

大规模可分凸优化问题的非精确自适应步随机原始对偶算法

周晓艳

重庆师范大学数学科学学院, 重庆

收稿日期: 2024年3月6日; 录用日期: 2024年3月26日; 发布日期: 2024年4月30日

摘要

本文研究了可分优化问题, 针对其目标函数的可分性, 分裂算法将目标函数分解成更小、更容易处理的子问题, 如原始对偶混合梯度算法。本文探讨了目标函数的邻近算子的非精确求解策略, 并基于此提出了一个非精确自适应步随机原始对偶算法。我们分析了误差序列选取方式对算法收敛速率的影响, 发现不同的误差序列选择会导致算法在收敛速度和稳定性方面表现出显著的差异。此外, 该算法在实际应用中也展现出了更高的效率和灵活性。

关键词

可分凸优化问题, 随机算法, 原始对偶算法, 自适应策略, 非精确求解

Inexact Adaptive Stochastic Primal-Dual Algorithm for Large Scale Separable Convex Optimization Problems

Xiaoyan Zhou

School of Mathematical Sciences, Chongqing Normal University, Chongqing

Received: Mar. 6th, 2024; accepted: Mar. 26th, 2024; published: Apr. 30th, 2024

Abstract

The separable optimization problem refers to optimizing an objective function with a separable structure. Splitting algorithms address this characteristic by breaking down the objective function into smaller, more manageable subproblems, such as the primal-dual hybrid gradient algorithm. We explore the approximate solution strategy of the objective function's proximal operator and propose an inexact adaptive stochastic primal-dual algorithm. We analyze the impact of the selection of error sequences on the convergence rate of the algorithm and find that different selections of error sequences can lead to significant differences in convergence speed and stability. In addition, this algorithm has also demonstrated higher efficiency and flexibility in practical applications.

Keywords

Large Scale Separable Convex Optimization Problem, Stochastic Optimization, Primal-Dual Algorithms, Adaptive Step-Size, Inexact Proximal Operator

Copyright © 2024 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

具有可分结构的凸优化问题广泛应用于生活生产的多个领域. 例如, 在图像处理 [1, 2]中, 我们常常需要通过优化算法来解决图像恢复、图像分割等问题; 在机器学习 [3]中, 许多模型的训练过程可以被视为凸优化问题的求解过程; 而在统计学 [4, 5]中, 凸优化问题也被广泛应用于参数估计, 假设检验等方面. 其一般形式如下:

$$\min_{x \in X} \left\{ \Phi(x) := \sum_{i=1}^n f_i(A_i x) + g(x) \right\}.$$

其中, f_i 是真凸函数且下半连续, g 是凸函数, A_i 是有界线性算子. Chambolle 和Pock 在文献 [1]中提出了PDHG算法, 这一算法巧妙地运用了分裂技术, 能够将原本复杂的优化问题拆解为一系列相对简单的子问题, 并通过交替迭代子问题来逼近原始问题的最优解, 加速整体的收敛过程. 特别

地, PDHG算法的计算过程主要依赖于近似算子的求解和矩阵向量乘法的简单运算。当目标函数的邻近算子具备显式表达式时, PDHG相较于其他一阶或二阶算法, 在计算的简洁性和内存占用的经济性上展现出了优势。因此, 在图像处理、机器学习和信号处理等多个关键领域中, PDHG算法都展现出了卓越的性能和实用性。然而, 当目标函数的邻近算子没有显式解或者求解过程复杂时, PDHG算法的应用可能会受到一定的限制。

非精确邻近算子求解在优化算法领域中受到广泛关注, 尤其是在处理复杂问题时, 精确计算邻近算子因受到计算资源和时间的限制而变得不切实际。为了平衡计算成本和求解精度, 非精确求解策略崭露头角。大量学者针对实际应用中的非精确邻近算子问题进行了深入研究, 例如文献 [6, 7]提出了采用绝对误差准则来控制子问题误差的方法, 其中误差容限由一个可求和的误差序列来调控。此外, 文献 [8] 引入了相对误差准则, 进一步放宽了对误差序列的要求, 即参数需满足平方可求和的条件。在算法应用方面, Villa 等人 [9, 10]将非精确计算邻近算子策略应用于前向-后向分裂算法, 并深入分析了算法的收敛性能。同时, Lin等人 [11, 12]提出了一种加速的近端坐标梯度方法, 该方法适用于大规模优化问题。

受到先前研究的启发, 本文在随机原始对偶算法的基础上, 进一步考虑了非精确求解目标函数邻近算子的情况。针对这一情境, 我们创新性地提出了非精确自适应随机原始对偶算法, 并对其收敛速率与误差序列之间的关系进行了深入分析。我们发现, 当误差序列满足一定的条件时, 算法仍能保持原有的收敛速率 $O(1/N)$ 。这一重要发现不仅极大地拓展了算法的应用范围, 而且为处理实际计算中的误差问题提供了坚实的理论支撑。

本文首先介绍了目标问题的相关定义和假设; 然后详细阐述了非精确自适应步随机原始对偶算法的框架和收敛性分析; 接着展示了该算法在正电子发射断层成像(PET)问题中的实际应用; 最后总结了全文内容并提出了未来研究方向。

2. 非精确求解大规模可分优化问题的原始对偶算法

针对大规模可分凸优化问题, 我们分析研究了非精确计算目标函数的邻近算子的情况。通过引入非精确计算邻近算子策略 [8, 13, 14], 我们能够在保证算法收敛性的同时, 降低对计算精度的要求, 从而进一步提高算法在实际应用中的效率和稳定性。此外, 我们对非精确计算下的PDHG算法的收敛性进行了详细的分析和论证。通过理论推导和实验验证, 我们证明了在非精确计算的情况下, 算法依然能够保持良好的收敛性能, 为实际应用提供了坚实的理论基础。

2.1. 模型描述

设 $Y_i, i = 1, \dots, n$ 是任意维的实希尔伯特空间, 定义积空间 $Y = \prod_{i=1}^n Y_i$, 对任意的 $y \in Y$, 记 $y = (y_1, y_2, \dots, y_n)$, 其中 $y_i \in Y_i$ 。在 Y 空间上定义内积 $\langle y, z \rangle = \sum_{i=1}^n \langle y_i, z_i \rangle$, 其中 $y_i, z_i \in Y_i$ 。自然地, 在 Y 空间上由内积诱导的范数为 $\|y\|^2 = \sum_{i=1}^n \|y_i\|^2$ 。

考虑 $A : X \rightarrow Y$ 的一个有界线性算子, 根据积空间的定义, 有 $(Ax)_i = A_i x$, 其中 $A_i : X \rightarrow Y_i$ 是线性算子; 则 A 的伴随算子满足 $A^* y = \sum_{i=1}^n A_i^* y_i$ 。记 $f(y) = \sum_{i=1}^n f_i(y_i)$, 函数 $f : Y \rightarrow R_\infty$ 是真凸函数,

且下半连续. $g: X \rightarrow R_\infty$ 是凸函数.

根据上面的定义及假设, 考虑以下优化问题:

$$\min_{x \in X} \left\{ \Phi(x) := \sum_{i=1}^n f_i(A_i x) + g(x) \right\}. \quad (2.1.1)$$

在介绍具体的求解算法之前, 为了提供更为全面和清晰的理解, 我们首先需要介绍一些相关的定义和背景知识.

定义2.1.1. [15] 若函数 $f: E \rightarrow (-\infty, \infty]$, 则对任意的 $x \in E$, 邻近映射 $\text{prox}_f(x)$ 定义为:

$$\text{prox}_f(x) = \arg \min_{u \in E} \left\{ \frac{1}{2} \|u - x\|^2 + f(u) \right\}.$$

特别地, $\text{prox}_f^\tau(x) = \arg \min_{u \in E} \left\{ \frac{1}{2} \|u - x\|_{\tau^{-1}}^2 + f(u) \right\}$, 其中 $\|x\|_{\tau^{-1}}^2 = \langle \tau^{-1}x, x \rangle$.

性质2.1.1. [15] 若函数 $f: E \rightarrow (-\infty, \infty]$ 是一个闭真凸函数, 则对任意的 $x, u \in E$, 以下结果等价:

- (1) $u = \text{prox}_f(x)$;
- (2) $x - u \in \partial f(u)$;
- (3) 对任意的 $y \in E$, $\langle x - u, y - u \rangle \leq f(y) - f(u)$.

