arXiv:2301.02511v3 [math.OC] 4 Dec 2023

带自适应步长的随机原始-对偶混合梯度算法

Antonin Chambolle* Claire Delplancke† Matthias J. Ehrhardt‡
Carola-Bibiane Sch¨onlieb§ Junqi Tang¶

2023年12月5日

摘要

本文提出了一种具有自适应步长的新型原始-对偶算法。由于具备良好的可扩展性，采用固定步长的随机原始-对偶混合梯度（SPDHG）算法已广泛应用于许多科学领域的大规模凸优化问题。为确保收敛，原始步长与对偶步长的乘积必须满足上界约束，而步长比值的选择在实际应用中至关重要。迄今为止，尚无系统且有效的方法来选择 SPDHG 的原始步长和对偶步长。本文提出了一类通用的自适应SPDHG（A-SPDHG）算法，并在较弱的假设条件下证明了其收敛性。我们还提出了满足理论假设的具体参数更新策略，从而得到收敛算法。计算机断层扫描数值示例展示了所提方案的有效性。

1 引言

[8] 中提出的随机原始-对偶混合梯度（SPDHG）算法是原始-对偶混合梯度（PDHG）算法的随机版本，后者也称为 Chambolle–Pock 算法 [9]。对于大规模非光滑凸逆问题框架中的各类问题，SPDHG 已被证明比 PDHG 更高效 [13, 22, 24, 27]。事实上，SPDHG 在每次迭代时仅使用部分数据，从而降低了计算前向算子及其伴随算子的成本；因此，在相同计算开销下，SPDHG 的收敛速度快于 PDHG。这在医学成像领域尤为重要，因为该领域需要收敛速度符合临床标准的算法，同时还需要

*CEREMADE，巴黎多菲纳大学，马歇尔·德·拉特尔·德·塔西尼广场，75775 巴黎，以及 MOKA-PLAN，法国国家信息与自动化研究所巴黎分部，法国。

†法国电力集团巴黎-萨克雷实验室，萨克雷路，91300 帕莱索，法国。本文所述研究开展期间，CD 任职于巴斯大学数学科学系。

‡巴斯大学数学科学系，克莱弗顿唐，巴斯 BA2 7AY，英国。§剑桥大学应用数学与理论物理系，威尔伯福斯路，剑桥 CB3 0WA，英国。

¶伯明翰大学数学学院，埃德巴斯顿，伯明翰 B15 2TT，英国。

能够处理凸的非光滑先验，例如全变差（TV）；这类先验适用于病态成像逆问题，但会使得无法采用可扩展的基于梯度的方法。

与 PDHG 类似，在原始步长和对偶步长的乘积有界于一个取决于待求解问题的常数这一假设下，可以证明 SPDHG 收敛。另一方面，原始步长与对偶步长的比值是一个自由参数，其取值需要由用户选择。该参数可解释为控制原始与对偶收敛之间平衡的参数，其取值会对 PDHG 的收敛速度产生重大影响，SPDHG 也同样如此 [12]。这在实践中带来了一个重要挑战，因为目前没有已知的理论或经验规则可以指导该参数的选择。手动调参的计算成本很高，因为这需要在一系列取值下运行并比较算法，而且无法保证在一个数据集上能快速收敛的取值，在另一个数据集上仍是合适的选择。对于 PDHG， [14] 提出了一种在线原始-对偶平衡策略来解决这一问题，其中步长的取值会随迭代而变化。更一般而言，自适应步长已用于带回溯的 PDHG [14, 20], 、通过适应局部光滑性的 PDHG [25] ，以及其他多种算法，包括梯度方法 [19], 、次梯度方法 [3] 和分裂方法 [4, 6, 5, 7, 18] ，以提高收敛速度，并避免显式使用模型常数（如 Lipschitz 常数或算子范数）的需要。对于 SPDHG，已有研究在磁粒子成像中使用了一种经验性自适应方案，但没有给出收敛性证明 [27]。

在理论方面，证明基于近端算子的凸优化算法收敛的一种标准方法，是利用 F´ejer 单调性的概念。采用恒定步长会得到固定度量设置，而采用自适应步长则会得到可变度量设置。文献 [11] 讨论了可变度量设置下确定性 F´ejer 单调序列的收敛性，而文献 [10] 关注固定度量设置下随机 F´ejer 单调序列的收敛性。

在本文中，我们提出并研究 SPDHG 的一个自适应版本。具体而言：

- 我们提出了一大类策略，用于自适应地选择 SPDHG 的步长。该类策略包括但不限于自适应原始-对偶平衡策略；在该策略中，步长之比控制着原始变量与对偶变量的收敛平衡，并会在线调整。

- 我们证明了在该类策略下 SPDHG 几乎必然收敛。为此，我们引入了 C-稳定性的概念，将其推广了 F´ejer 单调性的概念，并证明了可变度量设置下随机 C-稳定序列的收敛性，从而推广了文献 [11] 和 [10]中的结果。随后，我们采用与文献 [16, 1]中几乎必然收敛证明类似的策略，表明所提出的算法属于这一全新的理论框架。

- 我们在大规模成像逆问题（稀疏视角计算机断层成像、有限角度 CT、低剂量 CT）上，比较了 SPDHG 在各种自适应方案和已知固定步长方案下的性能。我们发现，原始-对偶平衡自适应策略始终不逊于其他所有策略，甚至更快。尤其是，当固定步长虽处于理论收敛范围内、但选择不当时，该策略相较于固定步长策略能够持续显著提升收敛速度。这一点尤其重要，因为若不进行成本高昂的比较测试，就无法判断固定步长选择得是否恰当。

即使在 SPDHG 的固定步长经过良好调优的情况下，也就是说，这些步长处于观察到自适应步长收敛至的范围内，我们仍发现，经过一定次数的迭代后，我们的自适应方案依然能够加速标准 SPDHG 的收敛。最后，我们特别关注自适应方案所用的超参数。这些超参数实质上控制着算法的自适应程度，并且各自都有明确的含义，在实践中也易于选择。通过大量数值测试，我们发现，在所提供的经验范围内，自适应方案的收敛速度对这些参数的选择具有稳健性，因此无需微调即可直接应用于当前问题，并解决用户面临的步长选择难题。

本文其余部分安排如下。第 2 节介绍采用自适应步长的 SPDHG，陈述收敛定理并给出证明。第3节提出实现自适应性的具体方案，随后在第4节对 CT 数据进行数值测试。第5节总结全文。最后，第 6 节汇集了一些有用的引理及其证明。

2 理论

2.1 收敛定理

待求解的变分问题具有以下形式 m:

$$\min_{x\in X}\sum_{i=1}^{n} f_i(A_i x)+g(x),$$

其中， X 和$(Y_i)_{i\in\{1,\dots,n\}}$ 是希尔伯特空间， $A_i: X\to Y_i$ 是有界线性算子， $f_i: Y_i\to\mathbb{R}\cup\{+\infty\}$和 $g\colon X\to\mathbb{R}\cup\{+\infty\}$是凸函数。我们定义 $Y=Y_1\times\cdots\times Y_n$，其元素为$y=(y_1,\dots,y_n)$，并定义 $A\colon X\to Y$，使得 $Ax=(A_1x,\dots,A_nx)$。相应的鞍点问题写作

$$\min_{x\in X}\sup_{y\in Y}\sum_{i=1}^{n}\langle A_i x,y_i\rangle-f_i^*(y_i)+g(x), \tag{2.1}$$

其中， f_i^*表示 f_i的Fenchel共轭。(2.1)的解集记为 $\mathcal{C}$，非负整数集记为 $\mathbb{N}$和 $[\![1,\ n]\!]$表示 $\{1,\dots,n\}$。 $\mathcal{C}$中的元素(x^*,y^*)称为鞍点，其特征为

$$A_i x^*\in\partial f_i^*(y_i^*),\ i\in[\![1,n]\!];\quad -\sum_{i=1}^{n}A_i^*y^*\in\partial g(x^*). \tag{2.2}$$

为求解鞍点问题，我们在算法 2.1 中引入自适应随机原始-对偶混合梯度（A-SPDHG）算法。在每次迭代 $k\in\mathbb{N}$中，A-SPDHG包含以下五个步骤:

- 更新原始步长 τ^k 和对偶步长$(\sigma_i^k)_{i\in[\![1,\ n]\!]}$ （第 4 行）；
- 通过步长为 τ^{k+1} 的近端步更新原始变量 x^k （第 5 行）；

- 以概率 p_i 随机选择一个索引 i （第 6 行）；
- 通过步长为 σ_i^{k+1} 的近端步更新对偶变量 y_i^k（第 7 行）；
- 计算外推对偶变量（第 8 行）。

A-SPDHG 具有自适应性，具体而言，步长值在每次迭代时都会根据更新规则进行更新，该规则会考虑截至当前迭代的原始变量和对偶迭代点 x^l 与 y^l 的值。由于迭代点具有随机性，步长本身也具有随机性，因此在理论分析中必须谨慎处理这一点。

算法 2.1：A-SPDHG（可变步长，串行抽样）

1: 输入：对偶步长 $(\sigma_i^0)_{i\in[\![1,\ n]\!]}$，原始步长 τ^0，更新规则；概率 $(p_i)_{i\in[\![1,\ n]\!]}$；
   原始变量 x^0，对偶变量 y^0
2: 初始化 $\bar{y}^0 = y^0$
3: 对于 $k \in [\![0, K-1]\!]$，执行
4: 根据更新规则以及 $(\sigma_i^l)_{i\in[\![1,n]\!]}$, τ^l 的取值，确定 $(\sigma_i^{k+1})_{i\in[\![1,n]\!]}$, τ^{k+1}，
   x^l 和 y^l，对于 $l \in [\![0,k]\!]$。
5: $x^{k+1} =$近端算子$_{\tau^{k+1}g}(x^k - \tau^{k+1}A^*\bar{y}^k)$
6: 以概率 p_i 随机选取 $i \in [\![1,n]\!]$
7: $y_j^{k+1} = \begin{cases} \text{近端算子}_{\sigma_i^{k+1}f_i^*}(y_i^k + \sigma_i^{k+1}A_i x^{k+1}) & \text{若 } j = i \\ y_j^k & \text{若 } j \neq i \end{cases}$
8: $\bar{y}_j^{k+1} = \begin{cases} y_i^{k+1} + \frac{1}{p_i}\left(y_i^{k+1} - y_i^k\right) & \text{若 } j = i \\ y_j^k & \text{若 } j \neq i \end{cases}$
9: 结束循环
10: 返回 x^K

在讨论 A-SPDHG 的收敛性之前，我们先回顾一下先进的 SPDHG 的一些事实。SPDHG 的每次迭代都涉及从 $[\![1,n]\!]$ 中随机抽取一个子集。在随机子集为单元素集合的串行抽样情况下，SPDHG 算法 [8] 是算法 2.1 的一个特例，其更新规则为

$$\begin{cases} \sigma_i^{k+1} = \sigma_i^k\ (=\sigma_i),\ i \in [\![1,n]\!], \\ \tau^{k+1} = \tau^k\ (=\tau_i), \end{cases} \qquad k \in \mathbb{N}.$$

在以下条件下

$$\tau\sigma_i < \frac{p_i}{\|A_i\|^2}, \quad i \in [\![1,n]\!], \tag{2.3}$$

SPDHG 迭代点几乎必然收敛到鞍点问题 (2.1) 的一个解 [1, 16].

现在我们来讨论 A-SPDHG 的收敛性。下文的主要定理——定理 2.1——给出了 A-SPDHG 可证明收敛所需的更新规则条件。简单来说，这些条件有三方面：

(i) 步骤 $k+1$ 的步长 $(\sigma_i^{k+1})_{i\in[\![1,\ n]\!]}$ 以及 τ^{k+1}，仅取决于截至步骤 k，

(ii) 步长满足条件 (2.3) 的一个一致版本，

(iii) 对于 $i \in [\![1、\ n]\!]$，步长序列 $(\tau^k)_{k\geq 0}$ 和 $(\sigma_i^k)_{k\geq 0}$ 不会下降得太快。更准确地说，按照下文定义的意义，它们一致地几乎必然准递增。

为了严谨地陈述该定理，我们先引入一些有用的记号和定义。对于所有 $k \in \mathbb{N}$，由截至迭代点 k 的迭代点生成的 σ-代数 $(x^l, y^l)\ l \in [\![0,k]\!])$ 记为 $\mathcal{F}^k$。如果对于所有 $k \in \mathbb{N}$，u^k 关于 $\mathcal{F}^k$ 可测，则称序列 $(u^k)_{k\in\mathbb{N}}$ 是 $\left(\mathcal{F}^k\right)_{k\in\mathbb{N}}$-适应的。

如果存在一个取值于 $[0,1)$ 的序列 $(\eta^k)_{k\in\mathbb{N}}$，称为 $(u^k)_{k\in\mathbb{N}}$ 上的控制序列，使得 $\sum_{k=1}^{\infty}\eta^k<\infty$ 且：则称正实数序列 $(u^k)_{k\in\mathbb{N}}$ 准递增。

$$u^{k+1}\geq(1-\eta^k)u^k,\quad k\in\mathbb{N}. \tag{2.4}$$

推广来说，若存在一个取值于 $[0,1)$ 的确定性序列 $(\eta^k)_{k\in\mathbb{N}}$，使得 $\sum_{k=1}^{\infty}\eta^k<\infty$ 且上文中的公式 (2.4) 几乎必然（a.s.）成立，则称随机正实数序列 $(u^k)_{k\in\mathbb{N}}$ 一致几乎必然准递增。

