), 第 – 页。

[12] B. He, F. Ma, S. Xu, and X. Yuan, *A generalized primal-dual algorithm with improved convergence condition for saddle point problems* 15 2022 1157 1183 , SIAM 成像科学杂志, ( ), 第 – 页。

[13] B. He, F. Ma, and X. Yuan, *Convergence study on the symmetric version of ADMM with larger step sizes* 9 2016 1467 1501 , SIAM 成像科学杂志, ( ), 第 – 页。

[14] B. He, F. Ma and X. Yuan *An algorithmic framework of generalized primal–dual hybrid gradient methods for saddle point problems* , , , 数学成像与视觉杂志, ( ), 第 – 293 页。

[15] B. He, F. Ma, and X. Yuan, *Optimal proximal augmented Lagrangian method and its application to full Jacobian splitting for multi-block separable convex minimization problems*, IMA 数值分析杂志, 40 (2020), 第 1188–1216 页。

[16] B. He, M. Tao, and X. Yuan, *Alternating direction method with Gaussian back substitution for separable convex programming* 22 2012 313 340 , SIAM 优化杂志, ( ), 第 – 页。

[17] B. He, M. Tao, and X. Yuan, *Convergence rate analysis for the alternating direction method of multipliers with a substitution procedure for separable convex programming* , 数学运算研究, 42 (2017), 第 662–691 页。

[18] B. He, S. Xu, and X. Yuan, *Extensions of ADMM for separable convex optimization problems with linear equality or inequality constraints* 2107 01897, 2021 , arXiv 预印本 arXiv: . ( )。

[19] B. He and X. Yuan, *On the $O(1/t)$ convergence rate of the alternating direction method*, SIAM 数值分析杂志, 50 (2012), 第 700–709 页。

[20] B. He and X. Yuan, *On non-ergodic convergence rate of DouglascRachford alternating direction method of multipliers* 130 2015 567 577 , 数值数学, ( ), 第 – 页。

[21] B. He and X. Yuan, *On construction of splitting contraction algorithms in a prediction-correction framework for separable convex optimization* 2204 11522, 2022 , arXiv: . ( )。

[22] X. He, R. Hu, and Y. Fang, *Inertial primal-dual methods for linear equality constrained convex optimization problems* 90 2022 1669 1690 , 数值算法, ( ), 第 – 页。

[23] M. R. Hestenes, *Multiplier and gradient methods*, 优化理论与应用杂志, 4 (1969), 第 303–320 页。

[24] R. Ioan Bot, E. Csetnek, and D. Nguyen, *Fast augmented Lagrangian method in the convex regime with convergence guarantees for the iterates* 200 2023 147 , 数学规划, ( ), 第 – 197 页。

[25] F. Jiang, Z. Zhang, and H. He, *Solving saddle point problems: a landscape of primal-dual algorithm with larger stepsizes* 85 2023 821 846 , 全局优化杂志, ( ), 第 – 页。

[26] H. Li and Z. Lin, *Accelerated alternating direction method of multipliers: an optimal $O(1/K)$ nonergodic analysis*, 科学计算杂志, 79 (2019), 第 671–699 页。

[27] H. Li and M. Yan, *On the improved conditions for some primal-dual algorithms*,

arXiv:2201.00139v1, (2022). [28] Z. Lin, H. Li, and C. Fang, *Accelerated Optimization for Machine Learning: First-Order Algorithms* 2020 , 新加坡施普林格·自然出版社, 。[29] H. Luo、 *Accelerated primal-dual methods for linearly constrained convex optimization problems*, arXiv:2109.12604, (2021). [30] H. Luo, *A unified differential equation solver approach for separable convex optimization: splitting acceleration and nonergodic rate* 2109 13467, 2021 , , arXiv:。 (). [31] H. Luo、 *A primal-dual flow for affine constrained convex optimization*, ESAIM 控制、优化与变分法, 28（2022）。[32] H. Luo and L. Chen, *From differential equation solvers to accelerated first-order methods for convex optimization* 195 2022 735 781 , 数学规划, (), 第–页。[33] B. Martinet、 *Regularisation d'inequations variationelles par approximations successives*, 《法国计算机科学与运筹研究评论》, 4 (1970), 第154–159页。[34] Y. Nesterov, *A method for solving the convex programming problem with convergence rate O* 269 1983 543 547 $(1/k^2)$, 《苏联科学院院刊》, (), 第–页。[35] M. J. D. Powell、 *A method for nonlinear constraints in minimization problems*, 《优化》, (1969), 第283–298页。[36] R. T. Rockafellar, *Augmented Lagrangians and applications of the proximal point algorithm in convex programming* 1 1976 97 116 , 数学运筹研究, (), 第–页。[37] R. T. Rockafellar、 *Monotone operators and the proximal point algorithm*, SIAM 控制与优化杂志, 14（1976）, 第877–898页。[38] S. Sabach and M. Teboulle、 *Faster Lagrangian-based methods in convex optimization*, SIAM 优化杂志, 32（2022）, 第204–227页。[39] Q. Tran-Dinh and Y. Zhu, *Non-stationary first-order primal-dual algorithms with faster convergence rates* 30 2020 2866 2896 , SIAM 优化杂志, (), 第–页。[40] P. Tseng、 *Approximation accuracy gradient methods and error bound for structured convex optimization* 125 2010 263 295 , , , 《数学规划》, (), 第–页。[41] T. Valkonen、 *Inertial、 corrected、 primal-dual proximal splitting*, SIAM 优化杂志, 30（2020）, 第1391–1420页。[42] S. Xu and B. He、 *A parallel splitting ALM-based algorithm for separable convex programming*, 《计算与优化应用》, 80（2021）, 第831–851页。[43] J. Yang and X. Yuan, *Linearized augmented Lagrangian and alternating direction methods for nuclear norm minimization* 82 2013 301 329 , 《数学计算》, (), 第–页。