定义2.1.2. [15] 设函数 $f: E \rightarrow [-\infty, +\infty]$ 是广义实值函数, 则 f 的Fenchel共轭函数 $f^*: E^* \rightarrow [-\infty, +\infty]$ 定义为:

$$f^*(y) = \sup_{x \in R^*} \{ \langle x, y \rangle - f(x) \}, \quad y \in E^*,$$

特别地, 当 f 是闭真凸函数时, f^{**} 也是闭真凸函数, 且 $f(y) = f^{**}(y)$.

于是寻求问题(2.1.1)的解 x 等价于寻求以下鞍点问题的原始变量 x :

$$\min_{x \in X} \sup_{y \in Y} \left\{ \psi(x, y) := g(x) + \sum_{i=1}^n \langle A_i x, y_i \rangle - f_i^*(y_i) \right\}. \quad (2.1.2)$$

假设问题(2.1.2)的鞍点 $w^\# = (x^\#, y^\#) = (x^\#, y_1^\#, y_2^\#, \dots, y_n^\#)$ 存在, 根据鞍点定义原始间隙和对偶间隙:

$$P_{x^\#, y^\#}(x) = g(x) - g(x^\#) + \langle A^* y^\#, x - x^\# \rangle \geq 0 \quad \forall x \in X, \quad (2.1.3)$$

$$D_{x^\#, y^\#}(y) = f^*(y) - f^*(y^\#) - \langle A x^\#, y - y^\# \rangle \geq 0, \quad \forall y \in Y. \quad (2.1.4)$$

则原始对偶间隙可以表示为: $G_{x^\#, y^\#}(x, y) = P_{x^\#, y^\#}(x) + D_{x^\#, y^\#}(y) \geq 0$. 在本文讨论的大规模问题中, 函数 f 展现出了可分结构特性, 具体表现为 $f(y) = \sum_{i=1}^n f_i(y_i)$. 实际上, 对于这类可分函数, 其Fenchel共轭满足以下结论:

性质2.1.2. [15] 设 $Y := \prod_{i=1}^n Y_i$, $f(y) = \sum_{i=1}^n f_i(y_i)$, 其中 $f_i, i = 1, \dots, n$ 是真函数, 则

$$f^*(y_1, y_2, \dots, y_n) = \sum_{i=1}^n f_i^*(y_i)$$

定义2.1.3. [16] 函数 $g: \mathcal{X} \rightarrow (-\infty, +\infty]$ 在 $z \in \mathcal{X}$ 处的 ε -次微分定义为:

$$\partial_\varepsilon g(z) := \{p \in \mathcal{X} \mid g(x) \geq g(z) + \langle p, x - z \rangle - \varepsilon \quad \forall x \in \mathcal{X}\}.$$

定义2.1.4. 设近似精度 $\varepsilon \geq 0$. $z \in \mathcal{X}$ 是邻近算子 $\text{prox}_g^\tau(y)$ 的一个2-型非精确解, 若

$$z \approx_2^\varepsilon \text{prox}_g^\tau(y) \stackrel{\text{def}}{\iff} \frac{y - z}{\tau} \in \partial_\varepsilon g(z).$$

虽然有如此多的邻近算子的非精确解的定义, 但是2-型非精确解在实际问题中应用最为广泛. 事实上, 求解2-型非精确解问题可以转换为一个下降对偶间隙的问题.

推论2.1.1. [7] 若 $z \approx_2^\varepsilon \text{prox}_g^\lambda(y)$, 则存在 $v \in \text{dom } \omega^*$ 使得 $z = y - \lambda B^*v$.

推论2.1.2. [7] 设 $z \in \mathcal{Z}$, 则

$$\mathcal{G}(y - B^*z, z) \leq \varepsilon \Rightarrow y - B^*z \approx_2^\varepsilon \text{prox}_g^\tau(y).$$

上述推论对于非精确子问题的求解具有重要的实践意义, 在数值实验中我们采用FISTA 算法 [17]来求解子问题的对偶问题, 使得其对偶间隙满足预先设定的误差精度, 从而找到子问题的一个2-型非精确解.

2.2. 非精确原始对偶混合梯度算法

受到 [9–11, 13]等研究工作的启发, 本文结合以上非精确邻近算子的定义, 在上一节的算法基础上, 提出了非精确自适应随机原始对偶算法, 该算法不仅研究了非精确计算目标函数邻近算子的情况, 而且结合了自适应策略, 提高了算法的效率和稳定性. 通过对误差序列的分析, 我们评估了误差对算法收敛效果的影响, 从而为实际应用中如何选择合适的误差容限提供了指导. 此外, 我们将该算法应用于正电子发射断层成像 (PET) 问题, 成功地提高了重建图像的质量和计算效率, 验证了算法在实际应用中的可行性和有效性.

该算法的具体步骤如下:

算法 1. 非精确自适应步随机原始对偶算法

初始化: $x^{(0)} \in X$, $y^{(1)} \in Y$, $\tau_{(0)} > 0$, $\mu \in (0, 1)$, $\delta \in (0, 1)$, $\beta > 0$, $N \in \mathbb{N}^+$, $e_{(k)}$.

主循环: 对 $k = 1, 2, \dots, N$ 执行:

1. 计算 $x^{(k)} \approx_2^{e_{(k-1)}} \text{prox}_g^{\tau_{(k-1)}}(x^{(k-1)} - \tau_{(k-1)}A^*y^{(k)})$.

2. 任意选择 $\tau_{(k)} := \tau_{(k-1)}$ 及 $S_{(k)} \subset \{1, 2, \dots, n\}$.

(2a) 计算 $\theta_{(k)} = \frac{\tau_{(k)}}{\tau_{(k-1)}}$.

$$\bar{x}^{(k)} = x^{(k)} + \theta_{(k)}(x^{(k)} - x^{(k-1)}).$$

$$y_i^{(k+1)} = \begin{cases} \text{prox}_{f_i^*}^{\beta\tau_{(k)}}(y^{(k)} + \beta\tau_{(k)}A_i\bar{x}^{(k)}), & i \in S_{(k)}, \\ y_i^{(k)}, & i \notin S_{(k)}. \end{cases}$$

(2b) 线性搜索步长终止条件: $\sqrt{\beta}\tau_{(k)} \|A^*y^{(k+1)} - A^*y^{(k)}\| \leq \delta \|y^{(k+1)} - y^{(k)}\|$.

否则, 令 $\tau_{(k)} := \tau_{(k)}\mu$, 并返回(2a).

注2.2.1. 通过推论2.1.2来实现子问题 $x^{(k)} \approx_2^{e_{(k-1)}} \text{prox}_g^{\tau_{(k-1)}}(x^{(k-1)} - \tau_{(k-1)}A^*y^{(k)})$ 的非精确求解, 即下降 $\mathcal{G}(y - B^*z, z)$ 的值达到预先设定的精度 $e_{(k-1)}$.

注2.2.2. 误差序列 $e_{(k)}$ 的选择对算法的收敛性起着关键作用. 当 $e_{(k)} = O(N^{-\alpha}), \alpha > 1$ 时, 算法的收敛速率与精确求解的情况一致. 换句话说, 原始对偶间隙按期望以 $O(\frac{1}{N})$ 的速率收敛. 定理2.3.2详细说明了, 不同情况下的误差序列对收敛速率的影响.

2.3. 收敛性分析

引理2.3.1. 在非精确自适应步随机原始对偶算法中, 有以下结论成立:

- (1) 该算法的子循环迭代有限次后, 一定会满足终止条件 (2b);
- (2) 对任意的 $k \geq 0$, 存在 τ , 使得 $\tau_{(k)} \geq \tau$.

证明: (1) 每次子循环迭代执行 $\tau_{(k)} \cdot \mu (\mu < 1)$, 则存在 k' , 使得 $\tau_{(\mu')} \leq \frac{\delta}{\sqrt{\beta}\|A^*\|}$, 则:

$$\sqrt{\beta}\tau_{(k')} \left\| A^*y^{(k'+1)} - A^*y^{(k')} \right\| \leq \sqrt{\beta}\tau_{(k')} \|A^*\| \left\| y^{(k'+1)} - y^{(k')} \right\| \leq \delta \left\| y^{(k'+1)} - y^{(k')} \right\|.$$

线性搜索子循环终止.

