J. Vis. Commun. Image R. 38 (2016) 814–823

Contents lists available at ScienceDirect

J. Vis. Commun. Image R.

journal homepage: www.elsevier.com/locate/jvci

A simple primal–dual method for total variation image restoration ☆

Benxin Zhang [a], Zhibin Zhu [b,*], Shuo Wang [b]

[a] School of Electronic Engineering and Automation, Key Laboratory of Automatic Detecting Technology and Instruments, Guilin University of Electronic Technology, Guilin 541004, PR China

[b] School of Mathematics and Computing Science, Guangxi Colleges and Universities Key Laboratory of Data Analysis and Computation, Guilin University of Electronic Technology, Guilin 541004, PR China

ARTICLE INFO

Article history:
Received 22 March 2016
Revised 27 April 2016
Accepted 29 April 2016
Available online 2 May 2016

Keywords:
Primal–dual method
Saddle-point problem
Total variation
Image reconstruction

ABSTRACT

In this study we propose a simple primal–dual method for total variation minimization problems. A predictor–corrector scheme to the dual variable is used in our algorithms and convergence of the method is proved. We also show that the iterative scheme has $O(1/N)$ convergence rate in the ergodic sense, where N denotes the iteration number. Numerical results including image deblurring and computerized tomography reconstruction demonstrate the efficient of the new algorithms.

1. Introduction

In this paper, we consider the following total variation (TV) image restoration model:

$$\min_x \frac{\alpha}{2}\|Ax-b\|^2 + \|x\|_{TV}, \tag{1}$$

where $x \in X$ is the original image, $\|\cdot\|_{TV}$ is the TV norm. $X = R^n$, $\Omega = X \times X$ are finite dimensional vector spaces equipped with an inner product $\langle\cdot,\cdot\rangle$ and the Euclidean norm $\|\cdot\|$. A is a possibly large and ill-conditioned matrix representing a linear transform, such as blurring operator, Radon transform. b is the degraded image or data. $\alpha > 0$ is a weighting parameter. If A is an identity matrix, then the problem (1) is the well-known Rudin–Osher–Fatemi denoise model [34]. The total variation of x has the following equivalent dual form:

$$\|x\|_{TV} = \max_{y\in\Omega} \nabla x^T y = \max_{y\in\Omega} -x^T \nabla\cdot y, \tag{2}$$

where $y \in \{\Omega : |y| \le 1\}$ is the dual variable. ∇ is gradient operator and bounded. $\nabla\cdot$ is the divergence operator and also bounded. Using the dual formulation of the TV norm, the objective function of (1) can be written as

$$\min_x \max_{y\in\Omega} \quad Q(x,y) = \frac{\alpha}{2}\|Ax-b\|^2 - x^T\nabla\cdot y, \tag{3}$$

Some ideas from the duality were proposed first by Chan et al. [9], later by Chambolle [7]. Chambolle's project algorithm has been very popular for total variation image denoising. Based on the method, there are so many total variation minimization algorithms. In [46], the authors proposed a nonmonotone Chambolle gradient projection algorithm. They used the well known Barzilai–Borwein stepsize instead of the constant time stepsize in Chambolle's method. Further, they adopt adaptive nonmonotone line search proposed by Dai and Fletcher [16] to guarantee the global convergence. Thus, the approach cannot be directly applied to solve the minimax problem of (3).

A benchmark algorithm for the problem (1) is the alternating direction method of multipliers (ADMM) proposed in [21] or [20]. This algorithm has shown to be very efficient and useful for a large class of convex separable programming problems. The famous split Bregman algorithm [22] is also equivalent to ADMM. In [8], the authors shown that their primal–dual algorithm was equivalent to the preconditioned version of the ADMM [18]. An $O(1/N)$ efficiency estimate of ADMM has been established in [25,26,35] and many relevant references. However, recently, Chen et al. [12] proved that ADMM like methods do not directly extend to problems involving multi-block convex minimization problems. Davis and Yin introduce a new operator-splitting scheme for solving a variety of problems that are reduced to a monotone inclusion of three operators and give the simplest extension of the classic ADMM from 2 to 3 blocks of variables in the paper [15]. Tseng [38] and Nemirovski [28] proposed some prox methods for convex–concave optimization problem (3) which have a convergence rate of $O(1/N)$, where their methods are provided that the gradi-

☆ This paper has been recommended for acceptance by M.T. Sun.

