

基于邻近算子求解带凸集约束可分离凸优化问题的原始对偶不动点算法

陈培军^{1,2}, 黄建国¹, 张小群^{1,3}

(1. 上海交通大学数学系, 科学工程计算教育部重点实验室, 上海 200240)
(2. 太原科技大学数学系, 山西 太原 030024)
(3. 上海交通大学自然科学研究院, 上海 200240)

[摘要] 很多实际问题根据不同的物理背景, 解的取值是有一定限制的. 本文拟推广 PDFP²O 算法以求解带闭凸集约束的可分离凸优化问题. 通过将闭凸集约束表示成示性函数而加入目标函数中的技巧, 适当重组函数, 可直接利用 PDFP²O 算法求解, 再利用函数的可分离性, 即可得到闭凸集上的基于邻近算子的原始对偶不动点算法 (PDFP²O_C). 因为 PDFP²O_C 本质上就是利用 PDFP²O 求解与原问题等价的无约束问题, 根据 PDFP²O 的理论结果, 可以方便地得到 PDFP²O_C 的收敛性以及收敛速度. 最后通过 CT 重构说明了算法的有效性.

[关键词] 凸约束, 可分离凸优化, 邻近算子, 不动点算法

[中图分类号] O224; O29 [文献标志码] A [文章编号] 1001-4616(2013)03-0001-05

A Primal-Dual Fixed Point Algorithm Based on Proximity Operator for Convex Set Constrained Separable Problem

Chen Peijun^{1,2}, Huang Jianguo¹, Zhang Xiaoqun^{1,3}

(1. Department of Mathematics, MOE-LSC, Shanghai Jiao Tong University, Shanghai 200240, China)
(2. Department of Mathematics, Taiyuan University of Science and Technology, Taiyuan 030024, China)
(3. Institute of Natural Sciences, Shanghai Jiao Tong University, Shanghai 200240, China)

Abstract: In many real problems, the solutions may have constraints arising from their physical requirements. In this paper, we design an efficient algorithm for solving the separable convex minimization on closed convex set based on the algorithm PDFP²O. Precisely speaking, the constraint can be enforced by adding an indicator function to the objective function, and the function are reformulated and can be solved with PDFP²O. Using the separability of the function with respect to its variables, we thus get a primal-dual fixed point algorithm based on proximity operator on closed convex set (PDFP²O_C). Since the algorithm PDFP²O_C can be recast as the original PDFP²O for unconstrained problem, the convergence and convergence rate analysis can be obtained directly. Finally, we illustrate the efficiency of PDFP²O_C through computerized tomographic reconstruction.

Key words: convex constraint, convex separable minimization, proximity operator, fixed point algorithm

考虑带凸集约束的可分离凸优化问题:

$$\mathbf{x}^* = \arg \min_{\mathbf{x} \in C} (f_1 \circ \mathbf{B})(\mathbf{x}) + f_2(\mathbf{x}), \tag{1}$$

其中 C 为非空闭凸集, $f_1 \in \Gamma_0(\mathbf{R}^m)$, $f_2 \in \Gamma_0(\mathbf{R}^n)$, f_2 在 \mathbf{R}^n 上可微, 且其梯度满足 $1/\beta$ -Lipschitz 连续, $\beta \in (0, +\infty)$, $\mathbf{B}: \mathbf{R}^n \rightarrow \mathbf{R}^m$ 为线性变换. 对实 Hilbert 空间 χ , $\Gamma_0(\chi)$ 表示从 χ 到 $(-\infty, +\infty]$ 适当的下半连续凸函数集. 虽然式(1)形式简单, 但很多实际问题可以归结为该问题的求解^[1]. 典型问题包括图像还原和医学

收稿日期: 2013-03-26.
基金项目: 国家自然科学基金(11101277、11171219、11161130004)、上海市教育委员会 E-研究院建设计划项目(E03004)、上海浦江人才计划基金(11PJ1405900).
通讯联系人: 陈培军, 博士, 副教授, 研究方向: 科学计算. E-mail: chenpeijun@sjtu.edu.cn

图像重构中基于变分形式的稀疏还原模型:

$$\mathbf{x}^* = \arg \min_{\mathbf{x} \in C} \mu \|\mathbf{B}\mathbf{x}\|_{1,2} + \frac{1}{2} \|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2^2, \tag{2}$$