定理 2.1（A-SPDHG 的收敛性）。设 X 和 Y 为可分希尔伯特空间，$A_i\colon X\to Y_i$ 为有界线性算子，$f_i\colon Y_i\to\mathbb{R}\cup\{+\infty\}$ 和 $g\colon X\to\mathbb{R}\cup\{+\infty\}$ 为适当的凸下半连续函数，对所有 $i\in[\![1、n]\!]$ 均成立。假设鞍点集合 $\mathcal{C}$ 非空，且抽样是适当的，也就是说，对所有 $i\in[\![1、\ n]\!]$，均有 $p_i>0$。若满足以下条件：

(i) 步长序列 $(\tau^{k+1})_{k\in\mathbb{N}}$、$(\sigma_i^{k+1})_{k\in\mathbb{N}}$、$i\in[\![1、\ n]\!]$ 是 $\left(\mathcal{F}^k\right)_{k\in\mathbb{N}}$-适应的，

(ii) 存在 $\beta\in(0,1)$，使得对于所有指标 $i\in[\![1、\ n]\!]$ 和迭代点 $k\in\mathbb{N}$，

$$\tau^k\sigma_i^k\frac{\|A_i\|^2}{p_i}\leq\beta<1, \tag{2.5}$$

(iii) 对于所有指标 $i\in[\![1、\ n]\!]$，初始步长 τ^0 和 σ_i^0 均为正，且对于所有指标 $i\in[\![1、\ n]\!]$，步长序列 $(\tau^k)_{k\in\mathbb{N}}$ 和 $(\sigma_i^k)_{k\in\mathbb{N}}$ 均一致地几乎必然准递增，

则迭代点序列 $(x^k,y^k)_{k\in\mathbb{N}}$ 几乎必然收敛到 $\mathcal{C}$ 中的一个元素。

虽然条件 (i)-(iii) 足够一般，能够涵盖多种步长更新规则，但实际中我们将重点关注原始-对偶平衡策略，即在每次迭代时，分别将原始步长和对偶步长乘以一个因子的倒数。在这种情况下，更新规则取决于随机正序列 $(\gamma^k)_{k\in\mathbb{N}}$，其形式如下：

$$\tau^{k+1}=\frac{\tau^k}{\gamma^k},\quad \sigma_i^{k+1}=\gamma^k\sigma_i^k,\quad i\in[\![1,n]\!]. \tag{2.6}$$

引理 2.2（原始-对偶平衡）。设步长序列满足方程 (2.6)，并进一步假设 $(\gamma^k)_{k\in\mathbb{N}}$ 是 $\left(\mathcal{F}^k\right)_{k\in\mathbb{N}}$-适应的，且初始步长满足

$$\tau^0\sigma_i^0\frac{\|A_i\|^2}{p_i}<1,\quad i\in[\![1,n]\!],$$

并且为正；存在一个取值于 $[0,1)$ 的确定性序列 $(\epsilon^k)_{k\in\mathbb{N}}$，使得$\sum \epsilon^k < \infty$，且对所有 $k \in \mathbb{N}$ 和 $i \in [\![1, n]\!]$,

$$\min\left\{\gamma^k, (\gamma^k)^{-1}\right\} \geq 1 - \epsilon^k. \tag{2.7}$$

因此，步长序列满足定理 2.1 的假设 (i)-(iii)。

引理 2.2 在第 6 节中证明。

与相关文献的联系：

- 原始-对偶平衡策略已在 [14] 中为 PDHG 引入；事实上，对于 $n = 1$，我们利用引理 2.2 得到了 [14] 中提出的非回溯算法。因此，我们的定理还蕴含该算法的逐点收敛性，而 [14] 已证明该算法的收敛性，其意义是残差趋于零。

- 同样对于 PDHG，[20] 提出了一种更新规则（未给出证明），其中步长之比是准单调不增或准单调不减的。这一要求与我们的要求相似，但并无直接联系；我们的要求是步长本身准单调不增。

- 对于 SPDHG，[27] 中提出的角度约束步长规则（未给出收敛性证明）满足假设 (i)-(iii)。

证明概要：定理 2.1 将在以下各小节中证明。我们首先在第 2.2 节中定义与原始-对偶乘积空间上算法步长相关的度量。由于步长是自适应的，因此我们得到一个度量序列。定理 2.1 的证明策略与 [1] 和 [16] 的证明相似，但需要引入新的要素来处理度量的变化。在定理 2.5 中，我们给出配备随机变量度量的希尔伯特空间中抽象随机序列的收敛条件。在第 2.4 节和第 2.5 节中，我们将证明 A-SPDHG 属于定理 2.5 的适用范围。我们将在第 2.6 节汇总所有要素并完成证明。

2.2 变度量

对于希尔伯特空间 H，我们将 $\mathbb{S}(H)$ 称为从 H 到 H的有界自伴线性算子集合，并对所有 $M \in \mathbb{S}(H)$ 引入如下记号：

$$\|u\|_M^2 = \langle Mu, u\rangle, \quad u \in H.$$

略微滥用记号，我们用 $\|\cdot\|_\alpha^2 = \|\cdot\|_{\alpha\mathrm{Id}}^2$ 表示标量 $\alpha \in \mathbb{R}$。注意，如果 M 是正定的，那么 $\|\cdot\|_M$ 是 H 上的一个范数。此外，我们在 $\mathbb{S}(H)$ 上引入偏序 4，使得对于 M、$N \in \mathbb{S}(H)$，

$$N \preccurlyeq M \quad \text{if} \quad \forall u \in H, \|u\|_N \leq \|u\|_M.$$

我们将 $\mathbb{S}_\alpha(H)$ 称为 $\mathbb{S}(H)$ 的子集，该子集由满足 $\alpha\mathrm{Id} \preccurlyeq M$ 的 M 构成。此外，若存在一个确定性的非负序列 $(\eta^k)_{k\in\mathbb{N}}$，使得 $\sum_{k=1}^{\infty} \eta^k < \infty$ 且几乎必然成立，则称随机序列 $(M^k)_{k\in\mathbb{N}}$ 在 $\mathbb{S}(H)$ 中一致几乎必然准递减。

$$M^{k+1} \preccurlyeq (1 + \eta^k)M^k, \quad k \in \mathbb{N}.$$

回到 A-SPDHG，对于每次迭代 $k \in \mathbb{N}$ 和每个索引 $i \in [\![1$，我们定义 $\mathbb{S}(X \times Y_i)$ 的两个块算子 $n]\!]$:

$$M_i^k = \begin{pmatrix} \frac{1}{\tau^k}\mathrm{Id} & -\frac{1}{p_i}A_i^* \\ -\frac{1}{p_i}A_i & \frac{1}{p_i\sigma_i^k}\mathrm{Id} \end{pmatrix}, \quad N_i^k = \begin{pmatrix} \frac{1}{\tau^k}\mathrm{Id} & 0 \\ 0 & \frac{1}{p_i\sigma_i^k}\mathrm{Id} \end{pmatrix},$$

以及一个 $\mathbb{S}(X \times Y)$ 的块算子，如下:

$$N^k = \begin{pmatrix} \frac{1}{\tau^k}\mathrm{Id} & & & & (0) \\ & \frac{1}{p_1\sigma_1^k}\mathrm{Id} & & & \\ & & \ddots & & \\ & & & \frac{1}{p_i\sigma_i^k}\mathrm{Id} & \\ & & & & \ddots \\ (0) & & & & \frac{1}{p_n\sigma_n^k}\mathrm{Id} \end{pmatrix}. \tag{2.8}$$

下面的引理将定理 2.1 中关于变度量序列性质的假设 (i)-(iii) 转化为相应结论。

引理 2.3（变度量性质）。(a) 定理 2.1 的假设 (i) 蕴含 $(M_i^{k+1})_{k\in\mathbb{N}}$、$(N_i^{k+1})_{k\in\mathbb{N}}$、$i \in [\![1$、$n]\!]$ 和 $(N^{k+1})_{k\in\mathbb{N}}$ 是 $\left(\mathcal{F}^k\right)_{k\in\mathbb{N}}$-适应的。

(b) 定理 2.1 的假设 (ii) 等价于存在 $\beta \in (0,1)$，使得对所有指标 $i \in [\![1$、$n]\!]$ 以及迭代点 $k \in \mathbb{N}$,

$$(1-\sqrt{\beta})N_i^k \preccurlyeq M_i^k.$$

(c) 定理 2.1 的假设 (ii) 和 (iii) 蕴含 $(M_i^k)_{k\in\mathbb{N}}$、$(N_i^k)_{k\in\mathbb{N}}$、$i \in [\![1$、$n]\!]$ 和 $(N^k)_{k\in\mathbb{N}}$ 一致几乎必然拟递减。

(d) 定理 2.1 的假设 (ii) 和 (iii) 表明，对于所有 $i \in [\![1,n]\!]$，序列 $(\tau^k)_{k\in\mathbb{N}}$ 和 $(\sigma_i^k)_{k\in\mathbb{N}}$ 几乎必然有上界，且有正常数作为下界。特别地，这意味着存在 $\alpha > 0$，使得对于所有 $i \in [\![1,n]\!]$ 和 $k \in \mathbb{N}$，$N_i^k \in \mathbb{S}_\alpha(X \times Y_i)$；或者等价地，对于所有 $k \in \mathbb{N}$，$N^k \in \mathbb{S}_\alpha(X \times Y)$。

注 2.4（步长诱导的原始-对偶乘积空间度量）。该引理表明，M_i^k, N_i^k 和 N^k 是正定的，因此在相应空间上诱导出度量。若 $n = 1$ 且步长为常数，则 M_i^k 对应于 *[17]*, 中使用的度量；在该处，PDHG 被重新表述为原始-对偶乘积空间上一种非平凡度量下的近端点算法。

引理 2.3 的证明。该引理的断言 (a) 源于这样一个事实：对于所有迭代 $k \in \mathbb{N}$，算子 M_i^{k+1}、N_i^{k+1} 和 N^{k+1} 都属于由 $\left\{\tau^{k+1}, \sigma_i^{k+1}, i \in [\![1,n]\!]\right\}$ 生成的 σ-代数。断言 (b) 可由补充材料中的引理 6.1 的公式 (6.2) 得出。断言 (c) 的证明要复杂一些。假设定理 2.1 的假设 (iii) 成立

并设 $(\eta_0^k)_{k\in\mathbb{N}}$ 和 $(\eta_i^k)_{k\in\mathbb{N}}$ 分别是 $i\in[\![1,\ n]\!]$ 对应的 $(\tau^k)_{k\in\mathbb{N}}$ 和 $(\sigma_i^k)_{k\in\mathbb{N}}$ 的控制序列。我们定义序列 $(\eta^k)_{k\in\mathbb{N}}$ 如下：

$$\eta^k=\max\left\{\eta_i^k,\, i\in[\![0,n]\!]\right\},\quad k\in\mathbb{N}, \tag{2.9}$$

它是 $(\tau^k)_{k\in\mathbb{N}}$ 和 $(\sigma_i^k)_{k\in\mathbb{N}}$ 在 $i\in[\![1,n]\!]$ 下的一个共同控制序列，即有限个控制序列中的最大值。令 $k\in\mathbb{N}$ 和 $i\in[\![1,n]\!]$ 固定。由于有限个概率为一的可测事件的交仍然是概率为一的可测事件，因此几乎必然对所有 $(x,y_i)\in X\times Y_i$,

$$\begin{aligned}\|(x,y_i)\|_{N_i^{k+1}}^2&=\frac{1}{\tau^{k+1}}\|x\|^2+\frac{1}{p_i\sigma_i^{k+1}}\|y_i\|^2\\&\leq\frac{1}{1-\eta^k}\left(\frac{1}{\tau^k}\|x\|^2+\frac{1}{p_i\sigma_i^k}\|y_i\|^2\right)\\&=\left(1+\frac{\eta_k}{1-\eta^k}\right)\|(x,y_i)\|_{N_i^k}^2.\end{aligned}$$

因此，序列 $(N_i^k)_{k\in\mathbb{N}}$ 以控制序列 $\left(\eta^k(1-\eta^k)^{-1}\right)_{k\in\mathbb{N}}$ 一致拟递减，而该控制序列确实是一个有界和的正序列。（为说明 $\left(\eta^k(1-\eta^k)^{-1}\right)_{k\in\mathbb{N}}$ 有界和，注意到 $\left(\eta^k\right)_{k\in\mathbb{N}}$ 可求和，因此收敛于 0，故当整数 k 大于某个 K时，它小于 $1/2$；进而，当整数 k 大于 K时，项 $\eta^k(1-\eta^k)^{-1}$下界为 0，上界为 $2\eta^k$，因此可求和。）类似地可以证明，$(N^k)_{k\in\mathbb{N}}$ 以相同的控制序列一致拟递减。接下来考虑 $(M_i^k)_{k\in\mathbb{N}}$的情形，与前面一样，我们有：

$$M_i^{k+1}=\begin{pmatrix}\frac{1}{\tau^{k+1}}\mathrm{Id} & -\frac{1}{p_i}A_i^*\\ -\frac{1}{p_i}A_i & \frac{1}{p_i\sigma_i^{k+1}}\mathrm{Id}\end{pmatrix}\preccurlyeq M_i^k+\frac{\eta^k}{1-\eta^k}N_i^k\preccurlyeq\left(1+\frac{\eta^k}{1-\eta^k}\frac{1}{1-\sqrt{\beta}}\right)M_i^k$$