* Corresponding author.
E-mail address: optimization_zhu@163.com (Z. Zhu).

http://dx.doi.org/10.1016/j.jvcir.2016.04.025
1047-3203/

ents are Lipschitz continuous. We know that the optimal rate of any first order method is $O(1/\sqrt{N})$ for general nonsmooth objective functions. Although it is not improvable in general, Nesterov [29,30] recently study show that the optimal rate is able to improve to $O(1/N)$ by exploring the special structure of the objective function. Beck and Teboulle [5] proposed the famous fast iterative shrinkage-thresholding algorithm by the forward–backward splitting. Using the Nesterov's technique [31], their method has an improved complexity result of $O(1/N^2)$ which is an optima first order method for nonsmooth problems. Combettes and Pesquet [13] and Bot et al. [6] proposed primal–dual splitting algorithms for finding zeros of maximal monotone operators [2]. In [14], Davis analyzed a general monotone inclusion problem that captures a large class of primal–dual splittings as a special case and firstly shown the nonergodic convergence rates in the literature.

The organization of this paper is as follows. In Section 2, we illustrate our motivation of algorithmic design and show our new method. Then, we prove its convergence in Section 3. In Section 4, we give some extensions and discuss our scheme from proximal point algorithm perspective. In Section 5, some numerical results are given to illustrate the efficiency of the method. Finally, the conclusion is given.

2. Related work and proposed method

In this section, we present some related work for solving the problem (1) and (3) and give two new primal–dual algorithms.

2.1. Related work

Let B be gradient operator ∇, the problem (1) can be expressed as

$$\min_x \frac{\alpha}{2}\|Ax-b\|^2 + \|Bx\|_1. \tag{4}$$

By introducing a new intermediate variable z, The unconstrained optimization problem (4) is reformulated as

$$\min_{x,z} \frac{\alpha}{2}\|Ax-b\|^2 + \|z\|_1, \quad s.t.\ z = Bx. \tag{5}$$

For solving the problem (5), the iterative scheme of ADMM is

$$\begin{cases} x_{k+1} = \arg\ \min\limits_x \frac{\alpha}{2}\|Ax-b\|^2 + \langle \lambda_k, Bx - z_k\rangle + \frac{\beta}{2}\|Bx - z_k\|^2, \\ z_{k+1} = \arg\ \min\limits_z \|z\|_1 + \langle \lambda_k, Bx_{k+1} - z\rangle + \frac{\beta}{2}\|Bx_{k+1} - z\|^2, \\ \lambda_{k+1} = \lambda_k + \beta(Bx_{k+1} - z_{k+1}), \end{cases} \tag{6}$$

where λ_k is the Lagrange multiplier and $\beta > 0$ is a penalty parameter. We refer to this as an implicit algorithm. Instead of applying the augmented Lagrangian method (ALM) in [27,32] directly to (5), ADMM splits the subproblem of ALM in the Gauss–Seidel way such that the variables x and z can be minimized individually in alternating order.

Obviously, we need to compute the matrix of $(\alpha A^TA + \beta B^TB)^{-1}$ about the x subproblem of (6) which is quite time consuming when the dimension is large. In order to solve this issue, Zhang et al. [47] proposed a unified primal–dual algorithm framework based on Bregman iteration. The general idea of their algorithm is to replace the augmented Lagrangian primal minimizations x_{k+1} and z_{k+1} of (6) by proximal-like iterations. More precisely, the algorithm can be described as follows:

$$\begin{cases} x_{k+1} = \arg\ \min\limits_x \frac{\alpha}{2}\|Ax-b\|^2 + \langle \lambda_k, Bx - z_k\rangle + \frac{\beta}{2}\|Bx - z_k\|^2 + \frac{1}{2}\|x - x_k\|^2_{R_1}, \\ z_{k+1} = \arg\ \min\limits_z \|z\|_1 + \langle \lambda_k, Bx_{k+1} - z\rangle + \frac{\beta}{2}\|Bx_{k+1} - z\|^2 + \frac{1}{2}\|z - z_k\|^2_{R_2}, \\ \lambda_{k+1} = \lambda_k + \beta(Bx_{k+1} - z_{k+1}), \end{cases} \tag{7}$$