其中 $\|\cdot\|_2$ 表示向量的欧几里德范数, \mathbf{A} 为线性变换的 $p \times n$ 矩阵, $\mathbf{b} \in \mathbf{R}^p, \mu > 0$ 为正则化参数. $\|\mathbf{B}\mathbf{x}\|_{1,2}$ 为常用的基于 ℓ^1 的正则化, 使用该项来强化解在线性变换 \mathbf{B} 下的稀疏性. 在著名的 Rudin-Osher-Fatemi (ROF) 模型中 $\|\mathbf{B}\mathbf{x}\|_{1,2}$ 表示有界变差半范数, 用来恢复分片常数的图像, 其中 \mathbf{B} 为 $2n \times n$ 离散的差分矩阵^[2]. 具体来说, $\|\mathbf{B}\mathbf{x}\|_1$ 和 $\|\mathbf{B}\mathbf{x}\|_{1,2}$ 分别对应于各向异性和各向同性的有界变差. 对 CT 而言, 其还原结果分量不能为负值, 我们可以考虑非负的限制; 对灰度图像而言, 其像素值为 $[0, 255]$, 我们可以考虑还原结果的每个分量落在区间 $[0, 255]$ 中.

我们在文献[3]中对式(1)中 $C = \mathbf{R}^n$ 的情形进行了研究. 在邻近算子和 PFBS^[1]、不动点算法 FP^2O ^[4] 的基础上, 我们提出基于邻近算子的原始对偶不动点算法 (primal-dual fixed point algorithm based on proximity operator, PDFP²O):

$$(\text{PDFP}^2\text{O}) \begin{cases} \mathbf{v}_{k+1} = (\mathbf{I} - \text{prox}_{\frac{\gamma}{\lambda} f_1})(\mathbf{B}(\mathbf{x}_k - \gamma \nabla f_2(\mathbf{x}_k)) + (\mathbf{I} - \lambda \mathbf{B}\mathbf{B}^T)\mathbf{v}_k), \\ \mathbf{x}_{k+1} = \mathbf{x}_k - \gamma \nabla f_2(\mathbf{x}_k) - \lambda \mathbf{B}^T \mathbf{v}_{k+1}, \end{cases} \tag{3}$$

其中 $0 < \gamma < 2\beta, 0 < \lambda \leq 1/\lambda_{\max}(\mathbf{B}\mathbf{B}^T), \lambda_{\max}(\mathbf{B}\mathbf{B}^T)$ 表示 $\mathbf{B}\mathbf{B}^T$ 的最大特征值, \mathbf{I} 表示恒等算子或单位矩阵, prox_f 表示函数 f 的邻近算子, 即对任意 $\mathbf{x} \in \mathbf{R}^m$, 有

$$\text{prox}_f(\mathbf{x}) = \arg \min_{\mathbf{y} \in \mathbf{R}^m} f(\mathbf{y}) + \frac{1}{2} \|\mathbf{x} - \mathbf{y}\|_2^2.$$

文中也给出了 PDFP²O 的收敛性和收敛速度分析. 总体来讲, 算法 PDFP²O 与当前大家公认的图像恢复算法 ASB^[5]、CP^[6]、SIU^[7] 在性能上具有可比性, 而 PDFP²O 在参数选择上有相当的优势, 具有构造性的选取办法. 此外, PDFP²O 的理论分析框架建立在经典的不动点理论框架上, 简单、可拓展性强.

另一方面, 在很多实际应用中, 根据问题的物理背景, 解往往具有一定的约束, 因此研究式(1)是有必要和有意义的. 目前大多数算法普遍采用的做法是: 对迭代的最终结果做相应的处理, 比如投影操作, 以保证解满足相应的性质来近似求解原约束问题. 在本文中, 我们通过将闭凸集的约束表示成示性函数而加入目标函数中的技巧^[1,8], 适当重组函数, 可直接利用 PDFP²O 求解原问题, 再利用函数的可分离性, 最终得到一个求解带闭凸集约束优化问题的基于邻近算子的原始对偶不动点算法 (primal-dual fixed point algorithm based on proximity operator on closed convex set, PDFP²O_C), 详见算法 1, 其中 proj_C 表示闭凸集 C 上的投影算子. 因为 PDFP²O_C 本质上就是利用 PDFP²O 求解与式(1)等价的无约束问题, 因此相应的收敛性和收敛速度分析可由文献[3]中结果直接得到. 我们将该算法应用于 CT 重构以说明算法的有效性.

