

大规模可分凸优化问题的自适应步随机原始对偶算法*

周晓艳, 罗洪林

(重庆师范大学 数学科学学院, 重庆 401331)

摘要:针对大规模可分凸优化问题,提出自适应步随机原始对偶算法。首先将此问题等价地转换为一个对偶变量可分离的鞍点问题,接着随机选择鞍点问题的对偶变量更新,并按照一定规则自适应选取原始对偶步,证明得到该算法迭代点的遍历序列按期望以 $O(1/N)$ 的速率收敛。数值实验的结果表明该算法可以有效解决正电子发射断层成像问题。

关键词:大规模可分凸优化问题;随机优化;原始对偶算法;自适应步长

中图分类号:O224

文献标志码:A

文章编号:1672-6693(2024)03-0009-11

图像处理^[1-2]、机器学习^[3]、统计^[4]等领域中的很多问题都可以抽象成如下凸优化问题:

$$\min_{x \in X} \Phi(x) = f(Ax) + g(x),$$

其中: f 和 g 是凸函数, A 是有界线性算子。该问题已有众多成熟的算法来解决,如交替方向乘子法(ADMM)^[5]、原始对偶混合梯度算法(PDHG)^{[1][6-7]}。特别地,当 A 是恒等算子时,还可以选择向前-向后算法^[8]、Douglas-Rachford 分裂算法^[9]。

随着机器学习等学科的发展,问题的难点主要体现在规模大,计算量大等方面。因此,如何高效地解决如下大规模可分凸优化问题具有重要的研究意义:

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

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。近些年来,解决该问题的随机算法成为研究热点,如镜像下降随机近似算法^[10]、加速随机近似算法^[11]解决了问题(1)中 f_i 是非光滑的情况,并且这 2 种算法都以 $O(1/\sqrt{N})$ 的速率收敛。当 f_i 是光滑凸函数时,基于增量平均梯度方法^[12]和随机平均梯度方法^[13]解决问题(1)的收敛速率可以提升到 $O(1/N)$ 。又如当 g 是强凸函数时,文献^[14]针对问题(1)的对偶问题,提出了随机对偶坐标上升法。

Chambolle 等人在文献[1]中提出了原始对偶混合梯度算法(PDHG),该算法只涉及邻近算子求解和矩阵向量的乘法等简单的运算,当邻近算子具有显式表达式时,与其他一阶或二阶算法相比,该算法迭代求解的过程相对简单。在大规模问题(1)中,PDHG 每迭代一次,需要求解 $n+1$ 个邻近算子,当 n 的值较大时,对计算速度和内存空间都是极大的考验。

在原始对偶算法的基础上,越来越多的学者针对大规模可分凸优化问题提出了随机原始对偶算法:如 Dang^[15]、Zhang^[16]和 Zhu^[17]在每次迭代中随机选取部分对偶变量进行更新,有效减少了邻近算子的计算次数,这里随机抽样的方式为统一概率抽样。Fercoq 等人^[18]提出了一种随机原始对偶算法,并证明了此算法在非强凸假设下的次线性收敛速率,该算法利用先前所有迭代的平均信息来获得更大的步长,可是这需要更大的内存空间来实现。Chambolle 等人^[19]证明了在任意抽样方式下随机原始对偶算法的收敛性,其中保守的固定原始对偶步长的设置,在实际问题中可能影响收敛速度,特别是对未归一化的数据。

* 收稿日期:2023-07-25 修回日期:2024-01-30 网络出版时间:2024-06-07T14:07

资助项目:国家自然科学基金——重大项目(No. 11991024),自由申请项目(No. 11771064);重庆市创新领军人才团队项目(No. CQYC20210309536);重庆市高校创新研究群体项目(No. 20A110029);重庆市自然科学基金(No. cstc2021jcyj-msx300)

第一作者简介:周晓艳,女,研究方向为最优化理论与算法,E-mail:2469650052@qq.com;通信作者:罗洪林,男,教授,E-mail:luohonglin@cqnu.edu.cn

网络出版地址:https://link.cnki.net/urlid/50.1165.N.20240607.0921.006

在经典的原始对偶算法基础上,文献[7,20]从不同角度提出加速策略,如 Chang 等人^[7]将黄金比率融入到原始变量的凸组合中,使得该算法在实际应用中具有良好的性能。Chang^[7]和 Malitsky^[20]在算法结构中加入了线性搜索步长的过程,因此可以获取更大的步长值,从而提高收敛速度,并保证全局收敛性。He 等人^[21]对原始变量和对偶变量都执行凸组合的设置,这样 PDHG 就是完全对称的结构。

本文针对大规模可分凸优化问题,提出了自适应步随机原始对偶算法,并证明了遍历序列的原始对偶间隙按期期望以 $O(1/N)$ 的速率收敛。该算法在每次迭代中随机选取对偶变量空间的一个子集进行更新,相比文献[16]中的随机抽样的方式不局限于统一概率抽样,可以是任意抽样。在文献[19]基础之上,在没有强凸的假设条件下,本文所提出的算法依旧采用可变步长的设置,加快了收敛速率。此外,算法结构中的子循环使得原始对偶步长能够自适应地趋于适合实际问题的步长,避免了预先计算算子范数的困难。最后,本文将该算法应用于正电子发射断层成像(PET)问题,数值实验结果验证了该算法的有效性。

本文首先简要介绍了目标问题的相关定义和假设,再给出了自适应步随机原始对偶算法的框架和收敛性分析,将该算法应用于实际问题中,最后简要总结了全文的主要内容,并提出了未来研究的方向。