(2) 不失一般性, 假设 $\tau_{(0)} > \frac{\delta\mu}{\sqrt{\beta}\|A^*\|}$, 且 $\tau_{(k-1)} > \frac{\delta\mu}{\sqrt{\beta}\|A^*\|}$, 下面证明 $\tau_{(k)} > \frac{\delta\mu}{\sqrt{\beta}\|A^*\|}$.

设 $\tau_{(k)} = \tau_{(k-1)}\mu^j, j \in N$. 当 $j = 0$ 时, $\tau_{(k)} \geq \tau_{(k-1)} > \frac{\delta\mu}{\sqrt{\beta}\|A^*\|}$ 成立; 当 $j > 0$ 时, 则有 $\tau'_{(k)} = \tau_{(k-1)}\mu^{j-1}$ 不满足线性搜索终止条件(2b), 即:

$$\sqrt{\beta}\tau'_{(k)} \|A^*\| \left\| y^{(k+1)} - y^{(k)} \right\| \geq \sqrt{\beta}\tau'_{(k)} \left\| A^*y^{(k+1)} - A^*y^{(k)} \right\| > \delta \left\| y^{(k+1)} - y^{(k)} \right\|.$$

因此

$$\tau'_{(k)} > \frac{\delta}{\sqrt{\beta}\|A^*\|}, \quad \tau_{(k)} > \frac{\delta\mu}{\sqrt{\beta}\|A^*\|}.$$

非精确自适应步长随机原始对偶算法的收敛性分析是一个从确定到随机的过程.

引理2.3.2. 在非精确自适应步随机原始对偶算法中, 考虑对某个固定的 k , 对偶变量完全选取更新, 即:

$$\hat{y}_i^{(k+1)} = \text{prox}_{f_i^*}^{\beta\tau_k}(y^{(k)} + \beta\tau_k A_i \bar{x}^{(k)}),$$

则有

$$\begin{aligned} & \frac{1}{2}(\|x^{(k)} - x^\# \|^2 - \|x^{(k+1)} - x^\# \|^2) + \frac{1}{2\beta}(\|y^{(k)} - y^\# \|^2 - \|\hat{y}^{(k+1)} - y^\# \|^2) \\ & - \frac{1}{2}\|\bar{x}^{(k)} - x^{(k+1)}\|^2 - \frac{1-\delta^2}{2\beta}\|\hat{y}^{(k+1)} - y^{(k)}\|^2 \\ & \geq \varepsilon_{(k)} - \tau_{(k)}e_{(k-1)} - \theta_{(k)}\tau_{(k)}e_{(k-1)} - \tau_{(k)}e_{(k)}, \end{aligned}$$

其中, $\varepsilon_{(k)} = \tau_{(k)}((1 + \theta_{(k)})P_{x^\#, y^\#}(x^{(k)}) - \theta_{(k)}P_{x^\#, y^\#}(x^{(k-1)}) + D_{x^\#, y^\#}(\hat{y}^{(k+1)}))$.

证明: 假设 $(x^\#, y^\#)$ 是问题(2.1.2)的鞍点, 若对偶变量全部更新, 根据性质2.1.1中(3), 对任

意 $i = 1, 2, \dots, n$, 有

$$\tau_{(k)}(f_i^*(\hat{y}_i^{(k+1)}) - f_i^*(y_i^\#)) \leq \left\langle \frac{1}{\beta}(\hat{y}_i^{(k+1)} - y_i^{(k)}) - \tau_{(k)}A_i\bar{x}^{(k)}, y_i^\# - \hat{y}_i^{(k+1)} \right\rangle. \quad (2.3.1)$$

由于 $f_i^*(y_i^*) = \sup_{u_i \in Y_i} \{\langle y_i^*, u_i \rangle - f_i(u_i)\}$ (性质2.1.2), 则:

$$\begin{aligned} \sum_{i=1}^n f_i^*(y_i^\#) &= \sum_{i=1}^n \sup_{u_i \in Y_i} \left\{ \langle y_i^\#, u_i \rangle - f_i(u_i) \right\} = \sup_{u \in Y} \left\{ \sum_{i=1}^n \langle y_i^\#, u_i \rangle - f_i(u_i) \right\} \\ &= \sup_{u \in Y} \left\{ \langle y^\#, u \rangle - f(u) \right\} = f^*(y^\#). \end{aligned}$$

取(2.3.1) 式中 $i = 1, 2, \dots, n$ 相加可得:

$$\tau_{(k)}(f^*(\hat{y}^{(k+1)}) - f^*(y^\#)) \leq \left\langle \frac{1}{\beta}(\hat{y}^{(k+1)} - y^{(k)}) - \tau_{(k)}A\bar{x}^{(k)}, y^\# - \hat{y}^{(k+1)} \right\rangle. \quad (2.3.2)$$

由于 $x^{(k+1)} \approx_2^{e^{(k)}} \text{prox}_g^{\tau_{(k)}}(x^{(k)} - \tau_{(k)}A^*y^{(k+1)})$, 根据定义2.1.4可得:

$$\tau_{(k)}(g(x^{(k+1)}) - g(x^\#)) \leq \langle x^{(k+1)} - x^{(k)} + \tau_{(k)}A^*\hat{y}^{(k+1)}, x^\# - x^{(k+1)} \rangle + \tau_{(k)}e_{(k)}. \quad (2.3.3)$$

又因为 $x^{(k)} \approx_2^{e^{(k-1)}} \text{prox}_g^{\tau_{(k-1)}}(x^{(k-1)} - \tau_{(k-1)}A^*y^{(k)})$, 同理, 对 $\forall x \in X$, 有

$$\tau_{(k-1)}(g(x^{(k)}) - g(x)) \leq \langle x^{(k)} - x^{(k-1)} + \tau_{(k-1)}A^*y^{(k)}, x - x^{(k)} \rangle + \tau_{(k-1)}e_{(k-1)}. \quad (2.3.4)$$

分别取 $x = x^{(k+1)}$, $x = x^{(k-1)}$ 带入(2.3.4) 式, 得到:

$$\tau_{(k-1)}(g(x^{(k)}) - g(x^{(k+1)})) \leq \langle x^{(k)} - x^{(k-1)} + \tau_{(k-1)}A^*y^{(k)}, x^{(k+1)} - x^{(k)} \rangle + \tau_{(k-1)}e_{(k-1)}, \quad (2.3.5)$$

$$\tau_{(k-1)}(g(x^{(k)}) - g(x^{(k+1)})) \leq \langle x^{(k)} - x^{(k-1)} + \tau_{(k-1)}A^*y^{(k)}, x^{(k-1)} - x^{(k)} \rangle + \tau_{(k-1)}e_{(k-1)}. \quad (2.3.6)$$

(2.3.5)式乘以 $\theta_{(k)} = \frac{\tau_{(k)}}{\tau_{(k-1)}}$, (2.3.6) 式乘以 $\theta_{(k)}^2$, 再相加, 得到:

$$\begin{aligned} &\langle \theta_{(k)}(x^{(k)} - x^{(k-1)}) + \tau_{(k)}A^*y^{(k)}, x^{(k+1)} - x^{(k)} \rangle \\ &+ \langle \theta_{(k)}(x^{(k)} - x^{(k-1)}) + \tau_{(k)}A^*y^{(k)}, \theta_{(k)}(x^{(k-1)} - x^{(k)}) \rangle \\ &\geq \tau_{(k)}(g(x^{(k)}) - g(x^{(k+1)})) + \tau_{(k)}\theta_{(k)}(g(x^{(k)}) - g(x^{(k-1)})) - \tau_{(k)}e_{(k-1)} - \theta_{(k)}\tau_{(k)}e_{(k-1)}. \end{aligned}$$

由于 $\bar{x}^{(k)} = x^{(k)} + \theta_{(k)}(x^{(k)} - x^{(k-1)})$, 则:

$$\begin{aligned} &\langle \bar{x}^{(k)} - x^{(k)} + \tau_{(k)}A^*y^{(k)}, x^{(k+1)} - \bar{x}^{(k)} \rangle \\ &\geq \tau_{(k)}((1 + \theta_{(k)})g(x^{(k)}) - g(x^{(k+1)}) - \theta_{(k)}g(x^{(k-1)})) - \tau_{(k)}e_{(k-1)} - \theta_{(k)}\tau_{(k)}e_{(k-1)}. \end{aligned} \quad (2.3.7)$$