算法 1 PDFP²O_C

Step 1: 取 $\mathbf{x}_0, \mathbf{y}_0 \in \mathbf{R}^n, \mathbf{v}_0 \in \mathbf{R}^m, 0 < \lambda \leq 1/(\lambda_{\max}(\mathbf{B}\mathbf{B}^T) + 1), 0 < \gamma < 2\beta$.

Step 2: for $k = 0, 1, 2, \dots$

$$\begin{aligned} \mathbf{x}_{k+1/2} &= \mathbf{x}_k - \gamma \nabla f_2(\mathbf{x}_k), \\ \mathbf{v}_{k+1} &= (\mathbf{I} - \text{prox}_{\frac{\gamma}{\lambda} f_1})(\mathbf{B}(\mathbf{x}_{k+1/2} - \lambda(\mathbf{B}^T \mathbf{v}_k + \mathbf{y}_k)) + \mathbf{v}_k), \\ \mathbf{y}_{k+1} &= (\mathbf{I} - \text{proj}_C)(\mathbf{x}_{k+1/2} - \lambda(\mathbf{B}^T \mathbf{v}_k + \mathbf{y}_k) + \mathbf{y}_k), \\ \mathbf{x}_{k+1} &= \mathbf{x}_{k+1/2} - \lambda(\mathbf{B}^T \mathbf{v}_{k+1} + \mathbf{y}_{k+1}). \end{aligned}$$

end for

1 算法 PDFP²O_C 的导出

在本节中, 我们将在算法 PDFP²O 的基础上推导出算法 PDFP²O_C. 记 χ_C 表示定义在闭凸集 C 上的示性函数, 即

$$\chi_C(\mathbf{x}) = \begin{cases} 0, & \mathbf{x} \in C; \\ +\infty, & \mathbf{x} \notin C. \end{cases}$$

则式(1)等价于如下问题:

$$\mathbf{x}^* = \arg \min_{\mathbf{x} \in \mathbf{R}^n} (f_1 \circ \mathbf{B})(\mathbf{x}) + \chi_C(\mathbf{x}) + f_2(\mathbf{x}). \quad (4)$$

记

$$\tilde{\mathbf{B}} = \begin{pmatrix} \mathbf{B} \\ \mathbf{I} \end{pmatrix}, \quad (5)$$

$$\tilde{f}_1(\mathbf{v}, \mathbf{y}) = f_1(\mathbf{v}) + \chi_C(\mathbf{y}), \quad \forall (\mathbf{v}, \mathbf{y}) \in \mathbf{R}^m \times \mathbf{R}^n, \quad (6)$$

则有

$$(\tilde{f}_1 \circ \tilde{\mathbf{B}})(\mathbf{x}) = (f_1 \circ \mathbf{B})(\mathbf{x}) + \chi_C(\mathbf{x}), \quad \forall \mathbf{x} \in \mathbf{R}^n. \quad (7)$$

因此, (1) 等价于

$$\mathbf{x}^* = \arg \min_{\mathbf{x} \in \mathbf{R}^n} (\tilde{f}_1 \circ \tilde{\mathbf{B}})(\mathbf{x}) + f_2(\mathbf{x}). \quad (8)$$

根据文献[3], 我们可以使用 PDFP²O (见式(3)) 求解式(8), 其迭代格式为

$$\begin{cases} \tilde{\mathbf{v}}_{k+1} = (\mathbf{I} - \text{prox}_{\tilde{\chi}_{\tilde{f}_1}})(\tilde{\mathbf{B}}(\mathbf{x}_k - \gamma \nabla f_2(\mathbf{x}_k)) + (\mathbf{I} - \lambda \tilde{\mathbf{B}}\tilde{\mathbf{B}}^T)\tilde{\mathbf{v}}_k), \\ \mathbf{x}_{k+1} = \mathbf{x}_k - \gamma \nabla f_2(\mathbf{x}_k) - \lambda \tilde{\mathbf{B}}^T \tilde{\mathbf{v}}_{k+1}, \end{cases} \quad (9)$$

其中 $0 < \gamma < 2\beta, 0 < \lambda \leq 1/\lambda_{\max}(\tilde{\mathbf{B}}\tilde{\mathbf{B}}^T) = 1/(\lambda_{\max}(\mathbf{B}\mathbf{B}^T) + 1)$, $\tilde{\mathbf{v}}_k = \begin{pmatrix} \mathbf{v}_k \\ \mathbf{y}_k \end{pmatrix}$. 因为 \tilde{f}_1 关于变量 \mathbf{v}, \mathbf{y} 是可分离的, 且