根据 (b)。

最后证明断言 (d)。根据假设 (iii)，序列 $(\tau^k)_{k\in\mathbb{N}}$ 和 $(\sigma_i^k)_{k\in\mathbb{N}}$ 几乎必然一致地准递增。我们按照 (2.9) 定义一个公共控制序列 $(\eta^k)_{k\in\mathbb{N}}$。于是，序列 $(\tau^k)_{k\in\mathbb{N}}$ 和 $(\sigma_i^k)_{k\in\mathbb{N}}$ 几乎必然有相同的确定性常数$C=\min\left\{\tau^0,\,\sigma_i^0,\,i\in[\![1,n]\!]\right\}\prod_{j=0}^{\infty}(1-\eta^j)$作为下界；该常数为正，因为初始步长为正，且 $(\eta^k)_{k\in\mathbb{N}}$ 取值于 $[0,1)$ 并且其和有限。此外，根据假设 (ii)，序列 $(\tau^k)_{k\in\mathbb{N}}$ 和 $(\sigma_i^k)_{k\in\mathbb{N}}$ 的乘积几乎必然有上界。因此，序列 $(\tau^k)_{k\in\mathbb{N}}$和 $(\sigma_i^k)_{k\in\mathbb{N}}$各自几乎必然有上界。对所有 $i\in[\![1,$ 与 $N_i^k\in\mathbb{S}_\alpha(X\times Y_i)$ 的等价性，以及与 $N^k\in\mathbb{S}_\alpha(X\times Y)$ 的等价性，都是显而易见的。

□

2.3 随机 C-稳定序列在随机变量度量中的收敛

设 H是一个希尔伯特空间，$C\subset H$是 H的子集。设 $(\Omega\ \sigma(\Omega)\ \mathbb{P})$ 是一个概率空间。除非另有说明，以下所有随机变量都假定定义在 Ω 上，且关于 $\sigma(\Omega)$ 可测。设 $(Q^k)_{k\in\mathbb{N}}$ 是一个取值于 $\mathbb{S}(H)$ 的随机序列。

若对所有 $u \in C$，序列 $\left(\|u^k - u\|_{Q^k}\right)_{k\in\mathbb{N}}$ 几乎必然收敛，则称取值于 H 的随机序列 $(u^k)_{k\in\mathbb{N}}$ 关于目标 C相对于 $(Q^k)_{k\in\mathbb{N}}$ 稳定。以下定理给出了这类序列收敛的充分条件。

定理 2.5（C-稳定序列的收敛）。设 H 为可分希尔伯特空间，C 为 H的非空闭子集，$(Q^k)_{k\in\mathbb{N}}$ 为取值于 $\mathbb{S}(H)$的随机序列，且 $(u^k)_{k\in\mathbb{N}}$ 为取值于H的随机序列。若满足以下条件：

(i) 对给定的 $\alpha > 0$，$(Q^k)_{k\in\mathbb{N}}$ 取值于 $\mathbb{S}_\alpha(H)$，且一致几乎必然拟递减 ，

(ii) 相对于 $(Q^k)_{k\in\mathbb{N}}$，$(u^k)_{k\in\mathbb{N}}$ 关于目标 C 是稳定的，

(iii) $(u^k)_{k\in\mathbb{N}}$ 的每个弱序列聚点几乎必然属于 C，也就是说，存在 Ω 的一个可测子集 $\Omega^{(iii)}$，其概率为一，使得对于所有 $\omega \in \Omega$，$(u^k(\omega))_{k\in\mathbb{N}}$ 的每个弱序列聚点都属于 C。

则 $(u^k)_{k\in\mathbb{N}}$ 几乎必然弱收敛到 C中的一个随机变量。

关于目标集 C 的稳定性可由相对于 C 的 ´ejer 单调性和准 ´ejer 单调性推出；这两种性质此前分别针对随机序列 [10] 或在变度量 [11], 的框架下进行了研究，但据我们所知，尚未同时研究过。定理 2.5 的证明与 [10, 命题 2.3 (iii)] 的证明思路相同，并用到了 [11] 中的两个结果。

证明。集合 C 是可分希尔伯特空间 H 的子集，因此是可分的。由于 C 是闭且可分的，存在 $\{c^n, n \in \mathbb{N}\}$ C 的一个可数子集，其闭包等于 C。根据假设 (ii)，对所有 $n \in \mathbb{N}$，存在 Ω 的一个概率为一的可测子集 $\Omega^n_{(ii)}$，使得序列 $(\|u^k(\omega) - c^n\|_{Q^k(\omega)})_{k\in\mathbb{N}}$ 对所有 $\omega \in \Omega^n_{(ii)}$ 收敛。此外，设 $\Omega^{(i)}$ 为 Ω 的一个概率为一的可测子集，对应于假设 (i) 的几乎必然性质。令

$$\tilde{\Omega} = \left(\bigcap_{n\geq 0} \Omega^n_{(ii)}\right) \bigcap \Omega_{(i)} \bigcap \Omega_{(iii)}.$$

˜由于 Ω 是可数个概率为一的可测子集的交集，因此 Ω 本身也是 Ω 的˜ ˜可测集合，且 $\mathbb{P}(\Omega) = 1$。在余下的证明中，固定 $\omega \in \Omega$。

序列 $(Q^k(\omega))_{k\in\mathbb{N}}$ 在 $\mathbb{S}_\alpha(H)$ 中取值，对 $\alpha > 0$ 成立，且是准递减的，其控制序列为 $(\eta^k(\omega))_{k\in\mathbb{N}}$。此外，对所有 $k \in \mathbb{N}$,

$$\|Q^k(\omega)\| \leq \left(\prod_{j=0}^{k-1} \left(1 + \eta^j\right)\right) \|Q^0(\omega)\| \leq \left(\prod_{j=0}^{\infty} \left(1 + \eta^j\right)\right) \|Q^0(\omega)\|,$$

其中乘积 $\prod_{j=0}^{\infty} \left(1 + \eta^j\right)$ 是有限的，因为 $(\eta^k)_{k\in\mathbb{N}}$ 为正且可求和。根据 [11, 引理 2.3], $(Q^k(\omega))_{k\in\mathbb{N}}$ 逐点强收敛到某个 $Q(\omega) \in \mathbb{S}_\alpha(H)$。

此外，对于所有 $x \in C$，存在一个序列 $(x^n)_{n\in\mathbb{N}}$，其值属于 $\{c^n, n \in \mathbb{N}\}$，并强收敛于 x。根据假设，对于所有 $n \in \mathbb{N}$，序列 $(\|u^k(\omega) - x^n\|_{Q^k(\omega)})_{k\in\mathbb{N}}$ 收敛于一个

极限，该极限记作 $l^n(\omega)$。对于所有 $n \in \mathbb{N}$ 和 $k \in \mathbb{N}$，根据三角不等式，我们可以写出：

$$-\|x^n - x\|_{Q^k(\omega)} \leq \|u^k(\omega) - x\|_{Q^k(\omega)} - \|u^k(\omega) - x^n\|_{Q^k(\omega)} \leq \|x^n - x\|_{Q^k(\omega)}.$$

令 $k \to +\infty$ 取极限，可得：

$$\begin{aligned} -\|x^n - x\|_{Q(\omega)} &\leq \liminf_{k\to\infty} \|u^k(\omega) - x\|_{Q^k(\omega)} - l^n(\omega) \\ &\leq \limsup_{k\to\infty} \|u^k(\omega) - x\|_{Q^k(\omega)} - l^n(\omega) \leq \|x^n - x\|_{Q(\omega)}. \end{aligned}$$

现在令 $n \to +\infty$ 取极限，可知序列 $(\|u^k(\omega) - x\|_{Q^k(\omega)})_{k\in\mathbb{N}}$ 对所有 $x \in C$ 均收敛。另一方面，由于 $\omega \in \Omega^{(iii)}$，$(u^k(\omega))_{k\in\mathbb{N}}$ 的弱聚点属于 C。因此，根据 [11, 定理 3.3], ，序列 $(u^k(\omega))_{k\in\mathbb{N}}$ 几乎必然收敛到一点 $u(\omega) \in C$。 □

现在我们已具备证明定理 2.1 的条件。我们将在第 2.4 节和第 2.5 节分别证明 A-SPDHG 满足定理 2.5 的第（ii）点和第（iii）点，并在第 2.6 节完成证明。有趣的是，第（ii）点和第（iii）点的证明依赖于理解 A-SPDHG 的两种不同方式。第（ii）点依赖于凸优化论证：利用原始变量在第 $k+1$ 步相对于 $\mathcal{F}^k$ 的可测性，可以写出一个收缩型不等式，将第 $k+1$ 步迭代序列的范数的条件期望与第 k 步迭代序列的范数联系起来。第（iii）点依赖于单调算子理论：我们利用这样一个事实，即从半移位迭代 (y^k, x^{k+1}) 更新到 (y^{k+1}, x^{k+2})，可以解释为在 i 条件下，在 k 步随机选取索引 $X \times Y_i$ 后，对 $X \times Y_i$ 执行的一步近端点算法。

2.4 A-SPDHG 关于鞍点集合是稳定的

本节中，我们证明 $(x^k, y^k)_{k\in\mathbb{N}}$ 相对于上文公式 (2.8) 定义的变度量序列 $(N^k)_{k\in\mathbb{N}}$ 关于 $\mathcal{C}$ 是稳定的。我们引入算子 $P \in \mathbb{S}(Y)$ 和 $\Sigma^k \in \mathbb{S}(Y)$，它们分别定义为

$$(Py)_i = p_i y_i, \quad (\Sigma^k y)_i = \sigma_i^k y_i, \quad i \in [\![1, n]\!],$$

以及泛函 $(U^k)_{k\in\mathbb{N}}$、$(V^k)_{k\in\mathbb{N}}$，对所有 $(x, y) \in X \times Y$ 定义如下：

$$\begin{aligned} U^k(y) &= \|y\|^2_{(P\Sigma^k)^{-1}}, \\ V^k(x, y) &= \|x\|^2_{(\tau^k)^{-1}} - 2\langle P^{-1}Ax, y\rangle + \|y\|^2_{(P\Sigma^k)^{-1}}. \end{aligned}$$

我们首先回顾 SPDHG 迭代点所满足的基石不等式，该不等式最初发表于 [8] ，并在 [1] 中重新表述。

引理 2.6（[1], 引理 4.1）。对于每个鞍点（x^*， y^*），几乎必然对所有 $k \in \mathbb{N}\setminus\{0\}$ 均有，

$$\begin{aligned} &\mathbb{E}\left[V^{k+1}(x^{k+1} - x^*, y^{k+1} - y^k) + U^{k+1}(y^{k+1} - y^*)|\mathcal{F}^k\right] \\ \leq\ & V^{k+1}(x^k - x^*, y^k - y^{k-1}) + U^{k+1}(y^k - y^*) \\ & - V^{k+1}(x^{k+1} - x^k, y^k - y^{k-1}). \end{aligned} \tag{2.10}$$

第二步是将定理 2.1 的假设与式 (2.10) 中出现的泛函的性质联系起来。引入由至多一个非零分量的元素（y_1，.. .， y_n）构成的集合 $Y_{\text{sparse}} \subset Y$。

引理 2.7（感兴趣的泛函的性质）。在定理 2.1 的假设下，存在一个非负可求和序列 $(\eta^k)_{k\in\mathbb{N}}$，使得几乎必然地，对每个迭代 $k \in \mathbb{N}$ 以及 $x \in X$、$y \in Y$、$z \in Y_{sparse}$，均有：

$$U^{k+1}(y) \leq (1+\eta^k)U^k(y), \tag{2.11a}$$
$$V^{k+1}(x,z) \leq (1+\eta^k)V^k(x,z), \tag{2.11b}$$
$$\|(x,z)\|_{N^k}^2 \geq \alpha\|(x,z)\|^2, \tag{2.11c}$$
$$V^k(x,z) \geq (1-\beta)\|(x,z)\|_{N^k}^2, \tag{2.11d}$$
$$\left|\left\langle P^{-1}Ax, z\right\rangle\right| \leq \sqrt{\beta}\|x\|_{(\tau^k)^{-1}}\|z\|_{(P\Sigma^k)^{-1}}. \tag{2.11e}$$

证明。对所有 $i \in [\![1, n]\!]$，设 $(\eta_i^k)_{k\in\mathbb{N}}$ 和 $(\tilde{\eta}_i^k)_{k\in\mathbb{N}}$ 分别为 $(M_i^k)_{k\in\mathbb{N}}$ 和 $(N_i^k)_{k\in\mathbb{N}}$ 的控制序列。我们定义公共控制序列 $(\eta^k)_{k\in\mathbb{N}}$ 如下：

$$\eta^k = \max\left\{\max\left\{\eta_i^k, \tilde{\eta}_i^k\right\}, i \in [\![1,n]\!]\right\}, \quad k \in \mathbb{N}. \tag{2.12}$$

对所有 $y \in Y$，我们可以写成

$$U^{k+1}(y) = \sum_{i=1}^n \|(0,y_i)\|_{N_i^{k+1}}^2 \leq (1+\eta^k)\sum_{i=1}^n \|(0,y_i)\|_{N_i^k}^2 = (1+\eta^k)U^k(y),$$

这证明了 (2.11a)。现在固定 $x \in X$、$z \in Y_{\text{sparse}}$ 和 $k \in \mathbb{N}$。根据定义，存在 $i \in [\![1, n]\!]$，使得对所有 $j \neq i$，$z_j = 0$ 成立。通过写出以下不等式，我们得到 (2.11b)–(2.11d)：

$$\begin{aligned}
V^{k+1}(x,z) = \|(x,z_i)\|_{M_i^{k+1}}^2 &\leq (1+\eta^k)\|(x,z_i)\|_{M_i^k}^2 = (1+\eta^k)V^k(x,z),\\
\|(x,z)\|_{N^k}^2 = \|(x,z_i)\|_{N_i^k}^2 &\geq \alpha\|(x,z_i)\|^2 = \alpha\|(x,z)\|^2,\\
V^k(x,z) = \|(x,z_i)\|_{M_i^k}^2 &\geq (1-\beta)\|(x,z_i)\|_{N_i^k}^2 = (1-\beta)\|(x,z)\|_{N^k}^2.
\end{aligned}$$