1 预备知识

首先给出自适应步随机原始对偶算法涉及的定义、性质和一些假设。

定义 1^[22] 若函数 $f: E \rightarrow (-\infty, \infty]$, 则对任意的 $x \in E$, 邻近算子 $\text{prox}_f(x)$ 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, $\text{prox}_f^*(x) = \arg\min_{u \in E} \left\{ \frac{1}{2} \|u - x\|_{\tau^{-1}}^2 + f(u) \right\}$, 其中 $\|x\|_{\tau^{-1}}^2 = \langle \tau^{-1}x, x \rangle$ 。

性质 1^[22] 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^[22] 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:

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

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

设 $X, Y_i, i=1, \dots, n$ 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, Y 空间上由内积诱导的范数定义为 $\|y\| = \sqrt{\sum_{i=1}^n \|y_i\|^2}$ 。此外, 记 $W := X \times Y$ 为原始对偶变量空间。

设 $A: X \rightarrow Y$ 的一个有界线性算子。根据积空间 $Y = \prod_{i=1}^n Y_i$ 的特性, 有 $(Ax)_i = A_i x$, 其中 $A_i: X \rightarrow Y_i$ 是线性算子; 定义 A 的伴随算子 $A^*: Y \rightarrow X$, $A^* y = \sum_{i=1}^n A_i^* y_i$ 。函数 $f: Y \rightarrow \mathbf{R} \cup \{+\infty\}$ 和 $g: X \rightarrow \mathbf{R} \cup \{+\infty\}$ 是凸函数, 且 f 是下半连续的。特别地, 假设 f 是可分离的, 即 $f(y) = \sum_{i=1}^n f_i(y_i)$ 。

2 自适应步随机原始对偶算法及收敛性分析

首先, 本文不直接解决问题(1), 而是利用 Fenchel 共轭将它转换为一个鞍点问题。当 f 是下半连续的真凸函数时, 即 f 是闭真凸函数, 则:

$$f = f^{**} = \sup_{y \in Y} \left\{ \sum_{i=1}^n \langle A_i x, y_i \rangle - f_i^*(y_i) \right\},$$

于是寻求问题(1)的解 x 等价于寻求以下鞍点问题的原始变量 x :

$$\minsup_{x \in X, y \in Y} \left\{ \psi(x, y) := g(x) + \sum_{i=1}^n \langle A_i x, y_i \rangle - f_i^*(y_i) \right\}. \quad (2)$$

假设问题(2)的鞍点 $w^\# = (x^\#, y^\#) = (x^\#, y_1^\#, y_2^\#, \dots, y_n^\#)$ 存在,其一阶必要性条件为:

$$-A^* y^\# \in \partial g(x^\#), A_i x^\# \in \partial f_i^*(y_i^\#), i=1, 2, \dots, n.$$

根据鞍点定义原始间隙和对偶间隙:

$$P_{x^\#, y^\#}(x) = g(x) - g(x^\#) + \langle A^* y^\#, x - x^\# \rangle \geq 0, \forall x \in X, \quad (3)$$

$$D_{x^\#, y^\#}(y) = f^*(y) - f^*(y^\#) - \langle A x^\#, y - y^\# \rangle \geq 0, \forall y \in Y, \quad (4)$$

则原始对偶间隙可以表示为: $G_{x^\#, y^\#}(x, y) = P_{x^\#, y^\#}(x) + D_{x^\#, y^\#}(y) \geq 0$. 当 $(x^\#, y^\#)$ 确定时,显然 $g(x)$ 和 $f^*(y)$ 是凸函数,又因为 $g(x^\#)$ 和 $f^*(y^\#)$ 是定值, $\langle A^* y^\#, x - x^\# \rangle$ 和 $\langle A x^\#, y - y^\# \rangle$ 是线性的,所以 $P_{x^\#, y^\#}(\cdot)$, $D_{x^\#, y^\#}(\cdot)$, $G_{x^\#, y^\#}(\cdot, \cdot)$ 是凸函数.

解决问题(2)常用 PDHG^[1]方法,即:

$$\begin{aligned} y^{(k+1)} &= \text{prox}_{f^*}^\sigma(y^{(k)} + \sigma A \bar{x}^{(k)}), \\ x^{(k+1)} &= \text{prox}_g^\tau(x^{(k)} - \tau A^* y^{(k+1)}), \\ \bar{x}^{(k+1)} &= x^{(k+1)} + \theta(x^{(k+1)} - x^{(k)}). \end{aligned}$$

当 f 和 g 为凸函数时,文献[1]要求原始对偶步长满足条件 $\sigma\tau \|A\|^2 < 1$,并证明了该算法的收敛性.在此基础上,本文针对鞍点问题(2)的大规模、可分离等特点,将 PDHG 拓展到了随机的情况,并融入了自适应步长,从而加快收敛速率.

针对问题(2)提出如下自适应步随机原始对偶算法(ASPDHG).

算法 1 选取初始点 $x^{(0)} \in X, y^{(1)} \in Y$ 以及初始步长 $\tau_{(0)} > 0$,令 $k=0$;给定迭代次数 $N \in \mathbb{N}^+$ 和参数 $\mu \in (0, 1), \delta \in (0, 1)^+, \beta > 0$.

步骤 1,若 $k=N$,算法终止,否则计算 $x^{(k)} = \text{prox}_g^{\tau_{(k-1)}}(x^{(k-1)} - \tau_{(k-1)} A^* y^{(k)})$;

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

步骤 3,计算 $\theta_{(k)} = \frac{\tau_{(k)}}{\tau_{(k-1)}}$, $\bar{x}^{(k)} = x^{(k)} + \theta_{(k)}(x^{(k)} - x^{(k-1)})$, $y_i^{(k+1)} = \begin{cases} \text{prox}_{f_i^*}^{\beta\tau_{(k)}}(y^{(k)} + \beta\tau_{(k)} A_i \bar{x}^{(k)}), & i \in S_{(k)} \\ y_i^{(k)}, & i \notin S_{(k)} \end{cases}$;

步骤 4,判断 $\sqrt{\beta}\tau_{(k)} \|A^* y^{(k+1)} - A^* y^{(k)}\| \leq \delta \|y^{(k+1)} - y^{(k)}\|$ 是否成立,若成立,置 $k=k+1$,转步骤 1;否则令 $\tau_{(k)} := \tau_{(k)} \mu$,并转步骤 3.