$\text{prox}_{\tilde{\chi}_C} = \text{proj}_C^{[1]}$, 因此, 式(9)等价于

$$(\text{PDFP}^2\text{O}_C) \begin{cases} \mathbf{v}_{k+1} = (\mathbf{I} - \text{prox}_{\chi_{f_1}})(\mathbf{B}(\mathbf{x}_k - \gamma \nabla f_2(\mathbf{x}_k)) + (\mathbf{I} - \lambda \mathbf{B}\mathbf{B}^T)\mathbf{v}_k - \lambda \mathbf{B}\mathbf{y}_k), \\ \mathbf{y}_{k+1} = (\mathbf{I} - \text{proj}_C)((\mathbf{x}_k - \gamma \nabla f_2(\mathbf{x}_k)) - \lambda \mathbf{B}^T \mathbf{v}_k + (1 - \lambda)\mathbf{y}_k), \\ \mathbf{x}_{k+1} = \mathbf{x}_k - \gamma \nabla f_2(\mathbf{x}_k) - \lambda \mathbf{B}^T \mathbf{v}_{k+1} - \lambda \mathbf{y}_{k+1}. \end{cases} \quad (10)$$

根据式(10), 我们可以容易地得到算法 1. 因为上述算法等价于直接用 PDFP²O 求解式(8), 因此, 根据文献[3]中定理 3.2、3.5、3.7, 易知上述算法的收敛性和收敛速度. 容易验证, 如果 $C = \mathbf{R}^n$, 取 $\mathbf{y}_0 = \mathbf{0}$, 则迭代每一步中 $\mathbf{y}_{k+1} = \mathbf{0}$, 式(10)退化为式(3). 需要指出的是, 此时虽然二者形式上完全相同, 但它们允许的 λ 的取值范围不同, 分别为 $0 < \lambda \leq 1/\lambda_{\max}(\tilde{\mathbf{B}}\tilde{\mathbf{B}}^T)$ 和 $0 < \lambda \leq 1/\lambda_{\max}(\mathbf{B}\mathbf{B}^T)$.

2 数值实验

我们以 CT(computer tomographic) 重构算例^[9]来说明 PDFP²O_C 的计算效果. 在简化的并行光束断层成像问题中, 将待观测的物体的切片看成一个二维函数. 投影由线积分得到, 刻画了一束 X-射线穿过物体的衰减情况, 投影算子可表示为离散 Radon 变换. CT 重构问题可描述为通过有限的测量得到的线积分值来还原得到相应的函数. 临床应用中常用的算法是滤波反投影(Filtered Back Projection, FBP)算法. 在有噪音的情形下, 因为 Radon 变换是无界的和病态的, 求解该问题是非常困难的. 文献中, 也常用模型式(2)迭代重构. 这里 \mathbf{A} 为 Radon 变换, \mathbf{b} 为投影向量. 算子 $\mathbf{B}: \mathbf{R}^n \rightarrow \mathbf{R}^{2n}$ 为离散梯度算子. 我们采用各项同性的有界变差, 即对任意的 $\boldsymbol{\omega} = (\omega_1, \dots, \omega_n, \omega_{n+1}, \dots, \omega_{2n})^T \in \mathbf{R}^{2n}$, $f_1(\boldsymbol{\omega}) = \mu \|\boldsymbol{\omega}\|_{1,2}$, 其中

$$\|\boldsymbol{\omega}\|_{1,2} = \sum_{i=1}^n \sqrt{\omega_i^2 + \omega_{n+i}^2}.$$

令 $\mathbf{w}_i = (\omega_i, \omega_{n+i})^T$, $\|\mathbf{w}_i\|_2 = \sqrt{\omega_i^2 + \omega_{n+i}^2}$, $\varepsilon = \frac{\mu\gamma}{\lambda}$. 则 $\text{prox}_{\varepsilon \|\cdot\|_{1,2}}(\boldsymbol{\omega})$ 可被表示为

$$(\text{prox}_{\varepsilon \|\cdot\|_{1,2}}(\boldsymbol{\omega}))_{i,n+i} = \max\{\|\mathbf{w}_i\|_2 - \varepsilon, 0\} \frac{\mathbf{w}_i}{\|\mathbf{w}_i\|_2}, i = 1, \dots, n.$$