最后，通过如下推导得到不等式 (2.11e)：

$$\begin{aligned}
\left|\left\langle P^{-1}Ax, z\right\rangle\right| &= \frac{1}{p_i}\left|\langle A_i x, z_i\rangle\right|\\
&\leq \frac{\|A_i\|}{p_i}\|x\|\|z_i\|\\
&= \frac{\|A_i\|}{p_i} * \left(\tau^k\sigma_i^k p_i\right)^{1/2}\|x\|_{(\tau^k)^{-1}}\|z\|_{(P\Sigma^k)^{-1}}\\
&\leq \sqrt{\beta}\|x\|_{(\tau^k)^{-1}}\|z\|_{(P\Sigma^k)^{-1}},
\end{aligned}$$

其中，最后一个不等式由 (2.5) 得出。

□

引理 2.8（A-SPDHG 是 $\mathcal{C}$-稳定的）。在定理 2.1 的假设下，

(i) 算法 2.1 的序列 $(x^k, y^k)_{k\in\mathbb{N}}$ 相对于 $(N^k)_{k\in\mathbb{N}}$，关于 $\mathcal{C}$ 是稳定的 ，

(ii) 以下结果成立：

$$\mathbb{E}\left[\sum_{k=1}^{\infty}\left\|(x^{k+1}-x^k, y^k-y^{k-1})\right\|^2\right] < \infty \quad \textit{and a.s.} \quad \left\|x^{k+1}-x^k\right\| \to 0.$$

证明。我们先证明第 (i) 点。根据采用串行抽样的 A-SPDHG 的定义，两个连续对偶迭代点之差几乎必然是稀疏的：

$$\text{a.s. } \forall\, k \in \mathbb{N}\setminus\{0\}\,, y^k - y^{k-1} \in Y_{\text{sparse}}.$$

定义以下序列

$$a^k = V^k(x^k - x^*, y^k - y^{k-1}) + U^k(y^k - y^*), \quad b^k = V^{k+1}(x^{k+1} - x^k, y^k - y^{k-1}),$$

根据 (2.11c) 和 (2.11d)，这些量几乎必然非负。注意，原始迭代点 x^l 从$l=0$到 $l=k+1$关于 $\mathcal{F}^k$可测，而对偶迭代点 y^l从 $l=0$到 $l=k$ 关于 $\mathcal{F}^k$可测。因此， a^k 和 b^k 关于 $\mathcal{F}^k$可测。此外，不等式 (2.10)、(2.11a) 和 (2.11b) 表明，几乎必然地，对所有 $k\in\mathbb{N}\setminus\{0\}$,

$$\mathbb{E}\left[a^{k+1}|\mathcal{F}^k\right] \le (1+\eta^k)a^k - b^k.$$

根据 Robbins-Siegmund 引理， [23], (a^k) 几乎必然收敛，$\sup_k \mathbb{E}\left[a^k\right] < \infty$ 且 $\sum_{k=1}^{\infty}\mathbb{E}\left[b^k\right] <\infty$。特别地，根据最后一点，利用 (2.11d) 和单调收敛定理可得：

$$\begin{aligned}\mathbb{E}\left[\sum_{k=1}^{\infty}\left\|y^k-y^{k-1}\right\|^2_{(P\Sigma^{k+1})^{-1}}\right] &\le \mathbb{E}\left[\sum_{k=1}^{\infty}\left\|(x^{k+1}-x^k, y^k-y^{k-1})\right\|^2_{N^{k+1}}\right]\\ &\le (1-\beta)^{-1}\,\mathbb{E}\left[\sum_{k=1}^{\infty} b^k\right] = (1-\beta)^{-1}\sum_{k=1}^{\infty}\mathbb{E}\left[b^k\right] < \infty,\end{aligned}$$

因此 $\sum_{k=1}^{\infty}\|y^k-y^{k-1}\|^2_{(P\Sigma^{k+1})^{-1}}$ 几乎必然有限，故 $\left(\|y^k-y^{k-1}\|^2_{(P\Sigma^{k+1})^{-1}}\right)_{k\in\mathbb{N}\setminus\{0\}}$ 进而 $\left(\|y^k-y^{k-1}\|_{(P\Sigma^{k+1})^{-1}}\right)_{k\in\mathbb{N}\setminus\{0\}}$ 几乎必然收敛到 0。此外，$\sup_k \mathbb{E}\left[a^k\right] < \infty$ 因此 $\sup_k \|x^k-x^*\|^2_{(\tau^k)^{-1}}$，进而 $\sup_k \|x^k-x^*\|_{(\tau^k)^{-1}}$ 均有限，并且根据 (2.11e)，对于 $k\in\mathbb{N}\setminus\{0\}$，可以写为：

$$\begin{aligned}\left|\left\langle P^{-1}A(x^k-x^*), y^k-y^{k-1}\right\rangle\right| &\le \sqrt{\beta}\|x^k-x^*\|_{(\tau^{k+1})^{-1}}\|y^k-y^{k-1}\|_{(P\Sigma^{k+1})^{-1}}\\ &\le \sqrt{\beta(1+\eta^k)}\|x^k-x^*\|_{(\tau^k)^{-1}}\|y^k-y^{k-1}\|_{(P\Sigma^{k+1})^{-1}}.\end{aligned}$$

我们知道，$(\eta^k)^{k\in\mathbb{N}}$ 可求和，因此收敛于 0。由此可得，

$$|\langle P^{-1}A(x^k-x^*), y^k-y^{k-1}\rangle| \to 0 \quad \text{almost surely.}$$

最后，借助恒等式

$$a^k = \|(x^k - x^*, y^k - y^*)\|_{N^k}^2 + \langle P^{-1}A(x^k - x^*), y^k - y^{k-1}\rangle, \quad k \in \mathbb{N} \setminus \{0\},$$

$(a^k)_{k\in\mathbb{N}}$ 几乎必然收敛，进而意味着 $(\|(x^k - x^*, y^k - y^*)\|_{N^k}^2)_{k\in\mathbb{N}}$ 几乎必然收敛。现在来看第（ii）点。第一个断言可直接由

$$\mathbb{E}\left[\sum_{k=1}^{\infty} b^k\right] = \sum_{k=1}^{\infty} \mathbb{E}\left[b^k\right] < \infty$$

以及界 (2.11c) 和 (2.11d) 得出。此外，这意味着 $\sum_{k=1}^{\infty} \left\|(x^{k+1} - x^k, y^k - y^{k-1})\right\|^2$ 几乎必然有限，因此 $\left(\left\|(x^{k+1} - x^k, y^k - y^{k-1})\right\|\right)$ 几乎必然收敛到 0，$\left(\left\|x^{k+1} - x^k\right\|\right)$ 也如此。 □

2.5 A-SPDHG 的弱聚点是鞍点

本节旨在证明 A-SPDHG 满足定理 2.5 的第 (iii) 点。在事件 $\left\{I^k = i\right\}$ 上，A-SPDHG 更新过程可改写为

$$y_i^{k+1} = \operatorname{prox}_{\sigma_i^{k+1} f_i^*}(y_i^k + \sigma_i^{k+1} A_i x^{k+1}), \quad \bar{y}_i^{k+1} = y_i^{k+1} + \frac{1}{p_i}\left(y_i^{k+1} - y_i^k\right), \quad \bar{y}_j^{k+1} = y_j^k, j \neq i$$
$$x^{k+2} = \operatorname{prox}_{\tau^{k+2} g}(x^{k+1} - \tau^{k+2} A^* \bar{y}^{k+1}).$$

我们定义 $T^{\sigma,\ \tau}_{i}: (x, y) \mapsto (\hat{x}, \hat{y}_i)$ 如下：

$$\hat{y}_i = \operatorname{prox}_{\sigma_i f_i^*}(y_i + \sigma_i A_i x), \quad \hat{x} = \operatorname{prox}_{\tau g}\left(x - \tau A^* y - \tau \frac{1 + p_i}{p_i} A_i^*(\hat{y}_i - y_i)\right),$$

因此，在事件 $\{I^k = i\}$ 上（且对于 $j \neq i$，$y_j^{k+1} = y_j^k$），$(x^{k+2}, y_i^{k+1}) = T^{\sigma_i^{k+1}\ ,\ \tau^{k+2}}_{i}(x^{k+1}, y^k)$。

引理 2.9（A-SPDHG 的聚点是鞍点）。设 $(\bar{x}, \bar{y})$ 几乎处处是 $(x^k, y^k)_{k\in\mathbb{N}}$ 的弱聚点（意即存在 Ω 的一个概率为一的可测子集 $\bar{\Omega}$，使得对于所有 $\omega \in \bar{\Omega}$，$(\bar{x}(\omega)\ \bar{y}(\omega))$ 都是 $(x^k(\omega)\ y^k(\omega))_{k\in\mathbb{N}}$ 的弱序列聚点），并假设定理 2.1 的假设成立。那么，$(\bar{x}, \bar{y})$ 几乎处处属于 $\mathcal{C}$。

证明。由引理 2.8-(ii) 和单调收敛定理可知，

$$\sum_{k=1}^{\infty} \mathbb{E}\left[\left\|(x^{k+1} - x^k, y^k - y^{k-1})\right\|^2\right] = \mathbb{E}\left[\sum_{k=1}^{\infty} \left\|(x^{k+1} - x^k, y^k - y^{k-1})\right\|^2\right] < \infty.$$

Now,

$$\begin{aligned}
\sum_{k=1}^{\infty} \mathbb{E}\left[\left\|(x^{k+1} - x^k, y^k - y^{k-1})\right\|^2\right] &= \sum_{k=1}^{\infty} \mathbb{E}\left[\mathbb{E}\left[\|(x^{k+1} - x^k, y^k - y^{k-1})\|^2 | I^{k-1}\right]\right] \\
&= \sum_{k=1}^{\infty} \sum_{i=1}^{n} \mathbb{P}(I^{k-1} = i) \mathbb{E}\left[\left\|T_i^{\sigma_i^k, \tau^{k+1}}(x^k, y_i^{k-1}) - (x^k, y_i^{k-1})\right\|^2\right] \\
&= \mathbb{E}\left[\sum_{i=1}^{n} p_i \sum_{k=1}^{\infty} \left\|T_i^{\sigma_i^k, \tau^{k+1}}(x^k, y^{k-1}) - (x^k, y_i^{k-1})\right\|^2\right].
\end{aligned}$$

因此我们可以推得

$$\mathbb{E}\left[\sum_{k=1}^{\infty}\sum_{i=1}^{n} p_i \left\| T_i^{\sigma_i^k,\tau^{k+1}}(x^k,y^{k-1}) - (x^k,y_i^{k-1})\right\|^2\right] < \infty.$$

由此可知，期望中的级数几乎必然有限；又由于 $p_i > 0$，我们推得几乎必然有，

$$\left\| T_i^{\sigma_i^k,\tau^{k+1}}(x^k,y^{k-1}) - (x^k,y_i^{k-1})\right\| \overset{k\to\infty}{\longrightarrow} 0 \tag{2.13}$$

对所有 $i = 1$、$\ldots n$ 均成立。考虑一个有界的样本 (x^k, y^k)，且其满足 (2.13)。对每个 i，令 $(\hat{x}^{i,\ k+1}, \hat{y}^{i,\ _i\ k}) = T^{\sigma_i^k,\ _i\ \tau^{k+1}}(x^k, y^{k-1})$，使得 $\|(\hat{x}^{i,\ k+1}, \hat{y}^{i,\ _i\ k}) - (x^k, y_i^{k-1})\| \to 0$ 对$i = 1$，$\ldots\ldots$，n成立。于是有

$$\begin{aligned}
\partial f_i^*(\hat{y}_i^{i,k}) &\ni \frac{y_i^{k-1} - \hat{y}_i^{i,k}}{\sigma_i^k} + A_i x^k =: A_i x^k + \delta_y^{i,k} \\
\partial g(\hat{x}^{i,k+1}) &\ni \frac{x^k - \hat{x}^{i,k+1}}{\tau^{k+1}} - A^* y^{k-1} - \frac{1+p_i}{p_i} A_i^*(\hat{y}_i^{i,k} - y_i^{k-1}) =: -A^* y^{k-1} + \delta_x^{i,k}
\end{aligned}$$

其中 $\delta^{i,\ _x\ k,\ _y} \to 0$ 当 $k \to \infty$ 时成立。给定一个测试点 (x, y)，对任意 k，可写为:

$$\begin{aligned}
f_i^*(y_i) &\geq f_i^*(\hat{y}_i^{i,k}) + \langle A_i x^k, y_i - y_i^{k-1}\rangle + \langle A_i x^k, y_i^{k-1} - \hat{y}_i^{i,k}\rangle + \langle \delta_y^{i,k}, y_i - \hat{y}_i^{i,k}\rangle, \quad i = 1,\ldots,n \\
g(x) &\geq g(\hat{x}^{1,k+1}) - \langle A^* y^{k-1}, x - x^k\rangle - \langle A^* y^{k-1}, x^k - \hat{x}^{1,k+1}\rangle + \langle \delta_x^{i,k}, x - \hat{x}^{1,k+1}\rangle
\end{aligned}$$

将这些不等式全部相加，可得:

$$g(x) + \sum_{i=1}^{n} f_i^*(y_i) \geq g(\hat{x}^{1,k+1}) + \sum_{i=1}^{n}\left(f_i^*(\hat{y}_i^{i,k}) + \langle A_i x^k, y_i\rangle\right) - \langle A^* y^{k-1}, x\rangle + \delta^k$$