注 1 在选取 $S_{(k)}$ 时,与文献[15]的算法相比,该算法的随机抽样可以是任意的,不必是统一概率抽样,但必须保证每一个对偶变量指标 i 被选取概率 $P_i > 0$. 随机子集选择方式虽然更加灵活,但某些不适合的抽样方式可能会导致一些数据点被频繁更新,而另一些数据点则很少更新,从而影响实际应用效果.

注 2 该算法将线性搜索步长结构与针对大规模问题的随机算法相结合,步骤 4 的判断条件,使得原始对偶步长能够自适应地趋于适合实际问题的步长,避免了预先计算算子范数的困难.与文献[20]不同的是,该算法

步骤 4 中步长的选择是一个单调递减并趋于稳定的过程,而文献[20]的步长 $\left(\tau_{(k)} \in \left[\tau_{(k-1)} \sqrt{\frac{\beta_{(k-1)}}{\beta_{(k)}}}, \tau_{(k-1)} \sqrt{\frac{\beta_{(k-1)}}{\beta_{(k)}}} (1 + \theta_{(k-1)}) \right] \right)$ 选取策略在理论上并不适用.

注 3 每进行一次原始变量邻近算子求值后,搜索步长的子循环不会是一个无限的过程,在经过有限次迭代之后,一定会满足步骤 4 的终止条件,在引理 1 中详细证明了这一点.在收敛性证明的过程中,步骤 4 中的终止条件所起的作用与文献[1]中的步长条件 $\sigma\tau \|A\|^2 < 1$ 所起的作用一致.

注 4 在以下算法中,参数 β 等价于文献[1,19]中的对偶步长与原始步长之间的比值 $\frac{\sigma}{\tau}$;在实际应用时, δ

的取值可以尽量靠近 1; 在步长搜索子循环的过程中, 步长值通过 $\tau_{(k)} := \tau_{(k)} \mu$ 来下降, 逐渐趋于适应算法和实际问题的步长; 在收敛性分析中已经证明遍历序列的原始对偶间隙按期望以 $O(1/N)$ 的速度收敛, 因此可以根据目标精度来确定迭代次数 N 的取值。

针对上述算法的特殊结构, 可以得出以下结论。

引理 1 在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, 一定会满足步骤 4 中的终止条件;
- 2) 对任意的 $k \geq 0$, 存在 τ , 使得 $\tau_{(k)} \geq \tau$ 。

证明 1) 每一次子循环迭代执行 $\tau_{(k)} \cdot \mu (\mu < 1)$, 则存在 k' , 使得 $\tau_{(k')} \leq \frac{\delta}{\sqrt{\beta} \|A^*\|}$, 则:

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

线性搜索步长迭代终止。

- 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 $\tau_{(k)} = \tau_{(k-1)} \mu^j, j \in \mathbf{N}$ 。当 $j=0$ 时, $\tau_{(k)} \geq \tau_{(k-1)} > \frac{\delta \mu}{\sqrt{\beta} \|A^*\|}$ 成立; 当 $j>0$ 时, 则有 $\tau'_{(k)} = \tau_{(k-1)} \mu^{j-1}$ 不满足步骤 4 中的终止条件, 即:

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

因此:

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

自适应步随机原始对偶算法的收敛性分析是由确定到随机的过程, 下面引理考虑了在第 k 次迭代中对偶变量完全更新的情况。

引理 2 在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, 即对任意的 $i=1, 2, \dots, n$, 有 $\hat{y}_i^{(k+1)} = \text{prox}_{f_i^*}^{\beta \tau_{(k)}}(y^{(k)} + \beta \tau_{(k)} A_i \bar{x}^{(k)})$, 则:

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

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

证明 设 $(x^\#, y^\#)$ 是问题(2)的鞍点, 当对偶变量全部都更新时, 由性质 1 中条件(3)可得:

$$\tau_{(k)} (g(x^{(k+1)}) - g(x^\#)) \leq [x^{(k+1)} - x^{(k)} + \tau_{(k)} A^* \hat{y}^{(k+1)}, x^\# - x^{(k+1)}], \quad (5)$$

$$\tau_{(k)} (f_i^*(\hat{y}_i^{(k+1)}) - f_i^*(y_i^\#)) \leq \langle \frac{1}{\beta} (\hat{y}_i^{(k+1)} - y_i^{(k)}) - \tau_{(k)} A_i \bar{x}^{(k)}, y_i^\# - \hat{y}_i^{(k+1)} \rangle, i=1, 2, \dots, n, \quad (6)$$

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

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

因此, 取式(6)中 $i=1, 2, \dots, n$ 相加可得:

$$\tau_{(k)} (f^*(\hat{y}^{(k+1)}) - f^*(y^\#)) \leq \langle \frac{1}{\beta} (\hat{y}^{(k+1)} - y^{(k)}) - \tau_{(k)} A \bar{x}^{(k)}, y^\# - \hat{y}^{(k+1)} \rangle. \quad (7)$$

又因为 $x^{(k)} = \text{prox}_g^{\tau_{(k-1)}}(x^{(k-1)} - \tau_{(k-1)} A^* y^{(k)})$, 再次由性质 1 中(3)可得:

$$\tau_{(k-1)} (g(x^{(k)}) - g(x)) \leq \langle x^{(k)} - x^{(k-1)} + \tau_{(k-1)} A^* y^{(k)}, x - x^{(k)} \rangle, \forall x \in X, \quad (8)$$