又因为

$$\tau_{(k)} \langle A^* \hat{y}^{(k+1)} - A^* y^\#, \bar{x}^{(k)} - x^* \rangle = \tau_{(k)} \langle A \bar{x}^{(k)} - A x^\#, \hat{y}^{(k+1)} - y^* \rangle. \quad (2.3.8)$$

将(2.3.3)式, (2.3.2)式, (2.3.7)式, (2.3.8)式相加可得到:

$$\begin{aligned} & \langle x^{(k+1)} - x^{(k)}, x^\# - x^{(k+1)} \rangle + \frac{1}{\beta} \langle \hat{y}^{(k+1)} - y^{(k)}, y^\# - \hat{y}^{(k+1)} \rangle + \langle \bar{x}^{(k)} - x^{(k)}, x^{(k+1)} - \bar{x}^{(k)} \rangle \\ & + \tau_{(k)} \langle A^* \hat{y}^{(k+1)} - A^* y^{(k)}, \bar{x}^{(k)} - x^{(k+1)} \rangle - \tau_{(k)} \langle A^* y^\#, \bar{x}^{(k)} - x^\# \rangle + \tau_{(k)} \langle A x^\#, \hat{y}^{(k+1)} - y^\# \rangle \\ & \geq \tau_{(k)} (f^*(\hat{y}^{(k+1)}) - f^*(y^\#) + (1 + \theta_{(k)})g(x^{(k)}) - \theta_{(k)}g(x^{(k-1)}) - g(x^\#)) \\ & \quad - \tau_{(k)} e_{(k-1)} - \theta_{(k)} \tau_{(k)} e_{(k-1)} - \tau_{(k)} e_{(k)}. \end{aligned} \quad (2.3.9)$$

由(2.1.3)式可得:

$$\begin{aligned} & (1 + \theta_{(k)})g(x^{(k)}) - \theta_{(k)}g(x^{(k-1)}) - g(x^\#) + \langle A^* y^\#, \bar{x}^{(k)} - x^\# \rangle \\ & = (1 + \theta_{(k)})(g(x^{(k)}) - g(x^\#) + \langle A^* y^\#, x^{(k)} - x^\# \rangle) \\ & \quad - \theta_{(k)}(g(x^{(k-1)}) - g(x^\#) + \langle A^* y^\#, x^{(k-1)} - x^\# \rangle) \\ & = (1 + \theta_{(k)})P_{x^\#, y^\#}(x^{(k)}) - \theta_{(k)}P_{x^\#, y^\#}(x^{(k-1)}). \end{aligned} \quad (2.3.10)$$

结合(2.1.4)式, (2.3.9)式和(2.3.10)式, 可得:

$$\begin{aligned} & \langle x^{(k+1)} - x^{(k)}, x^\# - x^{(k+1)} \rangle + \frac{1}{\beta} \langle \hat{y}^{(k+1)} - y^{(k)}, y^\# - \hat{y}^{(k+1)} \rangle \\ & + \langle \bar{x}^{(k)} - x^{(k)}, x^{(k+1)} - \bar{x}^{(k)} \rangle + \tau_{(k)} \langle A^* \hat{y}^{(k+1)} - A^* y^{(k)}, \bar{x}^{(k)} - x^{(k+1)} \rangle \\ & \geq \tau_{(k)} ((1 + \theta_{(k)})P_{x^\#, y^\#}(x^{(k)}) - \theta_{(k)}P_{x^\#, y^\#}(x^{(k-1)}) + D_{x^\#, y^\#}(\hat{y}^{(k+1)})) \\ & \quad - \tau_{(k)} e_{(k-1)} - \theta_{(k)} \tau_{(k)} e_{(k-1)} - \tau_{(k)} e_{(k)}. \end{aligned} \quad (2.3.11)$$

记 $\varepsilon_{(k)} = \tau_{(k)} ((1 + \theta_{(k)})P_{x^\#, y^\#}(x^{(k)}) - \theta_{(k)}P_{x^\#, y^\#}(x^{(k-1)}) + D_{x^\#, y^\#}(\hat{y}^{(k+1)}))$.

根据cosin rule:

$$2\langle a - b, c - a \rangle = \|b - c\|^2 - \|a - b\|^2 - \|a - c\|^2.$$

可得:

$$\begin{aligned} & \frac{1}{2} (\|x^{(k)} - x^\#\|^2 - \|x^{(k+1)} - x^\#\|^2 - \|x^{(k+1)} - x^{(k)}\|^2) \\ & + \frac{1}{2\beta} (\|y^{(k)} - y^\#\|^2 - \|\hat{y}^{(k+1)} - y^\#\|^2 - \|\hat{y}^{(k+1)} - y^{(k)}\|^2) \\ & + \frac{1}{2} (\|x^{(k+1)} - x^{(k)}\|^2 - \|\bar{x}^{(k)} - x^{(k)}\|^2 - \|x^{(k+1)} - \bar{x}^{(k)}\|^2) \\ & + \tau_{(k)} \langle A^* \hat{y}^{(k+1)} - A^* y^{(k)}, \bar{x}^{(k)} - x^{(k+1)} \rangle \\ & \geq \varepsilon_{(k)} - \tau_{(k)} e_{(k-1)} - \theta_{(k)} \tau_{(k)} e_{(k-1)} - \tau_{(k)} e_{(k)}. \end{aligned} \quad (2.3.12)$$

根据Cauchy-Schwarz不等式和搜索步长终止条件(2b)可得:

$$\begin{aligned} \tau_{(k)} \langle A^* \hat{y}^{(k+1)} - A^* y^{(k)}, \bar{x}^{(k)} - x^{(k+1)} \rangle &\leq \tau_{(k)} \|A^* \hat{y}^{(k+1)} - A^* y^{(k)}\| \|\bar{x}^{(k)} - x^{(k+1)}\| \\ &\leq \frac{\delta}{\sqrt{\beta}} \|\hat{y}^{(k+1)} - y^{(k)}\| \|\bar{x}^{(k)} - x^{(k+1)}\| \\ &\leq \frac{1}{2} \|\bar{x}^{(k)} - x^{(k+1)}\|^2 + \frac{\delta^2}{2\beta} \|\hat{y}^{(k+1)} - y^{(k)}\|^2. \end{aligned}$$

将上式代入(2.3.12)式得到:

$$\begin{aligned} &\frac{1}{2} (\|x^{(k)} - x^\# \|^2 - \|x^{(k+1)} - x^\# \|^2) + \frac{1}{2\beta} (\|y^{(k)} - y^\# \|^2 - \|\hat{y}^{(k+1)} - y^\# \|^2) \\ &\quad - \frac{1}{2} \|\bar{x}^{(k)} - x^{(k+1)}\|^2 - \frac{1-\delta^2}{2\beta} \|\hat{y}^{(k+1)} - y^{(k)}\|^2 \\ &\geq \varepsilon_{(k)} - \tau_{(k)} e_{(k-1)} - \theta_{(k)} \tau_{(k)} e_{(k-1)} - \tau_{(k)} e_{(k)}. \end{aligned} \quad (2.3.13)$$

下面给出本节最重要的结论.

定理2.3.1. 设 $(x^{(k)}, y^{(k)})$ 是由非精确自适应步随机原始对偶算法迭代产生的序列, $(x^\#, y^\#)$ 是问题(2.1.2)的鞍点, 则遍历序列 $(X_{(N)}, Y_{(N)})$ 的原始对偶间隙 $P(X_{(N)}) + D(Y_{(N)})$ 满足

$$\begin{aligned} E\{P(X_{(N)}) + D(Y_{(N)})\} &\leq \frac{1}{S_{(N)}} (\Delta_{(1)} + \tau_{(1)} \theta_{(1)} P(x^{(0)}) + \tau_{(1)} D^p(y^{(1)})) + 3\tau_{(0)} \sum_{k=0}^N e_{(k)} \\ &\leq \frac{1}{N} \cdot \left(\frac{1}{\tau} (\Delta_{(1)} + \tau_{(1)} \theta_{(1)} P(x^{(0)}) + \tau_{(1)} D^p(y^{(1)})) \right) + 3\tau_{(0)} \sum_{k=0}^N e_{(k)}, \end{aligned}$$

$$\text{其中, } X_{(N)} = \frac{\tau_{(N)}(1+\theta_{(N)})x^{(N)} + \sum_{k=1}^{N-1} (\tau_{(k)}(1+\theta_{(k)}) - \tau_{(k+1)}\theta_{(k+1)})x^{(k)}}{\tau_{(1)}\theta_{(1)} + S_{(N)}}, \quad Y_{(N)} = \frac{\sum_{k=1}^N \tau_{(k)}y^{(k+1)}}{\sum_{k=1}^N \tau_{(k)}}, \quad S_{(N)} = \sum_{k=1}^N \tau_{(k)},$$

$$\Delta_{(1)} = E\left\{ \frac{1}{2} \|x^{(1)} - x^\# \|^2 + \frac{1}{2\beta} \|y^{(1)} - y^\# \|^2 \right\}.$$