其中 $\delta^k \to 0$ 当 $k \to \infty$ 时成立。我们推得，若 $(\bar{x}, \bar{y})$ 是子序列 (x^{k_l}, y^{k_l-1}) 的弱极限（当然，$(x^{k_l}$, $y^{k_l})$ 也同样如此），则:

$$g(x) + \sum_{i=1}^{n} f_i^*(y_i) \geq g(\bar{x}) + \sum_{i=1}^{n}(f_i^*(\bar{y}_i) + \langle A_i\bar{x}, y_i\rangle) - \langle A^*\bar{y}, x\rangle.$$

由于 (x, y) 是任意的，我们发现 (2.2) 对 $(\bar{x}, \bar{y})$ 成立。

□

2.6 定理 2.1 的证明

在定理 2.1 的假设下，$\mathcal{C}$ 鞍点集合是闭且非空的，并且$X \times Y$是可分希尔伯特空间。根据引理 2.3，在 (2.8) 中定义的变度量序列 $(N^k)_{k\in\mathbb{N}}$满足定理 2.5 的条件 (i)。此外，根据引理 2.8 和引理 2.9，算法 2.1 的迭代点分别满足定理 2.5 的条件 (ii) 和 (iii)，因此几乎必然收敛到 $\mathcal{C}$中的一点。

3 算法设计与实际实现

本节介绍 A-SPDHG 算法的实际应用实例，并给出一种满足收敛性证明中假设的步长调整规则。我们将 [14], 提出的用于确定性PDHG的自适应步长平衡规则扩展到随机设置，并采用小批量近似以尽量减少计算开销。

3.1 A-SPDHG规则（a）–跟踪并平衡原始-对偶进展 ss

首先简要介绍我们第一个数值方案的基础。该方案建立在 Goldstein 等人 [14], 提出的确定性自适应 PDHG 算法之上，其迭代点为:

$$x^{k+1} = \mathrm{prox}_{\tau^{k+1}g}(x^k - \tau^{k+1}A^*y^k), \quad y^{k+1} = \mathrm{prox}_{\sigma^{k+1}f^*}(y^k + \sigma^{k+1}A(2x^{k+1} - x^k))$$

在 Goldstein 等人 [14], 的这项基础工作中，他们提出评估两个序列，以跟踪并平衡确定性 PDHG 的原始和对偶迭代点（此处分别记为 v_k^* 和 d_k^*）的进展:

$$v_k^* := \|(x^k - x^{k+1})/\tau^{k+1} - A^*(y^k - y^{k+1})\|_1, \quad d_k^* := \|(y^k - y^{k+1})/\sigma^{k+1} - A(x^k - x^{k+1})\|_1. \tag{3.1}$$

这两个序列衡量目标函数 $\min_{x\in X}\max_{y\in Y} g(x) + \langle Ax, y\rangle - f^*(y)$ 的原始和对偶次梯度长度，这一点可由近端算子的定义推得。确定性 PDHG 的原始更新可写为:

$$x^{k+1} = \arg\min_x \frac{1}{2}\|x - (x^k - \tau^{k+1}A^*y^k)\|_2^2 + \tau^{k+1}g(x). \tag{3.2}$$

上述目标函数的最优性条件表明:

$$0 \in \partial g(x^{k+1}) + A^*y^k + \frac{1}{\tau^{k+1}}(x^{k+1} - x^k). \tag{3.3}$$

在等式两边加上 $-A^*y^{k+1}$ 并重新整理各项，可得:

$$(x^k - x^{k+1})/\tau^{k+1} - A^*(y^k - y^{k+1}) \in \partial g(x^{k+1}) + A^*y^{k+1} \tag{3.4}$$

类似地，对于对偶更新，也可以推导出:

$$(y^k - y^{k+1})/\sigma^{k+1} - A(x^k - x^{k+1}) \in \partial f^*(y^{k+1}) - Ax^{k+1}, \tag{3.5}$$

这表明，由 (3.1) 给出的序列 v_k^* 和 d_k^* 应能有效跟踪确定性PDHG的原始进展和对偶进展。因此，Goldstein 等人 [14] 提出以此为基础，平衡 PDHG 的原始步长和对偶步长。

鉴于此，我们在算法 3.1 中提出了 A-SPDHG 的首个实用实现，称为规则 (a)：对所有迭代点 k 和索引 j 使用统一的对偶步长 $\sigma^k = \sigma_j^k$，并通过每次迭代 k 时定义的两个序列 v^k 和 d^k 来估计原始变量和对偶变量达到最优性的进展，其中 $I^k = i$ 为:

$$v_{k+1} := \|(x^k - x^{k+1})/\tau^{k+1} - \frac{1}{p_i}A_i^*(y_i^k - y_i^{k+1})\|_1, \quad d_{k+1} := \frac{1}{p_i}\|(y_i^k - y_i^{k+1})/\sigma^{k+1} - A_i(x^k - x^{k+1})\|_1, \tag{3.6}$$

这两个序列是针对我们的随机设置定制的 (3.1) 的小批量扩展。通过在适当时调整原始-对偶步长比，使它们动态保持平衡，我们可以促使算法在原始步和对偶步中取得相近的进展，从而改善收敛。具体而言，如算法 3.1 所示，在每次迭代中都会计算并比较 v_k 和 d_k 的值。如果 v_k （用于跟踪原始次梯度）的值显著大于 d_k （用于跟踪对偶次梯度），则说明原始进展慢于对偶进展，因此算法会增大原始步长，同时减小对偶步长。如果 v_k 明显小于 d_k ，算法则会采取相反的操作。

请注意，此处我们采用 ℓ_1-范数作为 v^k 和 d^k 的长度度量，这与 Goldstein 等人 [14, 15], 的做法一致，因为我们也通过数值实验观察到，相较于更直观的 ℓ_2-范数选择，这种选择更有优势。

对于全批量情形（$n = 1$），它简化为 [14, 15] 提出的自适应 PDHG。我们根据 v^k 与 d^k 的比值调整原始步长与对偶步长的比值；每当步长发生变化时，我们就将 α （用于控制变化幅度）缩小至原来的 $\eta \in$ 倍，其中 $\eta \in \in$(0, 1)——在我们的实验中，通常取 $\eta = 0.995$ 。对于 s 的选择，我们默认采用 $s = \|A\|$ 。[1]

3.1.1 通过子采样减少开销：

请注意，与确定性情形不同，后者无需额外进行矩阵-向量乘法，因为 A^*y^k 和 Ax^k 可以被记忆下来；而我们的随机扩展则需要计算 A_ix^k ，因为相邻两次迭代之间很可能会抽取不同的子集。采用这一策略时，浮点运算次数方面的开销最多为 50%，与其为 SPDHG 带来的显著加速相比，这在数值上可以忽略不计；尤其是在原始-对偶步长比并非最优时，更是如此，后续实验中我们将对此进行展示。此外，我们通过数值实验发现，利用子采样等近似技巧可以显著降低这一开销：

$$d^{k+1} \approx \frac{\rho}{p_i}\|S^k(y_i^k - y_i^{k+1})/\sigma^{k+1} - S^kA_i(x^k - x^{k+1})\|_1 \tag{3.7}$$

其中， S^k 是一个随机子采样算子，满足 $\mathbb{E}[(S^k)^TS^k] = \frac{1}{\rho}$Id。在实验中，我们为此近似采用 10% 的子采样，因此开销从 50% 降至仅 5%，可以忽略不计，而且在实践中不会影响收敛率。

3.2 A-SPDHG 规则（b）– 利用角度对齐

最近，Yokota 和 Hontani [26] 提出了一种 PDHG 自适应步长平衡方案的变体，利用子梯度 $\partial g(x^{k+1}) + A^*y^{k+1}$ 与更新差值 $x^k - x^{k+1}$之间的夹角。

如果这两个方向高度一致，就可以增大原始步长，以采取更大的步幅。如果这两个方向之间的角度较大，则应缩小原始步长。将这一方案扩展到随机设置后，我们便得到 SPDHG 的另一种自适应方案。

[1]对于规则（a）的收敛行为， s 的选择至关重要。数值实验表明，采用算子范数 $\|A\|$ 进行缩放优于依赖像素值范围的做法，后者是 [15]中建议的方法。

算法 3.1：A-SPDHG，规则（a）

输入：对偶步长 σ^0，原始步长 $\tau^0, \alpha^0 \in (0,1)$，$\eta \in (0,1)$，$\delta > 1$，概率 $(p_i)_{1\le i\le n}$；原始变量 x^0，对偶变量 y^0

初始化 $\bar{y}^0 = y^0, v^0 = d^0 = 0, s = \|A\|$

对于 $k \in [\![0, K-1]\!]$，执行

如果 $v^k > sd^k\delta$，则 $\tau^{k+1} = \frac{\tau^k}{1-\alpha^k}$, $\sigma^{k+1} = \sigma^k(1-\alpha^k)$, $\alpha^{k+1} = \alpha^k\eta$

如果 $v^k < sd^k/\delta$，则 $\tau^{k+1} = \tau^k(1-\alpha^k)$, $\sigma^{k+1} = \frac{\sigma^k}{1-\alpha^k}$, $\alpha^{k+1} = \alpha^k\eta$

如果 $sd^k/\delta \le v^k \le sd^k\delta$，则 $\tau^{k+1} = \tau^k$, $\sigma^{k+1} = \sigma^k$, $\alpha^{k+1} = \alpha^k$

$x^{k+1} =$近端$_{\tau^{k+1}g}(x^k - \tau^{k+1}A^*\bar{y}^k)$

以概率 p_i随机选取 $i \in [\![1, n]\!]$

$$y_j^{k+1} = \begin{cases} \text{近端算子}_{\sigma^{k+1}f_i^*}(y_i^k + \sigma^{k+1}A_i x^{k+1}) & \text{若 } j = i \\ y_j^k & \text{若 } j \ne i \end{cases}$$

$$\bar{y}_j^{k+1} = \begin{cases} y_i^{k+1} + \frac{1}{p_i}\left(y_i^{k+1} - y_i^k\right) & \text{若 } j = i \\ y_j^k & \text{若 } j \ne i \end{cases}$$

$v^{k+1} = \|(x^k - x^{k+1})/\tau^{k+1} - \frac{1}{p_i}A_i^*(y_i^k - y_i^{k+1})\|_1$

$d^{k+1} = \frac{1}{p_i}\|(y_i^k - y_i^{k+1})/\sigma^{k+1} - A_i(x^k - x^{k+1})\|_1$ ——或使用（3.7）近似此步骤

结束循环

返回 x^K

我们在算法 3.2 中将这一方案列为规则（b）。在迭代 k 且满足 $I^k = i$ 时，计算 e:

$$q^{k+1} = (x^k - x^{k+1})/\tau^{k+1} - \frac{1}{p_i}A_i^*(y_i^k - y_i^{k+1}), \tag{3.8}$$

作为 $\partial g(x^{k+1}) + A^*y^{k+1}$ 的估计值，然后测量它与 $x^k - x^{k+1}$ 之间夹角的余弦：

$$w^{k+1} = \frac{\langle x^k - x^{k+1}, q^{k+1}\rangle}{(\|x^k - x^{k+1}\|_2\|q^{k+1}\|_2)}. \tag{3.9}$$

余弦值的阈值 c（当余弦值达到该阈值时，原始步长会增大）通常需要非常接近 1（我们采用 $c = 0.999$），因为我们主要在高维问题中应用这类算法，这遵循了 [26] 中针对确定性 PDHG 的选择。

最近，Zdun 等人 [27] 提出了一种与我们的规则（b）类似的启发式方法，但他们选择 q^{k+1} 作为 $\partial g(x^{k+1})$ 中某个元素的近似，而不是 $\partial g(x^{k+1}) + A^*y^{k+1}$。我们的选择更接近 Yokota 和 Hontani [26] 的原始方案。通过数值实验，我们发现他们的方案在我们的设置中不具竞争力。

算法 3.2：A-SPDHG，规则（b）

输入：对偶步长 σ^0，原始步长 τ^0, $\eta \in (0,1)$，概率 $(p_i)_{1\le i\le n}$；原始变量 x^0，对偶变量 y^0

初始化 $\bar{y}^0 = y^0$, $w^0 = 0$, $\alpha^0 = 1$

对于 $k \in [\![0, K-1]\!]$，执行

如果 $w^k < 0$，则 $\tau^{k+1} = \frac{\tau^k}{1+\alpha^k}$, $\sigma^{k+1} = \sigma^k(1+\alpha^k)$, $\alpha^{k+1} = \alpha^k \eta$

如果 $w^k \ge c$，则 $\tau^{k+1} = \tau^k(1+\alpha^k)$, $\sigma^{k+1} = \frac{\sigma^k}{1+\alpha^k}$, $\alpha^{k+1} = \alpha^k \eta$

如果 $0 \le w^k < c$，则 $\tau^{k+1} = \tau^k$, $\sigma^{k+1} = \sigma^k$, $\alpha^{k+1} = \alpha^k$

$x^{k+1} = $近端算子$_{\tau^{k+1} g}(x^k - \tau^{k+1} A^* \bar{y}^k)$

以概率 p_i 随机选取 $i \in [\![1, n]\!]$

$$y_j^{k+1} = \begin{cases} \text{近端算子}_{\sigma^{k+1} f_i^*}(y_i^k + \sigma^{k+1} A_i x^{k+1}) & \text{如果 } j = i \\ y_j^k & \text{如果 } j \ne i \end{cases}$$