分别取 $x = x^{(k+1)}$, $x = x^{(k-1)}$ 代入式(8),得到:

$$\tau_{(k-1)}(g(x^{(k)}) - g(x^{(k+1)})) \leq \langle x^{(k)} - x^{(k-1)} + \tau_{(k-1)}A^*y^{(k)}, x^{(k+1)} - x^{(k)} \rangle, \quad (9)$$

$$\tau_{(k-1)}(g(x^{(k)}) - g(x^{(k-1)})) \leq \langle x^{(k)} - x^{(k-1)} + \tau_{(k-1)}A^*y^{(k)}, x^{(k-1)} - x^{(k)} \rangle, \quad (10)$$

式(9)乘以 $\theta_{(k)} = \frac{\tau_{(k)}}{\tau_{(k-1)}}$, 式(10)乘以 $\theta_{(k)}^2$, 再相加可得:

$$\begin{aligned} & \langle \theta_{(k)}(x^{(k)} - x^{(k-1)}) + \tau_{(k)}A^*y^{(k)}, x^{(k+1)} - x^{(k)} \rangle + \langle \theta_{(k)}(x^{(k)} - x^{(k-1)}) + \tau_{(k)}A^*y^{(k)}, \theta_{(k)}(x^{(k-1)} - x^{(k)}) \rangle \geq \\ & \tau_{(k)}(g(x^{(k)}) - g(x^{(k+1)})) + \tau_{(k)}\theta_{(k)}(g(x^{(k)}) - g(x^{(k-1)})), \end{aligned} \quad (11)$$

将 $\bar{x}^{(k)} = x^{(k)} + \theta_{(k)}(x^{(k)} - x^{(k-1)})$ 代入式(11), 则:

$$\langle \bar{x}^{(k)} - x^{(k)} + \tau_{(k)}A^*y^{(k)}, x^{(k+1)} - \bar{x}^{(k)} \rangle \geq \tau_{(k)}((1 + \theta_{(k)})g(x^{(k)}) - g(x^{(k+1)}) - \theta_{(k)}g(x^{(k-1)})), \quad (12)$$

又因为:

$$\tau_{(k)}\langle A^*\hat{y}^{(k+1)} - A^*y^\#, \bar{x}^{(k)} - x^\# \rangle = \tau_{(k)}\langle A\bar{x}^{(k)} - Ax^\#, \hat{y}^{(k+1)} - y^\# \rangle, \quad (13)$$

将式(5)、(7)、(12)、(13)相加可得:

$$\begin{aligned} & \langle x^{(k+1)} - x^{(k)}, x^\# - x^{(k+1)} \rangle + \frac{1}{\beta} \langle \hat{y}^{(k+1)} - y^{(k)}, y^\# - \hat{y}^{(k+1)} \rangle + \langle \bar{x}^{(k)} - x^{(k)}, x^{(k+1)} - \bar{x}^{(k)} \rangle + \\ & \tau_{(k)}\langle A^*\hat{y}^{(k+1)} - A^*y^{(k)}, \bar{x}^{(k)} - x^{(k+1)} \rangle - \tau_{(k)}\langle A^*y^\#, \bar{x}^{(k)} - x^\# \rangle + \tau_{(k)}\langle Ax^\#, \hat{y}^{(k+1)} - y^\# \rangle \geq \\ & \tau_{(k)}(f^*(\hat{y}^{(k+1)}) - f^*(y^\#) + (1 + \theta_{(k)})g(x^{(k)}) - \theta_{(k)}g(x^{(k-1)}) - g(x^\#)), \end{aligned} \quad (14)$$

由式(3)得到:

$$\begin{aligned} & (1 + \theta_{(k)})g(x^{(k)}) - \theta_{(k)}g(x^{(k-1)}) - g(x^\#) + \langle A^*y^\#, \bar{x}^{(k)} - x^\# \rangle = \\ & (1 + \theta_{(k)})(g(x^{(k)}) - g(x^\#) + \langle A^*y^\#, x^{(k)} - x^\# \rangle) - \theta_{(k)}(g(x^{(k-1)}) - g(x^\#) + \langle A^*y^\#, x^{(k-1)} - x^\# \rangle) = \\ & (1 + \theta_{(k)})P_{x^\#, y^\#}(x^{(k)}) - \theta_{(k)}P_{x^\#, y^\#}(x^{(k-1)}), \end{aligned} \quad (15)$$

结合式(4)、(14)和(15)有:

$$\begin{aligned} & \langle x^{(k+1)} - x^{(k)}, x^\# - x^{(k+1)} \rangle + \frac{1}{\beta} \langle \hat{y}^{(k+1)} - y^{(k)}, y^\# - \hat{y}^{(k+1)} \rangle + \\ & \langle \bar{x}^{(k)} - x^{(k)}, x^{(k+1)} - \bar{x}^{(k)} \rangle + \tau_{(k)}\langle A^*\hat{y}^{(k+1)} - A^*y^{(k)}, \bar{x}^{(k)} - x^{(k+1)} \rangle \geq \\ & \tau_{(k)}((1 + \theta_{(k)})P_{x^\#, y^\#}(x^{(k)}) - \theta_{(k)}P_{x^\#, y^\#}(x^{(k-1)}) + D_{x^\#, y^\#}(\hat{y}^{(k+1)})), \end{aligned} \quad (16)$$