在实现算法 1 中, 我们直接计算 $(\mathbf{I} - \text{prox}_{\chi_{f_1}})(\boldsymbol{\omega})$. 事实上,

$$(\mathbf{I} - \text{prox}_{\frac{\mu\gamma}{\lambda} \|\cdot\|_{1,2}})(\boldsymbol{\omega}) = \text{proj}_{\varepsilon}(\boldsymbol{\omega}),$$

其中 $\text{proj}_{\varepsilon}$ 为从 \mathbf{R}^{2n} 到半径 ε 的 $\ell_{2,\infty}$ 球的投影算子, 即

$$(\text{proj}_{\varepsilon}(\boldsymbol{\omega}))_{i,n+i} = \min\{\|\mathbf{w}_i\|_2, \varepsilon\} \frac{\mathbf{w}_i}{\|\mathbf{w}_i\|_2}, i = 1, \dots, n. \quad (11)$$

这里我们仍使用文献[3,7]中的测试实例, 即 \mathbf{A} 为给定的投影矩阵, \mathbf{b} 由 128×128 基于 Shep-Logan 模

型的图像和 \mathbf{A} 作用生成的 50 个一致定向的投影,然后加上均值为 0、方差为 1 的白噪声得到. $\lambda_{\max}(\mathbf{B}\mathbf{B}^T)=8, \lambda_{\max}(\widetilde{\mathbf{B}}\widetilde{\mathbf{B}}^T)=\lambda_{\max}(\mathbf{B}\mathbf{B}^T)+1=9.$ β 可取为 $1/\lambda_{\max}(\mathbf{A}\mathbf{A}^T)$, 经计算 $\lambda_{\max}(\mathbf{A}\mathbf{A}^T)\approx 1.5086$, 因此 PDFP²O 中我们可以取 $0<\gamma<2/\lambda_{\max}(\mathbf{A}\mathbf{A}^T)\approx 1.3257, 0<\lambda\leq 1/8$. 根据文献[3]中我们的观察结果:在允许的范围内, γ, λ 接近上界时,算法的性能较好. 和文献[3]相同, PDFP²O 中仍取 $\lambda=1/8, \gamma=1.3, \mu=0.1$, FBP 得到的结果为 $\mathbf{x}_0, \mathbf{v}_0=\mathbf{0}$. PDFP²O_C 中 $\lambda=1/9, \mathbf{y}_0=\mathbf{0}$, 其余参数和 PDFP²O 相同. 由于真实图像像素值的取值范围为 $0\sim 255$, 因此,我们要求解的分量位于区间 $[0, 255]$. 对于每迭代步, PDFP²O_C 比 PDFP²O 增加了 3 次向量加(减)法运算和 1 次向量每个分量往 $[0, 255]$ 的投影操作. 相对于矩阵向量积运算,其增加的计算量很小. 通过图 1 可以看到关于指标 PSNR, PDFP²O_C 优于 PDFP²O, 尤其是对于前面迭代步的结果. 最后二者得到的图像差不多,但 PDFP²O_C 仍有稍高的 PSNR. 我们分别给出两种方法迭代 50、100、500 步得到的图像, 见图 2. 从中可以清楚地看到 PDFP²O_C 已经得到了基本很清晰的图像, 而 PDFP²O 得到的图像仍有很多的伪影. 能量下降方面 PDFP²O_C 一开始优于 PDFP²O, 最后的值比 PDFP²O 所得结果大.