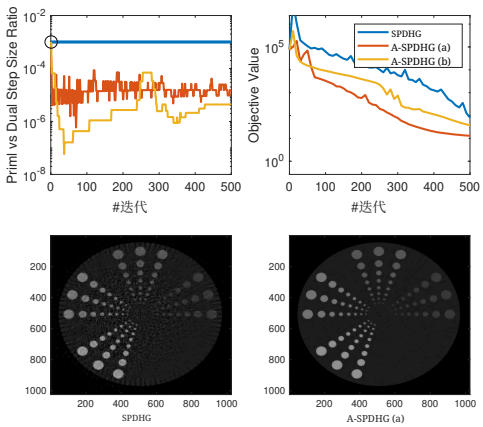
$$\bar{y}_j^{k+1} = \begin{cases} y_i^{k+1} + \frac{1}{p_i}\left(y_i^{k+1} - y_i^k\right) & \text{如果 } j = i \\ y_j^k & \text{如果 } j \ne i \end{cases}$$

$q^{k+1} = (x^k - x^{k+1})/\tau^{k+1} - \frac{1}{p_i} A_i^*(y_i^k - y_i^{k+1})$

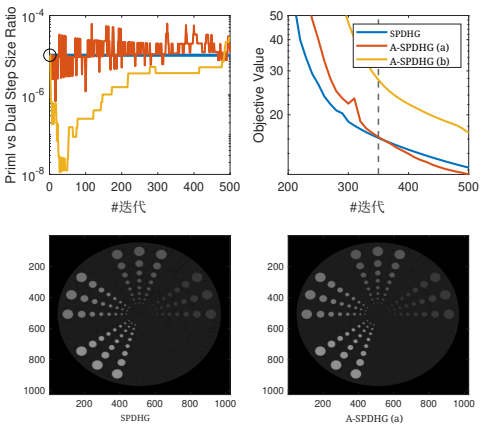
$w^{k+1} = \langle x^k - x^{k+1}, q^{k+1} \rangle / (\|x^k - x^{k+1}\|_2 \|q^{k+1}\|_2)$

结束循环

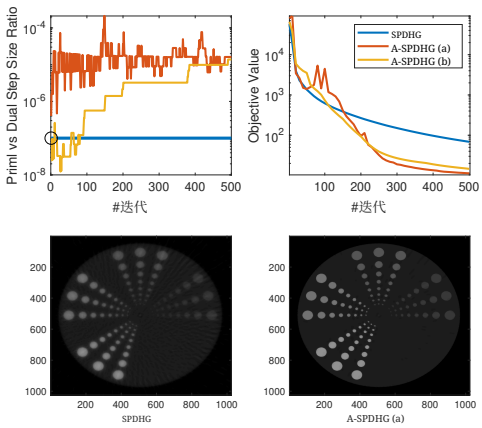
返回 x^K



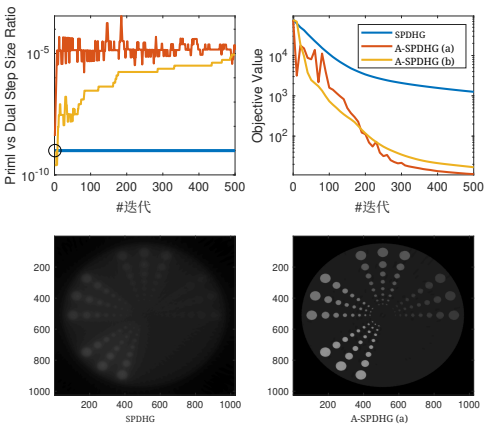
(a) 初始比率 10^{-3}



(b) 初始比率 10^{-5}

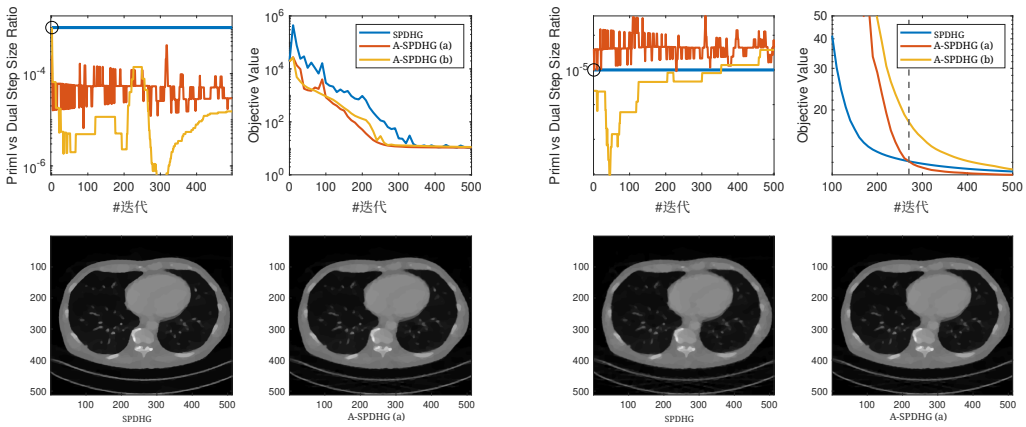


(c) 初始比率 10^{-7}



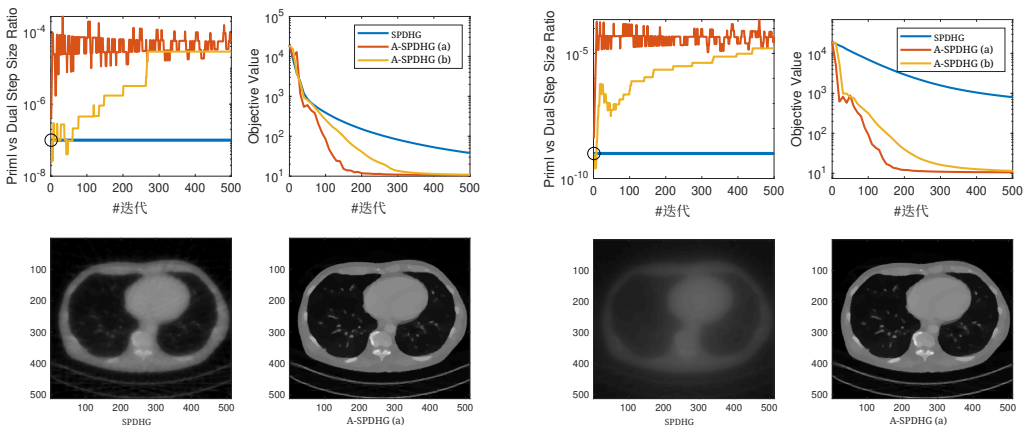
(d) 初始比率 10^{-9}

图 1：在稀疏视角计算机断层成像（示例 1）中，比较 SPDHG 和 A-SPDHG 在多种初始原始-对偶步长比下的表现。这里的前向算子为 $A \in \mathbb{R}^{m\times d}$，其中维度为 $m = 368640$、$d = 1048576$。我们还展示了算法终止时（第50个周期）重建的图像。每个子图的第一幅图中，黑色圆圈表示所有算法的初始步长比；后续图中也同样如此。



(a) 初始比率 10^{-3}

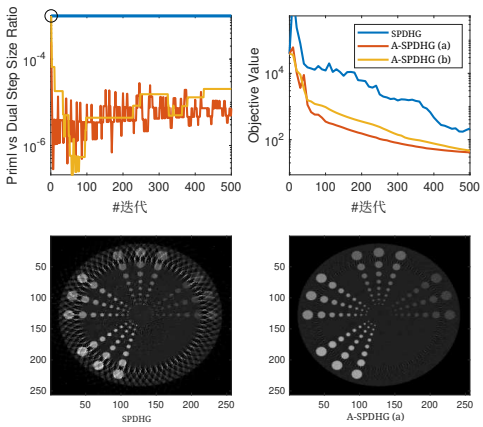
(b) 初始比率 10^{-5}



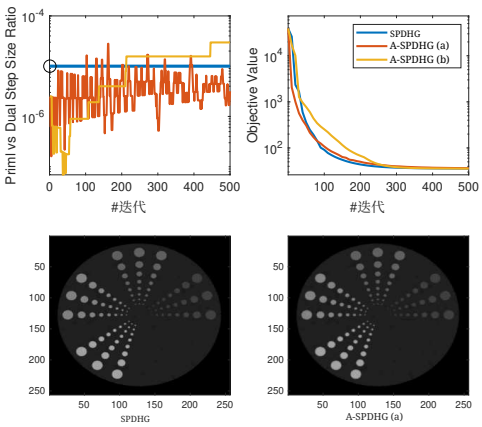
（c）初始比率 10^{-7}

(d) 初始比率 10^{-9}

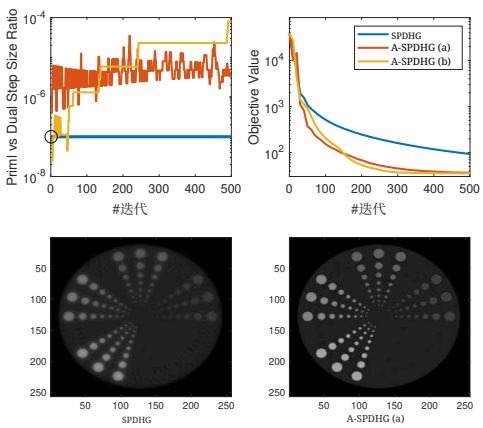
图2：比较不同初始原始-对偶步长比下，SPDHG 与 A-SPDHG 在稀疏视角计算机断层成像（例2）中的表现。此处前向算子为 $A \in \mathbb{R}^{m\times d}$，其中维度为 $m = 92160$、$d = 262144$。我们展示了算法终止时（第50个周期）重建的图像。



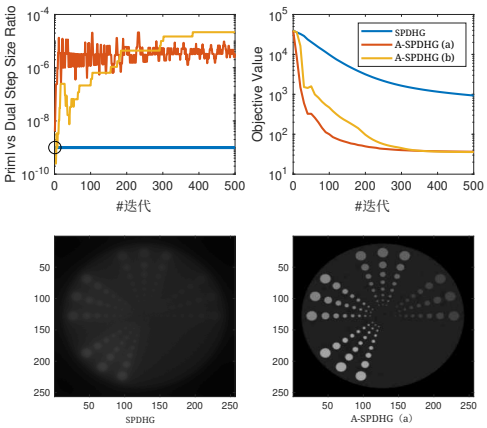
(a) 初始比率 10^{-3}



(b) 初始比率 10^{-5}

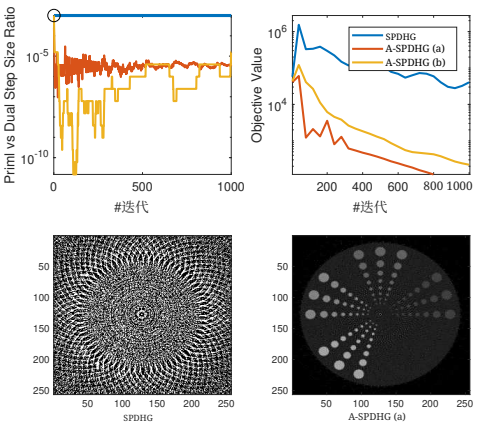


(c) 初始比率 10^{-7}

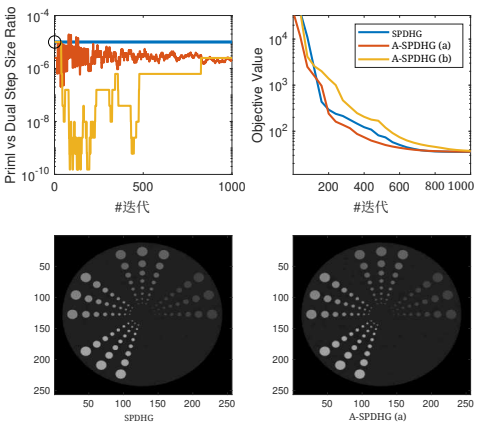


（d）初始比率 10^{-9}

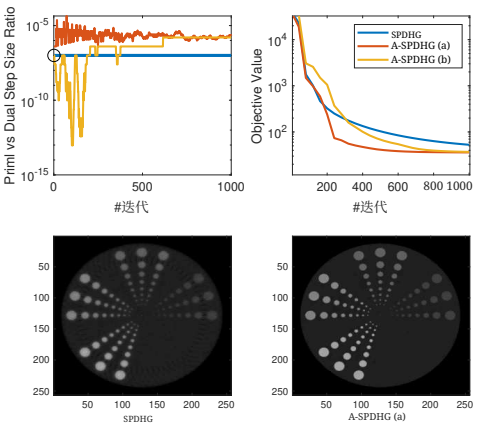
图3：在低剂量计算机断层扫描（其中我们使用了大量噪声很大的X射线测量）中，对比 SPDHG 与 A-SPDHG，并采用多种初始原始-对偶步长比。此处前向算子 $A \in \mathbb{R}^{m\times d}$，其中维度为 $m = 184320$、 $d = 65536$。我们将模体图像调整为 256×256。图中包含算法在终止时（第50个周期）重建的图像。