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

将 $2\langle a - b, c - a \rangle = \|b - c\|^2 - \|a - b\|^2 - \|a - c\|^2$ 代入式(16), 则:

$$\begin{aligned} & \frac{1}{2}(\|x^{(k)} - x^\#\|^2 - \|x^{(k+1)} - x^\#\|^2 - \|x^{(k+1)} - x^{(k)}\|^2) + \\ & \frac{1}{2\beta}(\|y^{(k)} - y^\#\|^2 - \|\hat{y}^{(k+1)} - y^\#\|^2 - \|\hat{y}^{(k+1)} - y^{(k)}\|^2) + \\ & \frac{1}{2}(\|x^{(k+1)} - x^{(k)}\|^2 - \|\bar{x}^{(k)} - x^{(k)}\|^2 - \|x^{(k+1)} - \bar{x}^{(k)}\|^2) + \\ & \tau_{(k)}\langle A^*\hat{y}^{(k+1)} - A^*y^{(k)}, \bar{x}^{(k)} - x^{(k+1)} \rangle \geq \epsilon_{(k)}, \end{aligned} \quad (17)$$

根据 Cauchy-Schwarz 不等式和步骤 4 中的终止条件可得:

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

将上式代入式(17)得到:

$$\begin{aligned} & \frac{1}{2}(\|x^{(k)} - x^\#\|^2 - \|x^{(k+1)} - x^\#\|^2) + \frac{1}{2\beta}(\|y^{(k)} - y^\#\|^2 - \|\hat{y}^{(k+1)} - y^\#\|^2) - \\ & \frac{1}{2}\|\bar{x}^{(k)} - x^{(k+1)}\|^2 - \frac{1 - \delta^2}{2\beta}\|\hat{y}^{(k+1)} - y^{(k)}\|^2 \geq \epsilon_{(k)}. \end{aligned} \quad (18)$$

因此引理 2 成立。

证毕

下面给出本文的重要结论。

定理 1 设 $(x^{(k)}, y^{(k)})$ 是由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)的鞍点, 则遍历序列 $(X_{(N)}, Y_{(N)})$ 的原始对偶间隙 $P(X_{(N)}) + D(Y_{(N)})$ 按期望以 $O(1/N)$ 的速率收敛到 0, 且满足:

$$E\{P(X_{(N)}) + D(Y_{(N)})\} \leq \frac{1}{S_{(N)}} (\Delta_{(1)} + \tau_{(1)} \theta_{(1)} P(x^{(0)}) + \tau_{(1)} D^p(y^{(1)})) \leq \frac{1}{N} \cdot \left(\frac{1}{\tau} (\Delta_{(1)} + \tau_{(1)} \theta_{(1)} P(x^{(0)}) + \tau_{(1)} D^p(y^{(1)})) \right),$$

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

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

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 的概率更新有, $E^{(k+1)}\varphi(y_i^{(k+1)}) = p_i\varphi(\hat{y}_i^{(k+1)}) + (1-p_i)\varphi(y_i^{(k)})$, 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

$$\begin{aligned} \|\hat{y}^{(k+1)} - y^{\#}\|^2 &= E^{(k+1)}\|y^{(k+1)} - y^{\#}\|_Q^2 - \|y^{(k)} - y^{\#}\|_Q^2, \\ \|\hat{y}^{(k+1)} - y^{(k)}\|^2 &= E^{(k+1)}\|y^{(k+1)} - y^{(k)}\|_Q^2, \\ D(\hat{y}^{(k+1)}) &= E^{(k+1)}D^p(y^{(k+1)}) - D^p(y^{(k)}) + E^{(k+1)}D(y^{(k+1)}), \end{aligned} \quad (19)$$

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

将式(19)代入到式(18)得到:

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

两边同时对 $y^{(k)}$ 求期望, 得到:

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

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:

$$\begin{aligned} &E\left\{ \frac{1}{2} \|x^{(k)} - x^{\#}\|^2 + \frac{1}{2\beta} \|y^{(k)} - y^{\#}\|_Q^2 + \tau_{(k)}\theta_{(k)}P(x^{(k-1)}) - \tau_{(k)}(1 + \theta_{(k)})P(x^{(k)}) + \tau_{(k)}D^p(y^{(k)}) \right\} \geq \\ &E\left\{ \frac{1}{2} \|x^{(k+1)} - x^{\#}\|^2 + \frac{1}{2\beta} \|y^{(k+1)} - y^{\#}\|_Q^2 + \tau_{(k)}(D^p(y^{(k+1)}) + D(y^{(k+1)})) \right\}, \end{aligned} \quad (20)$$

令 $\Delta_{(k)} = E \left\{ \frac{1}{2} \|x^{(k)} - x^\# \|^2 + \frac{1}{2\beta} \|y^{(k)} - y^\# \|^2 \right\}$, 则式(20)可以表示为:

$$\Delta_{(k)} \geq \Delta_{(k+1)} + E \{ \tau_{(k)} ((1 + \theta_{(k)}) P(x^{(k)}) - \theta_{(k)} P(x^{(k-1)}) - D^p(y^{(k)}) + D^p(y^{(k+1)}) + D(y^{(k+1)})) \}, \quad (21)$$

取式(21)中的 $k=1, 2, \dots, N$, 相加得到:

$$\Delta_{(1)} - \Delta_{(N+1)} \geq \sum_{k=1}^N E \{ \tau_{(k)} ((1 + \theta_{(k)}) P(x^{(k)}) - \theta_{(k)} P(x^{(k-1)})) + \tau_{(k)} (D^p(y^{(k+1)}) + D(y^{(k+1)}) - D^p(y^{(k)})) \}. \quad (22)$$