证明:考虑对偶变量以一定的概率随机更新, 即:

$$y_i^{(k+1)} = \begin{cases} \text{prox}_{f_i^*}^{\beta\tau_{(k)}}(y^{(k)} + \beta\tau_{(k)}A_i\bar{x}^{(k)}) & i \in S_{(k)}, \\ y_i^{(k)}, & i \notin S_{(k)}. \end{cases}$$

对任意映射 φ , 假设 $y_i^{(k+1)}$ 以 p_i 的概率更新, 则:

$$E^{(k+1)}\varphi(y_i^{(k+1)}) = p_i\varphi(\hat{y}_i^{(k+1)}) + (1 - p_i)\varphi(y_i^{(k)}).$$

即:

$$\begin{aligned} \varphi(\hat{y}_i^{(k+1)}) &= \frac{1}{p_i} E^{(k+1)}\varphi(y_i^{(k+1)}) - \left(\frac{1}{p_i} - 1\right)\varphi(y_i^{(k)}) \\ &= \left(\frac{1}{p_i} - 1\right)E^{(k+1)}\varphi(y_i^{(k+1)}) - \left(\frac{1}{p_i} - 1\right)\varphi(y_i^{(k)}) + E^{(k+1)}\varphi(y_i^{(k+1)}). \end{aligned}$$

因此

$$\begin{aligned} \|\hat{y}^{(k+1)} - y^\# \|^2 &= E^{(k+1)}\|y^{(k+1)} - y^\# \|^2_Q - \|y^{(k)} - y^\# \|^2_{(Q-I)}, \\ \|\hat{y}^{(k+1)} - y^{(k)} \|^2 &= E^{(k+1)}\|y^{(k+1)} - y^{(k)} \|^2_Q, \\ D(\hat{y}^{(k+1)}) &= E^{(k+1)}D^p(y^{(k+1)}) - D^p(y^{(k)}) + E^{(k+1)}D(y^{(k+1)}). \end{aligned} \quad (2.3.14)$$

其中, $D^p(y^{(k)}) = \sum_{i=1}^n (\frac{1}{p_i} - 1)D_{i(x^\#, y^\#)}(y_i^{(k)})$, $Q = \text{diag}(\frac{1}{p_1}I, \dots, \frac{1}{p_n}I)$, $E^{(k+1)}$ 表示对变量 $y^{(k+1)}$ 求期望.

将(2.3.14)式代入到(2.3.13)式得到:

$$\begin{aligned} & \frac{1}{2} \|x^{(k)} - x^\#\|^2 + \frac{1}{2\beta} \|y^{(k)} - y^\#\|_Q^2 - \frac{1}{2} \|\bar{x}^{(k)} - x^{(k)}\|^2 \\ & + \tau_{(k)}\theta_{(k)}P(x^{(k-1)}) - \tau_{(k)}(1 + \theta_{(k)})P(x^{(k)}) + \tau_{(k)}D^p(y^{(k)}) \\ \geq & \frac{1}{2} \|x^{(k+1)} - x^\#\|^2 + \frac{1}{2\beta} E^{(k+1)} \|y^{(k+1)} - y^\#\|_Q^2 + \frac{1-\delta^2}{2\beta} E^{(k+1)} \|y^{(k+1)} - y^{(k)}\|_Q^2 \\ & + \tau_{(k)}E^{(k+1)}(D^p(y^{(k+1)}) + D(y^{(k+1)})) - \tau_{(k)}e_{(k-1)} - \theta_{(k)}\tau_{(k)}e_{(k-1)} - \tau_{(k)}e_{(k)} \\ = & E^{(k+1)} \left\{ \frac{1}{2} \|x^{(k+1)} - x^\#\|^2 + \frac{1}{2\beta} \|y^{(k+1)} - y^\#\|_Q^2 + \frac{1-\delta^2}{2\beta} \|y^{(k+1)} - y^{(k)}\|_Q^2 \right. \\ & \left. + \tau_{(k)}(D^p(y^{(k+1)}) + D(y^{(k+1)})) - \tau_{(k)}e_{(k-1)} - \theta_{(k)}\tau_{(k)}e_{(k-1)} - \tau_{(k)}e_{(k)} \right\}. \end{aligned}$$

两边同时对 $y^{(k)}$ 求期望, 得到:

$$\begin{aligned} & E^{(k)} \left\{ \tau_{(k)}\theta_{(k)}P(x^{(k-1)}) - \tau_{(k)}(1 + \theta_{(k)})P(x^{(k)}) + \tau_{(k)}D^p(y^{(k)}) \right. \\ & \left. + \frac{1}{2} \|x^{(k)} - x^\#\|^2 + \frac{1}{2\beta} \|y^{(k)} - y^\#\|_Q^2 - \frac{1}{2} \|\bar{x}^{(k)} - x^{(k)}\|^2 \right\} \\ \geq & E^{(k,k+1)} \left\{ \frac{1}{2} \|x^{(k+1)} - x^\#\|^2 + \frac{1}{2\beta} \|y^{(k+1)} - y^\#\|_Q^2 + \frac{1-\delta^2}{2\beta} \|y^{(k+1)} - y^{(k)}\|_Q^2 \right. \\ & \left. + \tau_{(k)}(D^p(y^{(k+1)}) + D(y^{(k+1)})) - \tau_{(k)}e_{(k-1)} - \theta_{(k)}\tau_{(k)}e_{(k-1)} - \tau_{(k)}e_{(k)} \right\}. \end{aligned}$$

方便起见, 记 E 为对所有对偶变量 y 求期望, 则对任意的 $k \geq 0$ 都有:

$$\begin{aligned} & E \left\{ \frac{1}{2} \|x^{(k)} - x^\#\|^2 + \frac{1}{2\beta} \|y^{(k)} - y^\#\|_Q^2 + \tau_{(k)}\theta_{(k)}P(x^{(k-1)}) \right. \\ & \left. - \tau_{(k)}(1 + \theta_{(k)})P(x^{(k)}) + \tau_{(k)}D^p(y^{(k)}) \right\} \\ \geq & E \left\{ \frac{1}{2} \|x^{(k+1)} - x^\#\|^2 + \frac{1}{2\beta} \|y^{(k+1)} - y^\#\|_Q^2 + \tau_{(k)}(D^p(y^{(k+1)}) + D(y^{(k+1)})) \right. \\ & \left. - \tau_{(k)}e_{(k-1)} - \theta_{(k)}\tau_{(k)}e_{(k-1)} - \tau_{(k)}e_{(k)} \right\}. \end{aligned} \tag{2.3.15}$$

令 $\Delta_{(k)} = E \left\{ \frac{1}{2} \|x^{(k)} - x^\#\|^2 + \frac{1}{2\beta} \|y^{(k)} - y^\#\|_Q^2 \right\}$, 则(2.3.15)式可以表示为:

$$\begin{aligned} \Delta_{(k)} \geq & \Delta_{(k+1)} + E \left\{ \tau_{(k)}((1 + \theta_{(k)})P(x^{(k)}) - \theta_{(k)}P(x^{(k-1)}) - D^p(y^{(k)}) + D^p(y^{(k+1)})) \right. \\ & \left. + D(y^{(k+1)}) - \tau_{(k)}e_{(k-1)} - \theta_{(k)}\tau_{(k)}e_{(k-1)} - \tau_{(k)}e_{(k)} \right\}. \end{aligned} \tag{2.3.16}$$