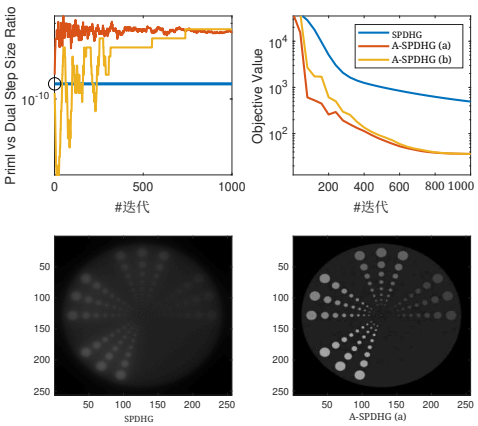
(a) 初始比率 10^{-3}



(b) 初始比率 10^{-5}

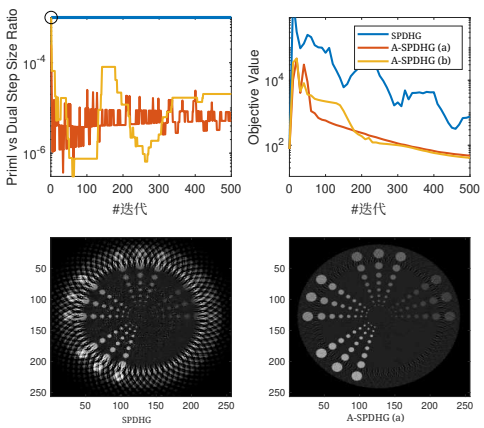


(c) 初始比率 10^{-7}

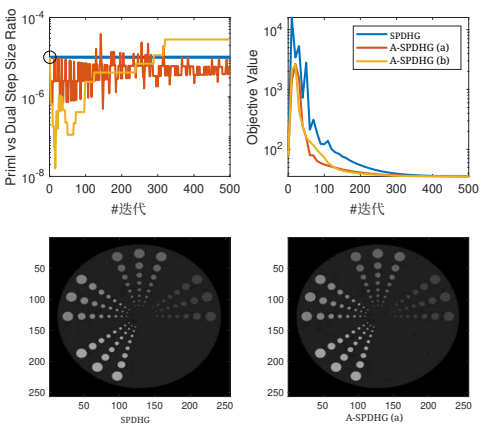


(d) 初始比率 10^{-9}

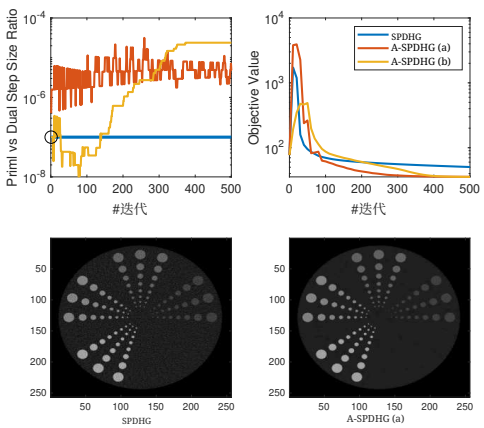
图4：在低剂量计算机断层扫描中，将数据划分为40个小批次，对比SPDHG与A-SPDHG。与图3中使用10个小批次的结果相比，我们得到的结果相近，而且A-SPDHG的表现仍优于SPDHG。



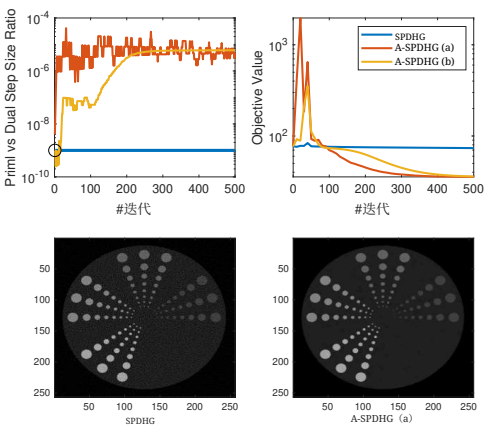
(a) 初始比率 10^{-3}



(b) 初始比率 10^{-5}

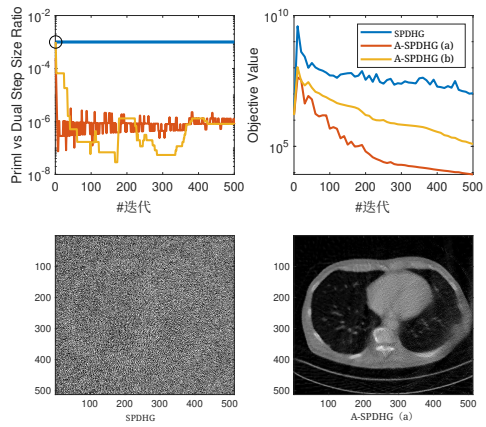


（c）初始比率 10^{-7}

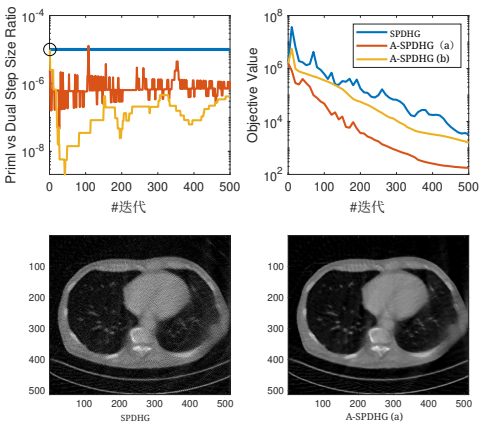


（d）初始比率 10^{-9}

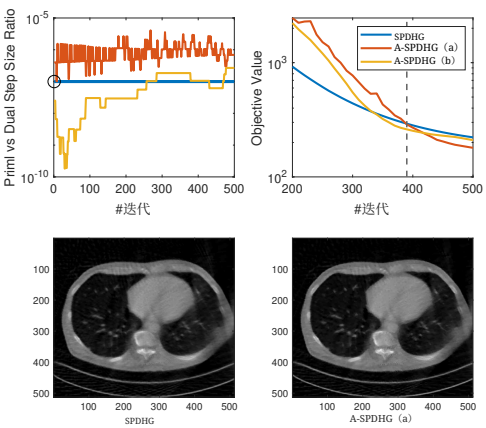
图 5：比较使用滤波反投影（FBP）对低剂量计算机断层扫描进行热启动时，SPDHG 与 A-SPDHG 的表现。与图3所示未采用热启动的结果相比，采用热启动后，我们的方法实际上表现得更为出色。另请注意，由于算法的随机性，函数值在早期出现跃升符合我们的预期。我们展示了算法在终止时（第50个周期）重建的图像。



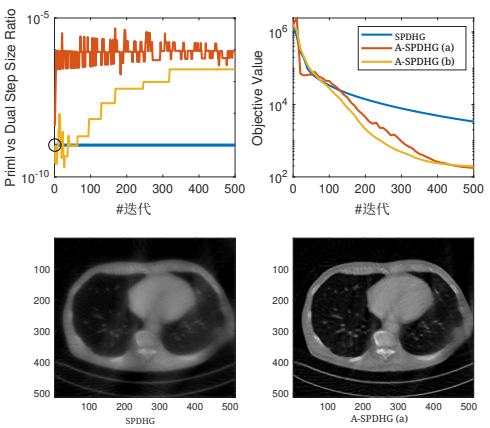
（a）初始比率 10^{-3}



(b) 初始比率 10^{-5}



（c）初始比率 10^{-7}



(d) 初始比率 10^{-9}

图6：在有限角度 CT（例2）中，对比 SPDHG 和 A-SPDHG 在多种初始原始-对偶步长比下的表现。这里的前向算子为 $A \in \mathbb{R}^{m\times d}$，其中维度为 $m = 92160$、 $d = 262144$。我们还展示了算法终止时（第50个周期）重建的图像。

4 数值实验

本节将通过数值研究，展示所提出的方案在求解一种最典型的成像逆问题——计算机断层扫描（CT）中的表现。我们在原始步长和对偶步长的不同初始比率下，将 A-SPDHG 算法与原始 SPDHG 进行比较。

在 CT 成像示例中，我们通过求解以下全变分正则化目标函数，从扇束 X 射线测量数据中重建断层扫描图像：

$$x^\star \in \arg\min_{x\in\mathbb{R}^d} \frac{1}{2}\|Ax-b\|_2^2 + \lambda\|Dx\|_1 \tag{4.1}$$

其中， D 表示二维微分算子， $A\in\mathbb{R}^{m\times d}$ 和 $x\in\mathbb{R}^d$。我们考虑三种扇束 CT 成像模式：稀疏视角 CT、低剂量计算机断层扫描和有限角度 CT。我们在两幅不同尺寸的图像上测试 A-SPDHG 和 SPDHG，并采用 4 种不同的初始比率（10^{-3}、 10^{-5}、 10^{-7} 和 10^{-9}）： 示例 1 使用尺寸为 1024×1024的模体图像，示例 2 使用来自梅奥诊所数据集的图像 [21]，尺寸为 512×512。我们将测量数据和算子交错划分为 $n=10$ 个小批次，供两种算法使用。具体而言，我们首先收集所有 X 射线测量数据，并按从 0 度到 360 度的顺序依次排列，构成完整的 A和 b；然后每隔 10 个测量值交错分组为一个小批次，形成划分 $\{A_i\}_{i=1}^{10}$ 和 $\{b_i\}_{i=1}^{10}$。

对于 A-SPDHG，我们选择使用式 (3.7) 中给出的 d^k 近似步骤，并采用 10% 的子采样，因此在本实验中计算开销可以忽略不计。我们将所有算法的初始值设为零图像。

我们在图1、图2、图3和图6中展示数值结果。在这些图中，我们通过迭代次数比较各算法的收敛率（各算法每次迭代的执行时间几乎相同，因为 A-SPDHG 的计算开销在数值上微不足道）。其中，图1和图2报告了针对梅奥诊所数据集 [21], 中的模体图像和肺部 CT 图像开展的大规模稀疏视角 CT 实验结果；图3报告了低剂量计算机断层扫描实验结果，其中我们模拟了大量受到显著泊松噪声污染的测量；图6则报告了有限角度 CT 的结果，该实验仅包含 0 度至 150 度范围内测量角度的数据，其余 $[150,360]$ 度角度对应的测量数据均缺失。在所有这些示例中，我们都能一致地观察到，无论如何初始化原始-对偶步长比，A-SPDHG 都能自动且稳定地将步长比调整至最优值；对于这四种不同的 CT 问题，最优值约为 10^{-5} 或 10^{-7} 。当初始比率偏离最优范围时，A-SPDHG 的表现显著优于原始 SPDHG。与此同时，即使 SPDHG 算法的初始比率接近最优值，我们也能从大多数示例中一致地观察到，在经过一定次数的迭代后，我们的方法在局部仍优于原始 SPDHG 算法（相关子图中的竖直虚线标出了这一点），这进一步表明自适应性对这类算法的益处[2]。值得注意的是，在所有这些不同的示例中，我们始终只对 A-SPDHG 使用上一节建议的一组固定参数，这再次体现了我们方案的实用性。

对于低剂量计算机断层扫描示例，我们另外进行了两组实验：图4中使用更多的小批次划分（40 个），以及图5中从通过滤波反投影得到的更优初始化图像进行热启动

[2]这里最典型的例子是图1(b)：在收敛时，自适应方案选出的最优步长比几乎恰好为 10^{-5}，而我们已将 SPDHG 设置为以该比值运行。我们仍然可以观察到自适应方案带来的局部收敛加速效果。

。我们发现，在所有这些额外示例中，A-SPDHG 的表现始终优于标准 SPDHG，尤其是在原始-对偶步长比不理想时。有趣的是，我们发现热启动的效果并未明显影响 SPDHG 与 A-SPDHG 之间的相对表现。这主要是因为在实际应用中，采用不理想原始-对偶步长比的 SPDHG 在高精度阶段收敛非常缓慢（例如见图 5(d)），因此热启动在这里帮助不大。

还应指出，从概念上讲，我们的自适应方案中的所有超参数本质上都是控制算法自适应性的控制序列（在极端取值下，我们会得到原始的 SPDHG）。图 7 和图 9 展示了关于 A-SPDHG 算法规则（a）和规则（b）的超参数选择的一些数值研究。在这些实验中，我们为原始-对偶步长选择了固定的初始比率 10^{-7}。对于规则（a），我们发现，该规则对初始收缩率 α_0、收缩速度 η 和间隔 δ 的选择具有较强的鲁棒性。总体而言，我们发现这些参数对规则（a）的收敛性能影响较小，也很容易选取。

对于规则（b），我们发现，与规则（a）相比，其性能对参数 c 和 η 的选择更为敏感，不过这种依赖性仍然较弱。数值研究表明，规则（a）的表现优于规则（b），但两者各有一些小缺点（第一种规则会带来少量计算开销，可通过子采样方案部分缓解；而第二种规则似乎往往比第一种更慢），这仍有待进一步研究和改进。不过需要强调的是，所有这些参数本质上都是在控制算法的自适应程度，而且相当容易选取。值得注意的是，对于这些涉及不同尺寸、维度和成像模态的计算机断层扫描实验，我们在 A-SPDHG 中仅使用了一组固定的超参数，却已能稳定地观察到相较于原始 SPDHG 的数值改进。

5 结论

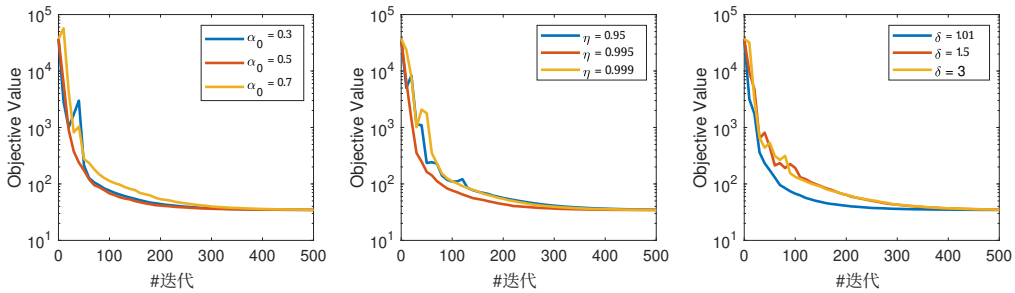
本文提出了一种新框架（A-SPDHG），用于随机原始-对偶混合梯度法中的自适应步长平衡。我们首先从理论上推导出一组关于自适应原始步长和对偶步长的充分条件，以确保算法在随机设置下收敛。随后，我们提出了若干满足收敛条件的实用方案；在成像逆问题上的数值结果也证实了所提方法的有效性。