由 P 的凸性可得:

$$\begin{aligned} & E \left\{ \sum_{k=1}^N [\tau_{(k)} (1 + \theta_{(k)}) P(x^{(k)}) - \tau_{(k)} \theta_{(k)} P(x^{(k-1)})] \right\} = \\ & E \left\{ \sum_{k=1}^{N-1} [\tau_{(k)} (1 + \theta_{(k)}) - \tau_{(k+1)} \theta_{(k+1)}] P(x^{(k)}) + \tau_{(N)} (1 + \theta_{(N)}) P(x^{(N)}) - \tau_{(1)} \theta_{(1)} P(x^{(0)}) \right\} \geq \\ & E \left\{ (\tau_{(1)} \theta_{(1)} + S_{(N)}) P \left(\frac{\tau_{(N)} (1 + \theta_{(N)}) x^{(N)} + \sum_{k=1}^{N-1} (\tau_{(k)} (1 + \theta_{(k)}) - \tau_{(k+1)} \theta_{(k+1)}) x^{(k)}}{\tau_{(1)} \theta_{(1)} + S_{(N)}} \right) - \tau_{(1)} \theta_{(1)} P(x^{(0)}) \right\} \geq \\ & E \{ S_N P(X_N) \} - \tau_1 \theta_1 P(x^0), \end{aligned} \quad (23)$$

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

同理, 由 D 的凸性可得到:

$$E \left\{ \sum_{k=1}^N \tau_{(k)} D(y^{(k+1)}) \right\} \geq E \left\{ \left(\sum_{k=1}^N \tau_{(k)} \right) \cdot D \left(\frac{\sum_{k=1}^N \tau_{(k)} y^{(k+1)}}{\sum_{k=1}^N \tau_{(k)}} \right) \right\} = E \{ S_{(N)} \cdot D(Y_{(N)}) \}, \quad (24)$$

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

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

$$\begin{aligned} & \sum_{k=1}^N E \{ \tau_{(k)} (D^p(y^{(k+1)}) - D^p(y^{(k)})) \} = \\ & E \left\{ \tau_{(N)} D^p(y^{(N+1)}) - \tau_{(1)} D^p(y^{(1)}) + \sum_{k=1}^{N-1} (\tau_{(k)} - \tau_{(k+1)}) D^p(y^{(k+1)}) \right\} \geq \\ & E \left\{ (\tau_{(1)} - \tau_{(N)}) D^p \left(\frac{\sum_{k=1}^{N-1} (\tau_{(k)} - \tau_{(k+1)}) y^{(k+1)}}{(\tau_{(1)} - \tau_{(N)})} \right) + \tau_{(N)} D^p(y^{(N+1)}) - \tau_{(1)} D^p(y^{(1)}) \right\} \geq \\ & E \{ \tau_{(N)} D^p(y^{(N+1)}) \} - \tau_{(1)} D^p(y^{(1)}) \geq -\tau_{(1)} D^p(y^{(1)}). \end{aligned} \quad (25)$$

将式(23)、(24)、(25)代入式(22)可得:

$$\begin{aligned} E \{ P(X_{(N)}) + D(Y_{(N)}) \} & \leq \frac{1}{S_{(N)}} (\Delta_{(1)} - \Delta_{(N+1)} + \tau_{(1)} \theta_{(1)} P(x^{(0)}) + \tau_{(1)} D^p(y^{(1)})) \leq \\ & \frac{1}{S_{(N)}} (\Delta_{(1)} + \tau_{(1)} \theta_{(1)} P(x^{(0)}) + \tau_{(1)} D^p(y^{(1)})), \end{aligned} \quad (26)$$

根据引理1中的结论(2), 对任意的 $k \geq 0$, 存在 τ , 使得 $\tau_{(k)} \geq \tau$, 得到:

$$S_{(N)} = \sum_{k=1}^N \tau_{(k)} \geq N \cdot \tau, \quad (27)$$

将式(27)代入式(26)有:

$$E\{P(X_{(N)})+D(Y_{(N)})\} \leq \frac{1}{S_{(N)}}(\Delta_{(1)}+\tau_{(1)}\theta_{(1)}P(x^{(0)})+\tau_{(1)}D^p(y^{(1)})) \leq \frac{1}{N} \cdot \left(\frac{1}{\tau}(\Delta_{(1)}+\tau_{(1)}\theta_{(1)}P(x^{(0)})+\tau_{(1)}D^p(y^{(1)}))\right).$$

即遍历序列的原始对偶间隙 $P(X_{(N)})+D(Y_{(N)})$ 按期望以 $O(1/N)$ 的速率收敛。

证毕

注 5 定理 1 的结果表明遍历序列的原始对偶间隙 $P(X_{(N)})+D(Y_{(N)})$ 按期望以 $O(1/N)$ 的速率收敛, 尽管遍历序列的定义方式与文献[2, 23]有所不同, 但他们的收敛速率一致。并且, 当随机选取退化完全选取时, 定理 2 的理论结果依然成立, 也就是说, 确定性和随机性两种情况下的结论都可以通过一个证明来解决。此外, 与文献[15]不同的是, 定理 1 中关于随机的证明不局限于统一概率抽样, 只需保证每一个对偶变量指标 i 被选取概率 $P_i > 0$ 。

3 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 $X = \mathbf{R}^{d_1 \times d_2}$, 其中 $d_1 = d_2 = 250$, 数据空间为 $Y = \prod_{i=1}^n Y_i$, $Y_i = \mathbf{R}^{|B_i|}$, 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)可抽象为问题(1)的标准形式:

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

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 $g(x) = 0.2 \|\nabla x\|_{2,1} + I_{\geq 0}(x)$, $\nabla x = (\nabla_1 x, \nabla_2 x) \in \mathbf{R}^{d_1 \cdot d_2 \times 2}$, 范数定义为 $\|\nabla x\|_{2,1} = \sum_j \sqrt{(\nabla_1 x_j)^2 + (\nabla_2 x_j)^2}$; 函数