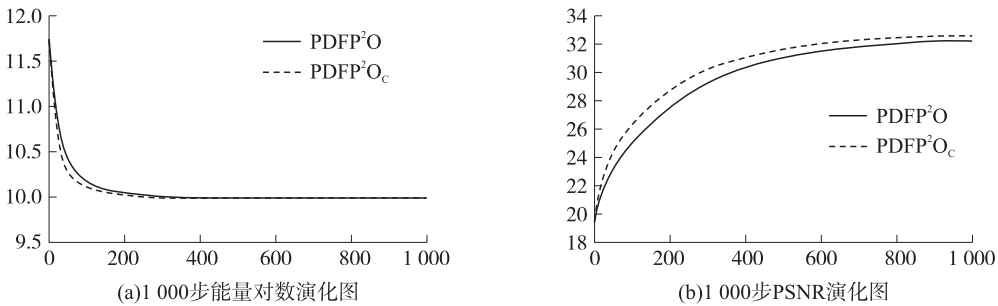


图 1 PDFP²O 与 PDFP²O_C 能量和 PSNR 的比较

Fig. 1 The comparison of energy and PSNR between PDFP²O and PDFP²O_C

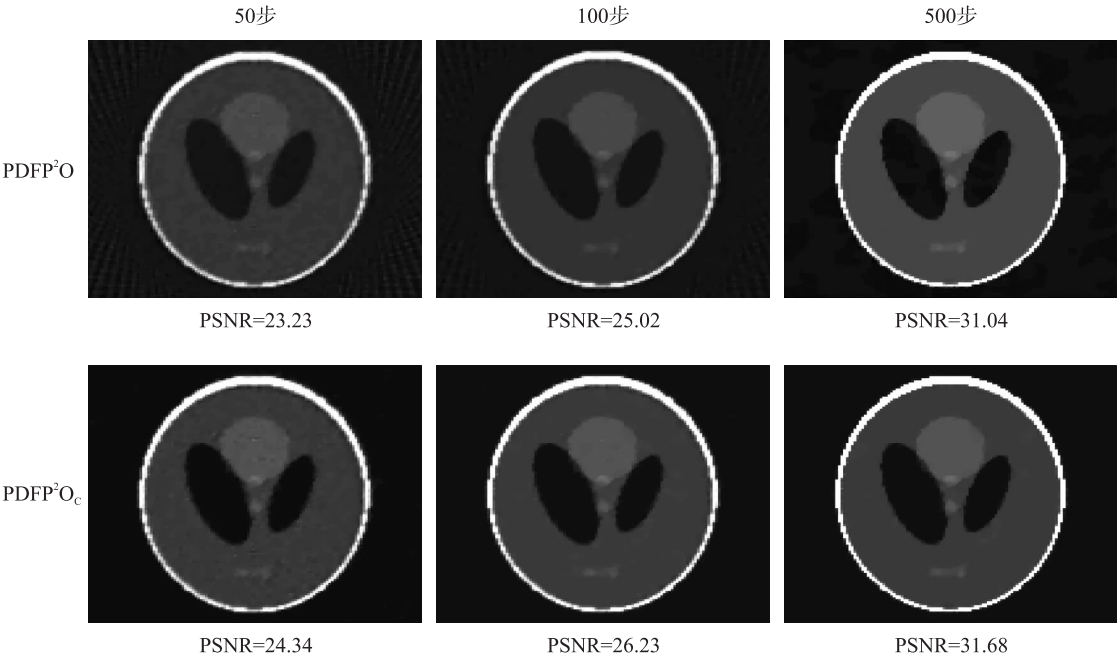


图 2 PDFP²O 与 PDFP²O_C 所得图像的比较

Fig. 2 The comparison of reconstruction images between PDFP²O and PDFP²O_C

3 结论

在本文中,我们基于算法 PDFP²O 获得求解带闭凸集约束问题式(1)的基于邻近算子的原始对偶不动点算法(PDFP²O_C). 该方法本质上仍然是 PDFP²O 算法,因此由 PDFP²O 算法的收敛性与收敛速度结果可自然得到 PDFP²O_C 算法的相应结果. 通过数值实验,我们可以看到,对一些应用,PDFP²O_C 确实优于

PDFP²O. 对 CP 和 SIU 等其它算法,我们也可以和 PDFP²O_c 一样类似地引入投影操作.

[参考文献]

- [1] Combettes P L, Wajs V R. Signal recovery by proximal forward-backward splitting[J]. Multiscale Modeling and Simulation, 2005, 4(4): 1 168–1 200.
- [2] Rudin L I, Osher S, Fatemi E. Nonlinear total variation based noise removal algorithms[J]. Physica D: Nonlinear Phenomena, 1992, 60(1/4): 259–268.
- [3] Chen P, Huang J, Zhang X. A primal-dual fixed point algorithm for convex separable minimization with applications to image restoration[J]. Inverse Problems, 2013, 29(2): 025011(1–33).
- [4] Micchelli C A, Shen L, Xu Y. Proximity algorithms for image models: denoising[J]. Inverse Problems, 2011, 27(4): 045009(1–30).
- [5] Goldstein T, Osher S. The split Bregman method for L_1 -regularized problems[J]. SIAM Journal on Imaging Sciences, 2009, 2(2): 323–343.
- [6] 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.
- [7] Zhang X, Burger M, Osher S. A unified primal–dual algorithm framework based on Bregman iteration[J]. Journal of Scientific Computing, 2011, 46(1): 20–46.
- [8] 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): 1 015–1 046.
- [9] Avinash C, Malcolm S. Principles of Computerized Tomographic Imaging[M]. Philadelphia: SIAM, 2001.

[责任编辑: 丁 蓉]