据我们所知，本文首次对随机原始-对偶算法的自适应步长进行了理论分析。我们目前的工作包括对采用线搜索方案的进一步加速随机原始-对偶方法进行理论分析和算法设计，以获得更快的收敛率。

6 第 2 节的补充材料

我们先介绍一个有用的引理。

引理 6.1。设 a、b 为正标量，$\beta \in$ 属于 (0, 1)，且 P 是从希尔伯特



(a) 对 α_0的选项进行测试　(b) 对 η 的选择进行测试　(c) 对 δ的选择进行测试

图 7：在 X射线低剂量扇束CT 示例上测试 A-SPDHG（rule-a）的不同参数选择，原始-对偶步长的初始比率为：10^{-7}。可以观察到，ASPDHG 的性能对这些参数选择的依赖很小。

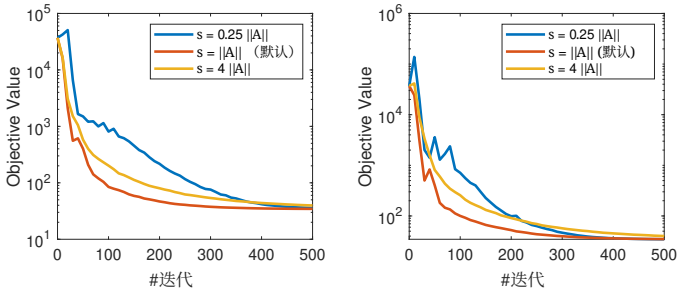


图 8：在 X射线低剂量扇束CT 示例上测试 A-SPDHG 的默认选择 $s = \|A\|$ （rule-a）。左图：原始-对偶步长的初始比率：10^{-7}。右图：原始-对偶步长的初始比率：10^{-5}。我们可以观察到，默认选择 s 确实是实践中合理的选择（至少接近最优），而偏离该选择可能会导致收敛变慢。

空间 X 到希尔伯特空间 Y 的有界线性算子。则，

$$(ab)^{-1/2}\|P\| \le 1 \quad \Leftrightarrow \quad \begin{pmatrix} a\,Id & P^* \\ P & b\,Id \end{pmatrix} \succcurlyeq 0. \tag{6.1}$$

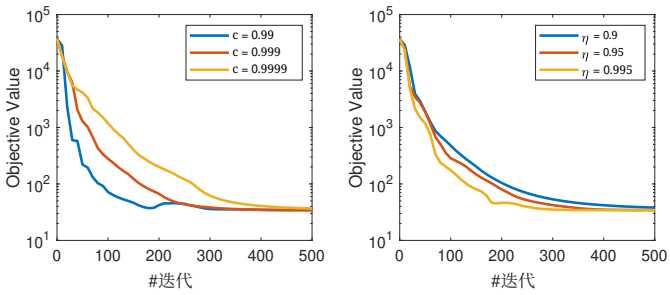
$$(ab)^{-1/2}\|P\| \le \beta \quad \Leftrightarrow \quad \begin{pmatrix} a\,Id & P^* \\ P & b\,Id \end{pmatrix} \succcurlyeq (1-\beta)\begin{pmatrix} a\,Id & 0 \\ 0 & b\,Id \end{pmatrix}. \tag{6.2}$$

证明。我们称

$$M = \begin{pmatrix} a\,\mathrm{Id} & P^* \\ P & b\,\mathrm{Id} \end{pmatrix}.$$

对于所有 $(x, y) \in X \times Y$，

$$\|(x,y)\|_M^2 \ge a\|x\|^2 + b\|y\|^2 - 2\|P\|\|x\|\|y\| = \|x\|_a^2 + \|y\|_b^2 - 2(ab)^{-1/2}\|P\|\|x\|_a\|y\|_b,$$



(a) 对 c选项的测试　　(b) 测试 η的选择

图9：在X射线低剂量扇束CT示例上测试A-SPDHG（规则-b）的不同参数选择，原始-对偶步长的初始比率： 10^{-7}。

这证明了（6.1）的直接蕴含。对于逆向蕴含，考虑满足 $\|Px\| = \|P\|\|x\|$ 且 $y = -\lambda Px$ 的 $x \in X \setminus \{0\}$，其中 λ 为标量。于是，多项式的非负性

$$\frac{\|(x,y)\|_M^2}{\|x\|^2} = b\|P\|^2\lambda^2 - 2\|P\|^2\lambda + a$$

对于所有 $\lambda \in \mathbb{R}$，意味着 $\|P\|^4 - ab\|P\|^2 \leq 0$，这等价于所需的结论（$ab$）$^{-1/2}\|P\| \leq 1$。

通过注意到以下事实，等价关系 (6.2) 显而易见：

$$\begin{pmatrix} a\,\mathrm{Id} & P^* \\ P & b\,\mathrm{Id} \end{pmatrix} \succcurlyeq (1-\beta)\begin{pmatrix} a\,\mathrm{Id} & 0 \\ 0 & b\,\mathrm{Id} \end{pmatrix} \Leftrightarrow \begin{pmatrix} \beta a\,\mathrm{Id} & P^* \\ P & \beta b\,\mathrm{Id} \end{pmatrix} \succcurlyeq 0.$$

□

现在来证明引理 2.2。

引理 2.2 的证明。假设步长满足引理中的条件。于是，定理 2.1 的假设 (i) 显然成立。此外，对于 $i \in [\![1,n]\!]$，由方程 (2.6) 可知，乘积序列 $(\tau^k\sigma_i^k)_{k\in\mathbb{N}}$ 在迭代过程中保持不变，并且对于迭代 $k = 0$ 满足方程 (2.5)，因此对于所有满足 $\beta = \max_i \left\{\tau^0\sigma_i^0\|A_i\|^2/p_i\right\}$ 的 $k \in \mathbb{N}$，均满足 (2.5)，这证明了假设 (ii)。最后，方程 (2.7) 表明假设 (iii) 成立。 □

7 声明

7.1 伦理批准

此声明不适用。

7.2 利益冲突

没有需要声明的利益冲突。

7.3 作者贡献

CD、MJE 和 AC 阐述了证明策略，CD 撰写了第 1、2 和 6 部分。JT 负责算法设计，进行了数值实验，并撰写了第 3 至第 5 部分。所有作者均审阅了手稿。

7.4 资助

CD 感谢英国工程与自然科学研究委员会（EPSRC）提供支持（EP/S026045/1）。MJE 感谢英国工程与自然科学研究委员会（EPSRC）提供支持（EP/S026045/1、EP/T026693/1、EP/V026259/1）以及 Leverhulme Trust 提供支持（ECF-2019-478）。CBS 感谢 Philip Leverhulme Prize、英国皇家学会 Wolfson Fellowship、英国工程与自然科学研究委员会（EPSRC）高级职业研究员奖学金 EP/V029428/1、英国工程与自然科学研究委员会（EPSRC）拨款 EP/S026045/1 和 EP/T003553/1、EP/N014588/1、EP/T017961/1、Wellcome创新者奖 215733/Z/19/Z 和 221633/Z/20/Z、欧盟地平线2020研究与创新计划 תרגסמב玛丽·斯克沃多夫斯卡-居里资助协议 No. 777826 NoMADS、Cantab Capital信息数学研究所及艾伦·图灵研究所提供的支持。

7.5 数据和材料的可用性

相关算法实现以及实验中使用的图像数据将在网站 https://junqitang.com 上提供。对于模体图像示例，我们使用 [8], 实验部分中的图像；对于肺部 CT 图像示例，我们使用梅奥诊所数据集 [21] 中的一张图像，该数据集可公开获取。

参考文献

[1] Ahmet Alacaoglu、Olivier Fercoq 和 Volkan Cevher。随机原始-对偶混合梯度的收敛性。SIAM 优化期刊，32(2):1288–1318，2022。[2] Heinz H. Bauschke 和 Patrick L. Combettes。希尔伯特空间中的凸分析与单调算子理论，第 408 卷。施普林格，2011。[3] Silvia Bonettini、Alessandro Benfenati 和 Valeria Ruggiero。epsilon-次梯度方法的缩放技术。SIAM 优化期刊，26(3):1741–1772，2016。[4] Silvia Bonettini、Federica Porta、Valeria Ruggiero 和 Luca Zanni。成像中前向–后向方法的变度量技术。计算与应用数学期刊，385:113192，2021。[5] Silvia Bonettini、Marco Prato 和 Simone Rebegoldi。一种基于线搜索的块坐标变度量近端梯度方法。计算优化与应用，71(1):5–52，2018。[6]Silvia Bonettini、Simone Rebegoldi 和 Valeria Ruggiero。用于非精确前向–后向算法的惯性变度量技术。SIAM科学计算期刊，40(5):A3180–A3210，2018。

[7] Silvia Bonettini 和 Valeria Ruggiero。用于全变分图像复原的原始–对偶混合梯度算法的收敛性。数学成像与视觉期刊，44(3):236–253，2012。[8] Antonin Chambolle、Matthias J Ehrhardt、Peter Richt´arik 和 Carola-Bibiane Sch¨onlieb。具有任意采样的随机原始-对偶混合梯度算法及其成像应用。SIAM 优化期刊，28(4):2783–2808，2018。[9] Antonin Chambolle 和 Thomas Pock。一种用于凸问题的一阶原始-对偶算法及其在成像中的应用。数学成像与视觉期刊，40(1):120–145，2011。[10] Patrick L. Combettes 和 Jean-Christophe Pesquet。具有随机扫描的随机准 Fej´er 块坐标不动点迭代。SIAM 优化期刊，25(2):1221–1248，2015。[11] Patrick L Combettes 和 Bang C V˜u。变度量准 Fej´er 单调性。非线性分析：理论、方法与应用，78:17–31，2013。[12] Claire Delplancke、Mark Gurnell、Jonas Latz、Pawel J Markiewicz、Carola-Bibiane Sch¨onlieb 和 Matthias J Ehrhardt。改进一种用于正则化 PET 图像重建的随机算法。载于《2020 年 IEEE 核科学研讨会与医学成像会议》（NSS/MIC），第 1–3 页。IEEE，2020。[13] Matthias J Ehrhardt、Pawel Markiewicz 和 Carola-Bibiane Sch¨onlieb。通过随机化和预条件，利用非光滑先验加速 PET 重建。Physics in Medicine & Biology，64(22):225019，2019。[14] Tom Goldstein、Min Li 和 Xiaoming Yuan。用于统计学习和图像处理的自适应原始-对偶分裂方法。Advances in Neural Information Processing Systems，(28):2089–2097，2015。[15] Tom Goldstein、Min Li、Xiaoming Yuan、Ernie Esser 和 Richard Baraniuk。用于鞍点问题的自适应原始-对偶混合梯度方法。arXiv 预印本 arXiv:1305.0546，2013。[16] Eric B. Guti´errez、Claire Delplancke 和 Matthias J. Ehrhardt。随机化原始-对偶算法的收敛性与抽样及其在并行 MRI 重建中的应用。arXiv 预印本 arXiv:2207.12291，2022。[17] Bingsheng He 和 Xiaoming Yuan。用于全变分图像复原的原始-对偶算法收敛性分析。技术报告，Citeseer，2010。[18] Yura Malitsky。用于变分不等式的黄金比例算法。Mathematical Programming，184(1):383–410，2020。[19] Yura Malitsky 和 Konstantin Mishchenko。无需下降的自适应梯度下降。载于 Hal Daum´e III 和 Aarti Singh 编，《第 37 届国际机器学习会议论文集》，机器学习研究论文集第 119 卷，第 6702–6712 页。PMLR，2020 年 7 月 13–18 日。[20] Yura Malitsky 和 Thomas Pock。一种带线搜索的一阶原始-对偶算法。SIAM 优化期刊，28(1):411–432，2018。

[21] Cynthia McCollough. Tu-fg-207a-04：低剂量 CT 重大挑战概述。Medical Physics，43(6Part35)：3759–3760，2016。[22]Evangelos Papoutsellis、Evelina Ametova、Claire Delplancke、Gemma Fardell、Jakob S Jørgensen、Edoardo Pasca、Martin Turner、Ryan Warr、William RB Lionheart 和 Philip J Withers。Core Imaging Library—第二部分：动态和能谱断层成像的多通道重建。Philosophical Transactions of the Royal Society A，379(2204)：20200193，2021。[23] Herbert Robbins 和 David Siegmund。关于非负几乎超鞅的收敛定理及其若干应用。载于《Optimizing Methods in Statistics》，第 233–257 页。爱思唯尔，1971。[24] Georg Schramm 和 Martin Holler。列表模式下稀疏泊松数据的快速、内存高效重建：采用非光滑先验，并应用于飞行时间正电子发射断层成像。Physics in Medicine & Biology，2022。[25] Maria-Luiza Vladarean、Yura Malitsky 和 Volkan Cevher。一种适应局部光滑性的一阶原始-对偶方法。Advances in Neural Information Processing Systems，34：6171–6182，2021。[26] Tatsuya Yokota 和 Hidekata Hontani。一种用于调整原始-对偶混合梯度法步长参数的高效方法，并将其应用于全变分正则化。载于 2017 年亚太信号与信息处理协会年度峰会暨会议（APSIPA ASC），第 973–979 页。IEEE，2017。[27] Lena Zdun 和 Christina Brandt。通过随机优化和数据驱动分裂，结合非光滑先验实现快速 MPI 重建。Physics in Medicine & Biology，66(17)：175004，2021。