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

则 $f_i(y)$ 的 Fenchel 共轭可表示为:

$$f_i^*(z) = \sum_{j \in B_i} \begin{cases} -z_j r_j - b_j \log(1 - z_j), & \text{如果 } z_j \leq 1, b_j = 0 \text{ 或 } z_j < 1; \\ \infty, & \text{否则。} \end{cases}$$

$f_i(y)$ 的邻近算子可表示为:

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

由于 g 的邻近算子没有显式表达式, 因此采用梯度投影算法^[24]近似求解。

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:

$$\text{PDHG 算法}^{[1]}: \gamma = 0.99, \sigma_i = \tau = \frac{\gamma}{\|A\|} \approx 6.9 \cdot 10^{-4};$$

$$\text{SPDHG 算法}^{[19]}: \gamma = 0.99, \tau = \frac{\gamma}{(n \max_i \|A_i\|)} \approx 2.2 \cdot 10^{-4}, \sigma_i = \frac{\gamma}{\|A_i\|} \approx 2.2 \cdot 10^{-3};$$

$$\text{ASPDHG 算法}: \delta = 0.85, \mu = 0.95, \beta = 5, \tau_{(0)} = 0.3, \sigma_{(1)} = \beta \cdot \tau = 1.5.$$

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 个子集进行测试, PDHG 算法、SPDH 算法和 ASPDHG 算法在经过 100 个轮次迭代之后得到了如下结果:

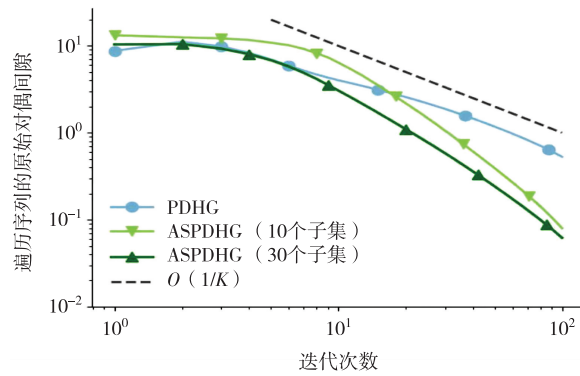


图1 遍历序列的原始对偶间隙

Fig. 1 Primal-dual gap of ergodic sequences

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。从图中得到,Chambolle等人提出的PDHG算法^[1]与本文提出的ASPDHG算法都以 $O(1/N)$ 的速率收敛,与PDHG算法不同的是,ASPDHG算法采用了可变步长的设置,因此在该问题中,ASPDHG算法的收敛速度更快。此外,当ASPDHG算法初始参数一致时,子集数取得越多,即每次迭代更新的对偶变量越少,收敛速度越快。

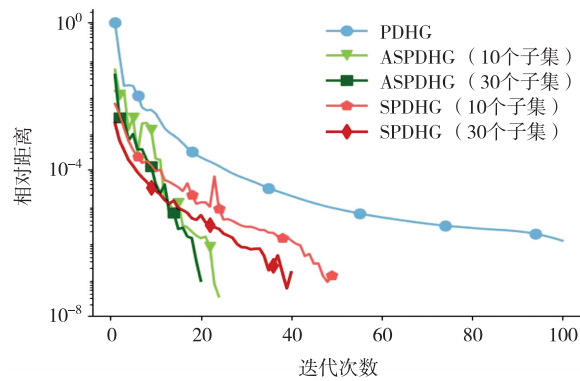
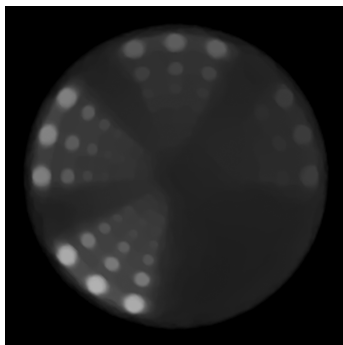


图2 迭代点与鞍点之间的相对距离

Fig. 2 Relative distance between iteration point and saddle point

图2中的相对距离定义为 $[\Phi(x^N) - \Phi(x^*)]/[\Phi(x^0) - \Phi(x^*)]$ 。由图可得,相比非随机算法PDHG^[1],2种随机算法SPDHG^[19]和ASPDHG的收敛效果更好;在非强凸的假设下,SPDHG算法的步长值固定不变,ASPDHG算法自适应步长选取规则能更好的适应实际问题,从而加快收敛速率。

此外,PDHG算法、SPDHG算法(10个子集)和ASPDHG算法(10个子集)在经过10个轮次迭代之后,PET图像重建的结果从左到右依次为:



a PDHG 算法重构 PET 图像



b SPDHG 算法重构 PET 图像



c ASPDHG 算法重构 PET 图像

图3 PET问题图像PDHG算法重构PET图像重建的结果

Fig. 3 The results of positron emission tomography (PET) reconstruction

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图像,如图 3 所示。经过相同轮次的迭代后,两种随机算法 SPDHG 和 ASPDHG 得到的图像较为清晰,且 ASPDHG 算法所得到的图像更加细致。从上图可以清楚地看出,本文提出的 ASPDHG 在实际应用中表现良好。

4 总结与展望

本文提出了一种自适应步随机原始对偶算法,用于解决大规模可分凸优化问题。该算法将线性搜索步长结构与针对大规模问题的随机算法相结合,得到了算法的全局收敛性结果。该算法通过随机选择部分对偶变量更新,有效地降低了计算和存储的成本,从而使得其在处理大规模问题时具有更好的可行性。此外,引入可变步长的策略克服了预先计算算子范数所带来的困难,并且在计算过程中动态调整步长,可以加快收敛速度。实验结果显示,在处理大规模问题时,相较于非随机算法,该算法收敛速度明显更快。即使与其他随机算法相比,在实际应用和收敛速度对比中,该算法展现出了更优秀的性能表现。