取(2.3.16)式中的 $k = 1, 2, \dots, N$, 相加得到:

$$\begin{aligned} \Delta_{(1)} - \Delta_{(N+1)} \geq & \sum_{k=1}^N E \left\{ \tau_{(k)}((1 + \theta_{(k)})P(x^{(k)}) - \theta_{(k)}P(x^{(k-1)}) + \tau_{(k)}(D^p(y^{(k+1)})) \right. \\ & \left. + D(y^{(k+1)}) - D^p(y^{(k)}) - \tau_{(k)}e_{(k-1)} - \theta_{(k)}\tau_{(k)}e_{(k-1)} - \tau_{(k)}e_{(k)} \right\}. \end{aligned} \tag{2.3.17}$$

由 P 的凸性可得:

$$\begin{aligned}
& E \left\{ \sum_{k=1}^N [\tau_{(k)}(1 + \theta_{(k)})P(x^{(k)}) - \tau_{(k)}\theta_{(k)}P(x^{(k-1)})] \right\} \\
&= E \left\{ \sum_{k=1}^{N-1} [\tau_{(k)}(1 + \theta_{(k)}) - \tau_{(k+1)}\theta_{(k+1)}]P(x^{(k)}) + \tau_{(N)}(1 + \theta_{(N)})P(x^{(N)}) - \tau_{(1)}\theta_{(1)}P(x^{(0)}) \right\} \\
&\geq E \left\{ (\tau_{(1)}\theta_{(1)} + S_{(N)})P\left(\frac{\tau_{(N)}(1+\theta_{(N)})x^{(N)} + \sum_{k=1}^{N-1} (\tau_{(k)}(1+\theta_{(k)}) - \tau_{(k+1)}\theta_{(k+1)})x^{(k)}}{\tau_{(1)}\theta_{(1)} + S_{(N)}}\right) \right\} - \tau_{(1)}\theta_{(1)}P(x^{(0)}) \\
&\geq E \{S_N P(X_N)\} - \tau_1\theta_1 P(x^0),
\end{aligned} \tag{2.3.18}$$

其中, $\sum_{k=1}^N \tau_{(k)} = S_{(N)}$, $X_{(N)} = \frac{\tau_{(N)}(1+\theta_{(N)})x^{(N)} + \sum_{k=1}^{N-1} (\tau_{(k)}(1+\theta_{(k)}) - \tau_{(k+1)}\theta_{(k+1)})x^{(k)}}{\tau_{(1)}\theta_{(1)} + S_{(N)}}$.

同理, 由 D 的凸性可得到:

$$E \left\{ \sum_{k=1}^N \tau_{(k)} D(y^{(k+1)}) \right\} \geq E \left\{ \left(\sum_{k=1}^N \tau_{(k)} \right) \cdot D\left(\frac{\sum_{k=1}^N \tau_{(k)} y^{(k+1)}}{\sum_{k=1}^N \tau_{(k)}} \right) \right\} = E \{S_{(N)} \cdot D(Y_{(N)})\}, \tag{2.3.19}$$

其中, $Y_{(N)} = \frac{\sum_{k=1}^N \tau_{(k)} y^{(k+1)}}{\sum_{k=1}^N \tau_{(k)}}$.

又因为 $\tau_{(k)} \geq \tau_{(k+1)}$, 由函数 $D^p(\cdot)$ 的凸性和非负性可得:

$$\begin{aligned}
& \sum_{k=1}^N E \{ \tau_k (D^p(y^{(k+1)}) - D^p(y^{(k)})) \} \\
&= E \left\{ \tau_{(N)} D^p(y^{(N+1)}) - \tau_{(1)} D^p(y^{(1)}) + \sum_{k=1}^{N-1} (\tau_{(k)} - \tau_{(k+1)}) D^p(y^{(k+1)}) \right\} \\
&\geq E \left\{ (\tau_{(1)} - \tau_{(N)}) D^p\left(\frac{\sum_{k=1}^{N-1} (\tau_{(k)} - \tau_{(k+1)}) y^{(k+1)}}{(\tau_{(1)} - \tau_{(N)})}\right) + \tau_{(N)} D^p(y^{(N+1)}) - \tau_{(1)} D^p(y^{(1)}) \right\} \\
&\geq E \{ \tau_{(N)} D^p(y^{(N+1)}) \} - \tau_{(1)} D^p(y^{(1)}) \\
&\geq -\tau_{(1)} D^p(y^{(1)}).
\end{aligned} \tag{2.3.20}$$

将(2.3.18)式, (2.3.19)式, (2.3.20)式代入(2.3.17)式可得:

$$\begin{aligned}
E \{P(X_{(N)}) + D(Y_{(N)})\} &\leq \frac{1}{S_{(N)}} \{ (\Delta_{(1)} - \Delta_{(N+1)} + \tau_{(1)}\theta_{(1)}P(x^{(0)}) + \tau_{(1)}D^p(y^{(1)})) \\
&\quad + \sum_{k=1}^N (\tau_{(k)}e_{(k-1)} + \theta_{(k)}\tau_{(k)}e_{(k-1)} + \tau_{(k)}e_{(k)}) \} \\
&\leq \frac{1}{S_{(N)}} \{ (\Delta_{(1)} + \tau_{(1)}\theta_{(1)}P(x^{(0)}) + \tau_{(1)}D^p(y^{(1)})) \\
&\quad + \sum_{k=1}^N (\tau_{(k)}e_{(k-1)} + \theta_{(k)}\tau_{(k)}e_{(k-1)} + \tau_{(k)}e_{(k)}) \}.
\end{aligned} \tag{2.3.21}$$

根据引理2.3.1中的(2), 对任意的 $k \geq 0$, 存在 τ , 使得 $\tau_{(k)} \geq \tau$, 则:

$$S_{(N)} = \sum_{k=1}^N \tau_{(k)} \geq N \cdot \tau. \quad (2.3.22)$$

那么, (2.3.21)式可以进一步放缩为:

$$\begin{aligned} E \{P(X_{(N)}) + D(Y_{(N)})\} &\leq \frac{1}{S_{(N)}} \left\{ (\Delta_{(1)} + \tau_{(1)}\theta_{(1)}P(x^{(0)}) + \tau_{(1)}D^p(y^{(1)})) + 3\tau_{(0)} \sum_{k=0}^N e_{(k)} \right\} \\ &\leq \frac{1}{N} \left\{ \frac{1}{\tau} (\Delta_{(1)} + \tau_{(1)}\theta_{(1)}P(x^{(0)}) + \tau_{(1)}D^p(y^{(1)})) + 3\tau_{(0)} \sum_{k=0}^N e_{(k)} \right\}. \end{aligned}$$

若误差序列是一个有界和序列, 那么遍历序列的原始对偶间隙按期望收敛.

引理2.3.3. [7] 对任意 $\alpha \in (0, 1)$, 记 $s_N := \sum_{k=1}^N k^{-\alpha}$, 则

$$s_N = \mathcal{O}(N^{1-\alpha}).$$

与文献 [7]中的推论1和文献 [10]中的定理4类似, 以下定理阐明了误差序列对非精确原始对偶混合梯度算法的收敛速率产生的影响.

定理2.3.2. 假设 $(x^\#, y^\#)$ 是问题(2.1.2)的鞍点, $(x_{(N)}, y_{(N)})$ 是由非精确自适应步随机原始对偶算法迭代产生的遍历序列, 非精确邻近算子的误差序列满足 $e_{(k)} = \mathcal{O}(k^{-\alpha})$, $k = 1, 2, \dots$, 则

$$E \{P(X_{(N)}) + D(Y_{(N)})\} = \begin{cases} \mathcal{O}(N^{-1}), & \alpha > 1, \\ \mathcal{O}(\ln(N)/N), & \alpha = 1, \\ \mathcal{O}(N^{-\alpha}), & \alpha \in (0, 1). \end{cases}$$

$$\text{其中, } X_{(N)} = \frac{\tau_{(N)}(1+\theta_{(N)})x^{(N)} + \sum_{k=1}^{N-1} (\tau_{(k)}(1+\theta_{(k)}) - \tau_{(k+1)}\theta_{(k+1)})x^{(k)}}{\tau_{(1)}\theta_{(1)} + S_{(N)}}, \quad Y_{(N)} = \frac{\sum_{k=1}^N \tau_{(k)}y^{(k+1)}}{\sum_{k=1}^N \tau_{(k)}}.$$