where R_1, R_2 are positive semi-define matrices. Their proposed algorithm can be classified as split inexact Uzawa (SIU) methods. If we choose $R_1 = \frac{1}{\theta} - \alpha A^TA - \beta B^TB, R_2 = 0,\ 0 < \theta < \frac{1}{\|\alpha A^A + \beta B^TB\|},\ \beta > 0.$ (7) can be further expressed as:

$$\begin{cases} x_{k+1} = x_k - \theta(\alpha A^T(Ax_k - b) + \beta B^T(Bx_k - z_k + \frac{\lambda_k}{\beta})), \\ z_{k+1} = shrink(Bx_{k+1} + \frac{1}{\beta}\lambda_k, \frac{1}{\beta}), \\ \lambda_{k+1} = \lambda_k + \beta(Bx_{k+1} - z_{k+1}), \end{cases} \tag{8}$$

where $shrink(t,\mu)$=sign(t)$\cdot\max\{|t| - \frac{1}{\mu}, 0\}$ and sign$(\cdot)$ is the sign function. In this case, the step corresponds to a one-step gradient descend and it is very efficient since it does not involve any operator inverting. We can also notice that this is a explicit algorithm. This type of approach is named generalized or proximal alternating direction method of multipliers. Convergence rate has been discussed in [19,40].

As analyzed in [8,18,48], an optimal solution x^* of the problem (1) can be obtained through a saddle point of $Q(x,y)$ which is defined by the Eq. (3). So more and more scholars have proposed some primal–dual algorithms for total variation image restoration problems (3). In [48], Zhu and Chan firstly proposed the famous primal–dual hybrid gradient (PDHG) algorithm as follows:

$$\begin{cases} y_{k+1} = P_\Omega(y_k + \tau_k \nabla x_k), \\ x_{k+1} = x_k - \theta_k(\alpha A^T(Ax^{k+1} - b) - \nabla\cdot y_{k+1}), \end{cases} \tag{9}$$

where the projection operate $P_\Omega(\cdot)$ is defined by

$$P_\Omega(y) = \arg\ \min\{\|z - y\|_2 : z \in \Omega\},$$

and $\tau_k,\ \theta_k$ are adaptive stepsize. For A is the matrix representation of a space-invariant blurring operator K, the Fourier transform of matrix multiplication by A becomes pointwise multiplication in the frequency domain. Hence, the second equation of (9) can be efficiently solved by

$$x_{k+1} = \mathcal{F}^{-1}\left(\frac{\mathcal{F}(x_k + \theta\nabla\cdot y_{k+1}) + \theta\alpha\mathcal{F}(K)^* \odot \mathcal{F}(b)}{1 + \theta\alpha\mathcal{F}(K)^* \odot \mathcal{F}(K)}\right), \tag{10}$$

where $\mathcal{F}(\cdot)$ and $\mathcal{F}^{-1}(\cdot)$ are the fast Fourier transform (FFT) and inverse FFT operators, respectively, "$*$" denotes the complex conjugate, and "$\odot$" is the pointwise multiplication operator. Though the algorithm is quite fast, the convergence is not proved. In [24], He et al. showed that PDHG algorithm with constant step sizes is indeed convergent if one of the functions of the saddle-point problem is strongly convex. But this condition of convergence is too strong to apply in most of image processing fields. Bonettini and Ruggiero [3] established the convergence of a general primal–dual method for nonsmooth convex optimization problems. In their paper, they showed that the convergence of the scheme can be considered as an ϵ-subgradient method on the primal formulation of the variational problem when the steplength parameters are a priori selected sequences. PDHG algorithm is the special case of their scheme.

Chambolle and Pock [8] proposed a primal–dual extrapolation algorithm (named CP) as follows:

$$\begin{cases} y_{k+1} = P_\Omega(y_k + \tau\nabla\bar{x}_k), \\ x_{k+1} = x_k - \theta(\alpha A^T(Ax^{k+1} - b) - \nabla\cdot y_{k+1}), \\ \bar{x}_{k+1} = 2x_{k+1} - x_k, \end{cases} \tag{11}$$

where $\tau\theta < \frac{1}{8}$. They showed that the convergence rate of this algorithm is $O(1/N)$. In [23], He and Yuan gave a novel study on these primal–dual algorithms from the perspective of contraction perspective. The method simplified the existing convergence analysis.

In [36], Tseng proposed a solution method for primal dual problem that alternated between a proximal step and a projection-type step. The splitting scheme for solving (3) is