经典原始对偶混合梯度算法对一般凸问题和强凸的处理方式有所不同,应用于强凸问题时,经典原始对偶混合梯度算法具有更好的收敛速度。因此,如何将本文提出的算法推广到强凸问题上,以提高算法的收敛效率,这是值得进一步研究的。

参考文献:

- [1] CHAMBOLLE A,POCK T. A first-order primal-dual algorithm for convex problems with applications to imaging[J]. Journal of Mathematical Imaging and Vision,2011,40(1):120-145.
- [2] CHAMBOLLE A,POCK T. An introduction to continuous optimization for imaging[J]. Acta Numerica,2016,25:161-319.
- [3] DARBON J, LANGLOIS G P. Accelerated nonlinear primal-dual hybrid gradient methods with applications to supervised machine learning[EB/OL]. (2021-09-28)[2022-12-16]. <https://arxiv.org/pdf/2109.12222v1.pdf>.
- [4] CHEN S, HALIMI A, REN X, et al. Learning Non-local spatial correlations to restore sparse 3D single-photon data[J]. IEEE Transactions on Image Processing,2020,29:3119-3131.
- [5] GABAY D,MERCIER B. A dual algorithm for the solution of nonlinear variational problems via finite element approximation [J]. Computers & Mathematics with Applications,1976,2(1):17-40.
- [6] ESSER E,ZHANG X,CHAN T F. A general framework for a class of first order primal-dual algorithms for convex optimization in imaging science[J]. SIAM Journal on Imaging Sciences,2010,3(4):1015-1046.
- [7] CHANG X,YANG J,ZHANG H. Golden ratio primal-dual algorithm with linesearch[J]. SIAM Journal on Optimization,2021,32:1584-1613.
- [8] VILLA S,SALZO S,BALDASSARRE L, et al. Accelerated and inexact forward-backward algorithms[J]. SIAM Journal on Optimization,2011,23(3):1607-1633.
- [9] GISELSSON P,BOYD S. Linear convergence and metric selection for Douglas-Rachford splitting and ADMM [J]. IEEE Transactions on Automatic Control,2014,62(2):532-544.
- [10] NEMIROVSKI A S,JUDITSKY A,LAN G, et al. Robust stochastic approximation approach to stochastic programming[J]. SIAM Journal on Optimization,2009,19(4):1574-1609.
- [11] LAN G. An optimal method for stochastic composite optimization[J]. Mathematical Programming,2012,133(1):365-397.
- [12] BLATT D,HERO A O,GAUCHMAN H. A convergent incremental gradient method with a constant step size[J]. SIAM Journal on Optimization,2007,18(1):29-51.
- [13] SCHMIDT M,ROUX N L,BACH F. Minimizing finite sums with the stochastic average gradient [J]. Mathematical Programming,2017,162(1/2):1-30.
- [14] SHALEY-SHWARTZ S,ZHANG T. Stochastic dual coordinate ascent methods for regularized loss[J]. The Journal of Machine Learning Research,2013,14(1):567-599.
- [15] DANG C,LAN G. Randomized methods for saddle point computation[EB/OL]. (2018-09-10)[2022-12-16]. <https://arxiv.org/pdf/1409.8625.pdf>.
- [16] ZHANG Y C,XIAN L. Stochastic primal-dual coordinate method for regularized empirical risk minimization[J]. The Journal of Machine Learning Research,2014,18:81-84.
- [17] ZHU Z,STORKEY A J. Adaptive stochastic primal-dual coordinate descent for separable saddle point problems[EB/OL].

- (2015-09-07)[2022-21-16]. <https://arxiv.org/pdf/1506.04093.pdf>.
- [18] FERCOQ O, BIANCHI P. A coordinate descent primal-dual algorithm with large step size and possibly non separable functions [J]. SIAM Journal on Optimization, 2019, 29(1): 100-134.
- [19] CHAMBOLLE A, EHRHARDT M J, RICHTÁRIK P, et al. Stochastic primal-dual hybrid gradient algorithm with arbitrary sampling and imaging applications[J]. Society for Industrial and Applied Mathematics, 2018, 28(4): 2783-2808.
- [20] MALITSKY Y, POCK T. A first-order primal-dual algorithm with linesearch[J]. SIAM Journal on Optimization, 2018, 28(1): 411-432.
- [21] HE B, FENG M, YUAN X. An algorithmic framework of generalized primal-dual hybrid gradient methods for saddle point problems[J]. Journal of Mathematical Imaging and Vision, 2017, 58(2): 279-293.
- [22] BECK A. First-order methods in optimization[M]. Philadelphia: SIAM, 2017.
- [23] CHAMBOLLE A, POCK T. On the ergodic convergence rates of a first-order primal-dual algorithm[J]. Mathematical Programming, 2016, 159(1/2): 253-287.
- [24] BECK A, TEOULLE M. A fast iterative shrinkage-thresholding algorithm for linear inverse problems[J]. SIAM Journal on Imaging Sciences, 2009, 2(1): 183-202.

Operations Research and Cybernetics

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

ZHOU Xiaoyan, LUO Honglin

(School of Mathematical Sciences, Chongqing Normal University, Chongqing 401331, China)

Abstract: For a class of large-scale separable convex optimization problems, an adaptive stochastic primal-dual algorithm is proposed. The optimization problem is reformulated as a saddle point problem with separable dual variables. Then, the dual variables of the saddle point problem are randomly updated with adaptively selected step size. The adaptive stochastic primal-dual algorithm almost surely converges with rate $O\left(\frac{1}{N}\right)$ in an ergodic sense. The results of numerical experiments indicate that the algorithm can effectively solve the problem of positron emission computed tomography.

Keywords: large scale separable convex optimization problem; stochastic optimization; primal-dual algorithms; adaptive step-size

(责任编辑 陈 乔)