证明: 由定理2.3.1可得:

$$\begin{aligned} &E \{P(X_{(N)}) + D(Y_{(N)})\} \\ &\leq \frac{1}{S_{(N)}} \left\{ (\Delta_{(1)} + \tau_{(1)}\theta_{(1)}P(x^{(0)}) + \tau_{(1)}D^p(y^{(1)})) + 3\tau_{(0)} \sum_{k=0}^N e_{(k)} \right\} \\ &\leq \frac{1}{N} \left\{ \frac{1}{\tau} (\Delta_{(1)} + \tau_{(1)}\theta_{(1)}P(x^{(0)}) + \tau_{(1)}D^p(y^{(1)})) + 3\tau_{(0)} \sum_{k=0}^N e_{(k)} \right\} \\ &\leq \frac{1}{N} \left\{ \frac{1}{\tau} (\Delta_{(1)} + \tau_{(1)}\theta_{(1)}P(x^{(0)}) + \tau_{(1)}D^p(y^{(1)})) + 3\tau_{(0)}e_{(0)} + 3\tau_{(0)} \sum_{k=1}^N e_{(k)} \right\}. \end{aligned} \quad (2.3.23)$$

显然存在有限数 $C > 0$, 使得

$$\Delta_{(1)} + \tau_{(1)}\theta_{(1)}P(x^{(0)}) + \tau_{(1)}D^p(y^{(1)}) + 3\tau_{(0)}e_{(0)} \leq C.$$

当 $\alpha > 1$ 时, $\sum_{k=1}^N e_{(k)}$ 收敛, 即存在有限数 $C_1 > 0$, 使得

$$3\tau_{(0)} \sum_{k=1}^N e_{(k)} \leq C_1.$$

因此

$$E \{P(X_{(N)}) + D(Y_{(N)})\} \leq \frac{C + C_1}{N}.$$

此时, $P(X_{(N)}) + D(Y_{(N)})$ 按期望以 $O(\frac{1}{N})$ 的速率收敛.

当 $\alpha = 1$ 时, $\frac{1}{k}$ 在区间 $[1, N + 1]$ 上单调递减, 因此:

$$\int_1^{N+1} \frac{1}{x} dx < \sum_{k=1}^N \frac{1}{k} < 1 + \int_1^N \frac{1}{x} dx.$$

对不等式两边求积分:

$$0 < \ln(N + 1) < \sum_{k=1}^N \frac{1}{k} < 1 + \ln(N).$$

考虑到对于任意 $x > 2$, $\ln x > 1$ 成立, 因此对任意 $N > 2$, 我们可以得到:

$$0 < \sum_{k=1}^N \frac{1}{k} < 2 \ln(N).$$

因此, $\sum_{k=0}^N e_{(k)} = O(\ln(N))$, 于是 $E \{P(X_{(N)}) + D(Y_{(N)})\} = O(\ln(N)/N)$.

当 $\alpha \in (0, 1)$ 时, 由引理 2.3.3 可知:

$$\sum_{k=0}^N e_{(k)} \leq O(N^{1-\alpha}).$$

根据 (2.3.23) 得到, $E \{P(X_{(N)}) + D(Y_{(N)})\} = O(N^{-\alpha})$.

2.4. 数值实验

正电子发射断层成像(PET), 是核医学领域比较先进的临床检查影像技术. PET 成像主要基于正电子发射原理, 使用放射性示踪剂在体内发出正电子, 当正电子与体内的电子相遇时, 会湮灭产生伽玛射线, 通过对伽玛射线的探测和记录, PET 系统重建出放射性示踪剂在人体组织中的分布情况. 设示踪剂分布(图像)空间为 $X = R^{d_1 \times d_2}$, 其中 $d_1 = d_2 = 250$, 数据空间为 $Y = \prod_{i=1}^n Y_i$, $Y_i = R^{|B_i|}$, 其中 $B_i \subset \{1, 2, \dots, 200 \cdot 250\}$, 对任意 $i \neq j$, $B_i \cap B_j = \emptyset$, 且 $\bigcup_{i=1}^n B_i = \{1, 2, \dots, 200 \cdot 250\}$, 则正电子发射断层成像(PET) 可抽象为如下标准形式:

$$\min_{x \in X} \left\{ \Phi(x) := \sum_{i=1}^n f_i(A_i x) + g(x) \right\}.$$

其中, 有界线性算子 $A : X \rightarrow Y$ 是200 个方向上的线积分, 且每个方向由250 个线积分构成; 函数 $g(x) = \alpha \|\nabla x\|_{2,1} + \iota_{\geq 0}(x)$, $\nabla x = (\nabla_1 x, \nabla_2 x) \in R^{d_1 d_2 \times 2}$, 范数定义为 $\|\nabla x\|_{2,1} = \sum_j \sqrt{(\nabla_1 x_j)^2 + (\nabla_2 x_j)^2}$; 函数

$$f_i(y) = \begin{cases} \sum_{j \in B_i} y_j + r_j - b_j + b_j \log\left(\frac{b_j}{y_j + r_j}\right), & y_j + r_j > 0, \\ \infty, & \text{else.} \end{cases}$$

则 $f_i(y)$ 的共轭可表示为:

$$f_i^*(z) = \sum_{j \in B_i} \begin{cases} -z_j r_j - b_j \log(1 - z_j), & \text{if } z_j \leq 1 \text{ and } (b_j = 0 \text{ or } z_j < 1), \\ \infty, & \text{else.} \end{cases}$$

$f_i(y)$ 的近似算子可表示为:

$$[\text{prox}_{f_i^*}^{\sigma_i}(z)]_j = \frac{1}{2}(z_j + 1 + \sigma_i r_j - \sqrt{(z_j - 1 + \sigma_i r_j)^2 + 4\sigma_i b_j}).$$

由于 g 的近似算子没有显式表达式, 根据算法 2, 我们需要非精确求解问题:

$$u^k \approx_2^{e(k-1)} \text{prox}_g^{\tau(k-1)}(u^{(k-1)} - \tau_{(k-1)} A^* y^{(k)}).$$

然后根据推论 2.1.2, 采用 FISTA 算法 [17] 近似求解该问题的对偶问题, 使得对应的对偶间隙在预先设定的误差值 $e(k) = O(k^{-\alpha})$ 之内.

本文数值实验运行环境为 AMD Ryzen 5 5600U 2.30 GHz 处理器, 6.0 GB 内存和 64 位 Windows 11 操作系统, 运用了 pycharm 17.0.4 验证了所提算法的可行性.

针对上述问题, 假设随机抽样方式为统一概率抽样, 并设置算法参数为: $\delta = 0.85$, $\mu = 0.95$, $\beta = 5$, $\tau_{(0)} = 0.3$, $\sigma_{(1)} = \beta \cdot \tau = 1.5$.

记对偶变量完全迭代更新一次为 1 个轮次, 将对偶变量指标 $\{1, 2, \dots, 200\}$ 分为 10 个子集, 30 个子集, 取 $\alpha = 1.1$, $\alpha = 0.7$ 进行测试, 算法 2 经过 100 个轮次迭代之后得到了如下结果:

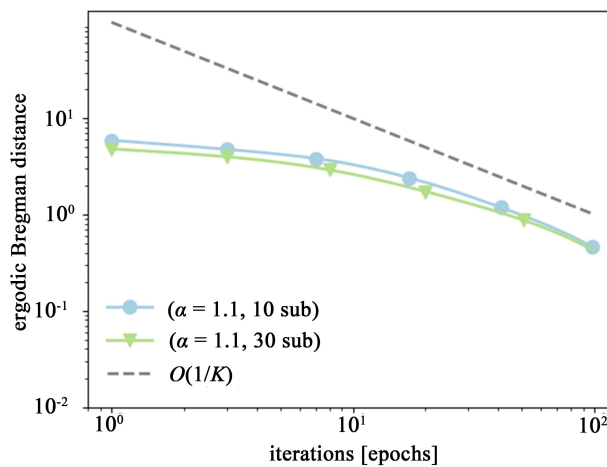


Figure 1. The gap of primal-dual in ergodic sense

图 1. 遍历序列的原始对偶间隙

图 1 以遍历序列的原始对偶间隙为纵坐标, 迭代次数(轮次) 为横坐标. 从图中可以发现, 当误差序列精度满足 $e_{(k)} = O(k^{-\alpha})$, $\alpha = 1.1$ 时, 本文提出的算法 2 以 $O(1/N)$ 的速率收敛, 这也印证了定理 2.3.2 的结论. 此外, 当 ASPDHG 初始参数一致时, 子集数取得越多, 即每次迭代更新的对偶变量越少, 收敛速度稍快.

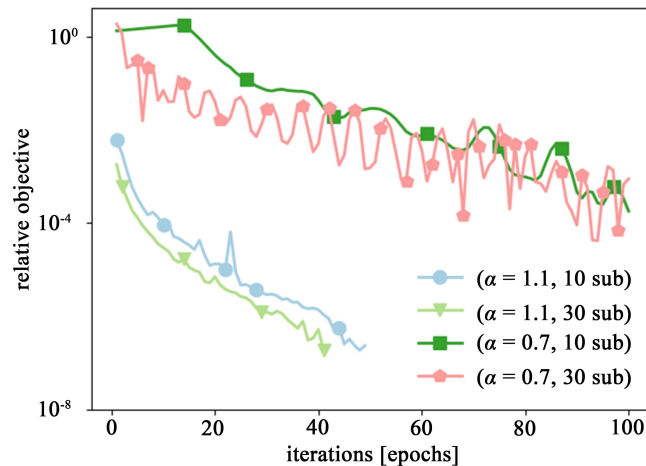


Figure 2. The relative distance between iteration point and the saddle point

图 2. 迭代点与鞍点之间的相对距离

图 2 中所展示的 relative objective 被定义为 $[\Phi(x^N) - \Phi(x^*)] / [\Phi(x^0) - \Phi(x^*)]$. 通过这一数值实验, 我们可以清晰地观察到误差序列的精度 (表示为 α) 对算法收敛速度的重要影响. 具体来说, 误差较小, 即 α 的取值较大时, 算法的收敛速度会显著加快. 此外, 在保持初始参数一致的情况下, 增加子集的数量意味着每次迭代中更新的对偶变量数量减少, 这同样会促使算法收敛速度的提升.

此外, 分别取 $\alpha = 1.1$, $\alpha = 0.7$, 在经过 40 个轮次迭代之后, PET 图像重建的结果从左到右依次为:

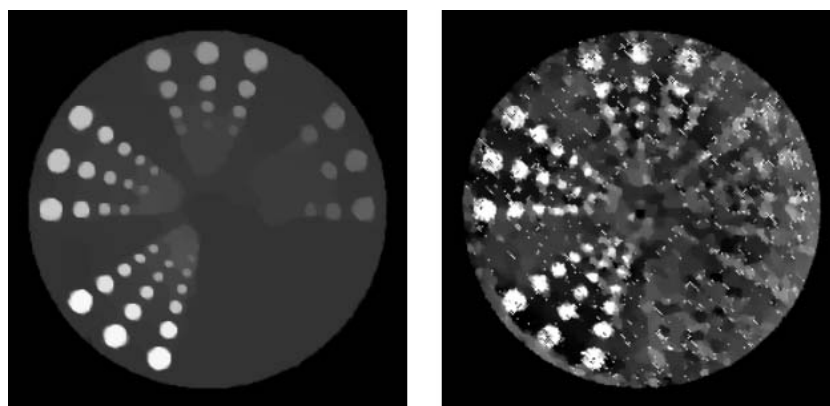


Figure 3. PET problem image reconstruction

图 3. PET 问题图像重建的效果

通过对放射性示踪剂产生的伽玛射线进行线积分测量, 重构出放射性示踪剂分布情况的图像, 如图 3 所示. 在历经 40 轮次的迭代后, 我们发现误差序列较小的成像效果更为清晰和精细. 这一现象

深刻揭示了在实际应用中,合理控制误差序列的取值对于提高成像质量的重要性.

3. 结论及展望

本文考虑了非精确求解目标函数邻近算子的情况,提出了一种非精确自适应步随机原始对偶算法.并深入探讨了误差序列选取方式对算法收敛速率的影响,以及非精确求解在实际应用中的实现方式.这些讨论在实际应用中具有重要的意义,例如当目标函数的邻近算子没有显式表达式或者精确求解难度较大时,非精确求解方法展现出其独特的灵活性和实用性.这些研究不仅丰富了算法的理论基础,也为实际应用提供了更加全面和深入的指导.

面对实际问题的多样性和复杂性,原始对偶算法的研究仍然充满无限的可能性.例如:未来的研究可以进一步探索如何根据目标函数本身的特性,结合线搜索策略,自适应地选择非精确求解的误差精度.这样的策略有助于避免过度求解子问题,从而提高算法的整体效率.

参考文献

- [1] Chambolle, A. and Pock, T. (2011) A First-Order Primal-Dual Algorithm for Convex Problems with Applications to Imaging. *Journal of Mathematical Imaging and Vision*, **40**, 120-145. <https://doi.org/10.1007/s10851-010-0251-1>
- [2] Chambolle, A. and Pock, T. (2016) An Introduction to Continuous Optimization for Imaging. *Acta Numerica*, **25**, 161-319. <https://doi.org/10.1017/S096249291600009X>
- [3] Jiu, M. and Pustelnik, N. (2021) A Deep Primal-Dual Proximal Network for Image Restoration. *IEEE Journal of Selected Topics in Signal Processing*, **15**, 190-203. <https://doi.org/10.1109/JSTSP.2021.3054506>
- [4] Chen, S., Halimi, A., Ren, X., *et al.* (2020) Learning Non-Local Spatial Correlations to Restore Sparse 3D Single-Photon Data. *IEEE Transactions on Image Processing*, **29**, 3119-3131. <https://doi.org/10.1109/TIP.2019.2957918>
- [5] Goldstein, T., Li, M. and Yuan, X. (2015) Adaptive Primal-Dual Splitting Methods for Statistical Learning and Image Processing. In: Cortes, C., Lawrence, N., Lee, D., Sugiyama, M. and Garnett, R., Eds., *Advances in Neural Information Processing Systems*, Vol. 28.
- [6] Eckstein, J. and Bertsekas, D.P. (1992) On the Douglas—Rachford Splitting Method and the Proximal Point Algorithm for Maximal Monotone Operators. *Mathematical Programming*, **55**, 293-318. <https://doi.org/10.1007/BF01581204>
- [7] Rasch, J. and Chambolle, A. (2020) Inexact First-Order Primal-Dual Algorithms. *Computational Optimization and Applications*, **76**, 381-430. <https://doi.org/10.1007/s10589-020-00186-y>
- [8] Ng, M.K., Wang, F. and Yuan, X. (2011) Inexact Alternating Direction Methods for Image Recovery. *Siam Journal on Scientific Computing*, **33**, 1643-1668. <https://doi.org/10.1137/100807697>

-
- [9] Villa, S., Salzo, S., Baldassarre, L., *et al.* (2013) Accelerated and Inexact Forward-Backward Algorithms. *Society for Industrial and Applied Mathematics*, **23**, 1607-1633.
<https://doi.org/10.1137/110844805>
- [10] Salzo, S. and Villa, S. (2012) Inexact and Accelerated Proximal Point Algorithms. *Journal of Convex Analysis*, **19**, 1167-1192.
- [11] Lin, H., Mairal, J. and Harchaoui, Z.O. (2017) Catalyst Acceleration for First-Order Convex Optimization: From Theory to Practice. *Machine Learning Research*, **18**, 7854-7907.
- [12] Lin, H., Mairal, J. and Harchaoui, Z. (2015) A Universal Catalyst for First-Order Optimization. *Advances in Neural Information Processing Systems*, **28**, 3384-3392.
- [13] Xie, Z. (2017) A Primal-Dual Method with Linear Mapping for a Saddle Point Problem in Image Deblurring. *Journal of Visual Communication and Image Representation*, **42**, 112-120.
<https://doi.org/10.1016/j.jvcir.2016.11.011>
- [14] Schmidt, M., Roux, N.L. and Bach, F.R. (2011) Convergence Rates of Inexact Proximal-Gradient Methods for Convex Optimization. Curran Associates Inc., Red Hook, 1458-1466.
- [15] Beck, A. (2017) First-Order Methods in Optimization. SIAM, Philadelphia, 269-304.
- [16] Aujol, J.F. and Dossal, C. (2015) Stability of Over-Relaxations for the Forward-Backward Algorithm Application to FISTA. *SIAM Journal on Optimization*, **25**, 2408-2433.
<https://doi.org/10.1137/140994964>
- [17] Beck, A. and Teboulle, M. (2009) A Fast Iterative Shrinkage-Thresholding Algorithm for Linear Inverse Problems. *SIAM Journal on Imaging Sciences*, **2**, 183-202.
<https://doi.org/10.1137/080716542>