Applied Mathematics and Computation 333 (2018) 76–94

Contents lists available at ScienceDirect

Applied Mathematics and Computation

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

A modified primal-dual method with applications to some sparse recovery problems

Yongchao Yu[a,*], Jigen Peng[b,c]

[a] School of Mathematics and Statistics, Xinyang Normal University, Xinyang 464000, China
[b] School of Mathematics and Statistics, Xi'an Jiaotong University, Xi'an 710049, China
[c] School of Mathematics and Information Sciences, Guangzhou University, 230 GuangZhou University City Outer Ring Road, Guangzhou 510006, China

A R T I C L E I N F O

Keywords:
Primal-dual method
Proximity operator
Sparse recovery problems

A B S T R A C T

In this paper, we first present a modified Chambolle–Pock primal-dual method (MCPPDM) to solve a convex composite optimization problem which minimizes the sum of two convex functions with one composed by a linear operator. It is well known that the Chambolle–Pock primal-dual method (CPPDM) with the combination parameter being 1 is an application of the proximal point algorithm and thus is convergent, however, when the combination parameter is not 1, the method may be not convergent. To choose flexibly the combination parameter, we develop a slightly modified version with little additional computation cost. In CPPDM, one variable is updated twice but another variable is updated only once at each iteration. However, in the modified version, two variables are respectively updated twice at each iteration. Another main task of this paper is that we reformulate some well-known sparse recovery problems as special cases of the convex composite optimization problem and then apply MCPPDM to address these sparse recovery problems. A large number of numerical experiments have demonstrated that the efficiency of the proposed method is generally comparable or superior to that of existing well-known methods such as the linearized alternating direction method of multipliers and the graph projection splitting algorithm in terms of solution quality and run time.

1. Introduction

In this paper, we consider the convex composite optimization problem

$$\text{minimize } f(x) + g(Ax), \tag{1}$$

where $f : \mathcal{X} \to \mathbb{R} \cup \{+\infty\}$ and $g : \mathcal{Y} \to \mathbb{R} \cup \{+\infty\}$ are proper, lower semicontinuous, convex functions defined respectively on finite-dimensional Hilbert spaces $\mathcal{X}$ and $\mathcal{Y}$ with the corresponding inner product $\langle \cdot,\cdot \rangle$ and the norm $\|\cdot\| = \sqrt{\langle \cdot, \cdot \rangle}$, $A : \mathcal{X} \to \mathcal{Y}$ is a linear operator with the adjoint A^* and the induced norm $\|A\|$. The input variable x can be a vector, matrix, or an element from a composite Hilbert space. By introducing an auxiliary variable $y \in \mathcal{Y}$, we can rewrite the problem (1) as the form

$$\text{minimize } f(x) + g(y) \tag{2}$$

* Corresponding author.
E-mail addresses: sparseelad@126.com (Y. Yu), jgpengxjtu@126.com (J. Peng).

https://doi.org/10.1016/j.amc.2018.03.089

$$\text{subject to} \quad y = Ax. \tag{2}$$

Parikh and Boyd in [1] first called the problem form (2) *graph form* since the variables x and y are constrained to lie in the graph $\{(x, y) \in \mathcal{X} \times \mathcal{Y} : y = Ax\}$ of the linear operator A. The problem (1) or graph form (2) includes many popular convex optimization problems such as linear and quadratic programming, general cone programming [3], and captures a wide variety of specific applications such as the lasso problem, logistic regression, support vector machine in statistics and machine learning [4], basis pursuit problem in signal processing [12], total-variation minimization problems in image processing [29], radiation treatment planning [5] and portfolio optimization [6], and so on. We invite the reader to consult [1,2] for more applications. In this paper, we will focus on three classes of more attractive sparse recovery problems.

Compressive sensing. The fundamental problem in compressive sensing [8–11] is to recover a high-dimensional sparse or approximately signal $\bar{x} \in \mathbb{R}^n$ from very few nonadaptive linear and noisy measurements $b \in \mathbb{R}^m$, that is,

$$b = \Psi\bar{x} + e,$$

where $\Psi \in \mathbb{R}^{m\times n}$ $(m \ll n)$ is a sensing matrix and $e \in \mathbb{R}^m$ is a noise vector. The theory of compressive sensing has indicated that one can recover accurately and efficiently the original sparse signal via solving the quadratically constrained Basis Pursuit (BP_ε) problem

$$\begin{aligned} &\text{minimize} \quad \|x\|_1 \\ &\text{subject to } \|\Psi x - b\|_2 \le \varepsilon, \end{aligned} \tag{3}$$

where ε is an estimated upper bound of the noise level. In the noiseless case ($e = 0$), we can set $\varepsilon = 0$ and then obtain the well-known Basis Pursuit (BP) problem [12]

$$\begin{aligned} &\text{minimize} \quad \|x\|_1 \\ &\text{subject to } b = \Psi x. \end{aligned} \tag{4}$$

Another more theoretically effective estimator for recovering sparse signals from noisy measurements is the Dantzig selector (DS) introduced in [13], which is the solution to the convex problem

$$\begin{aligned} &\text{minimize} \quad \|x\|_1 \\ &\text{subject to } \|\tilde{D}^{-1}\Psi^T(\Psi x - b)\|_\infty \le \gamma, \end{aligned} \tag{5}$$

where γ is a scalar related to the noise level, $||\cdot||_\infty$ is the ℓ_∞ norm and $\tilde{D}$ is the diagonal matrix whose diagonal entries are the ℓ_2 norms of the columns of Ψ.

Low-rank matrix completion. Some applications such as the well-known Netflix problem [14], collaborative filtering [16], system identification [15,18], global positioning [17], can be expressed as a low-rank matrix completion problem consisting of recovering an unknown low-rank matrix $M \in \mathbb{R}^{m\times n}$ from a give subset Ω of observed entries. The problem can be solved via the convex optimization

$$\begin{aligned} &\text{minimize} \quad \|X\|_* \\ &\text{subject to } X_{ij} = M_{ij}, \quad \forall\,(i, j) \in \Omega, \end{aligned} \tag{6}$$

where $\|\cdot\|_*$ is the nuclear norm of a matrix. By introducing the projection operator $\mathcal{P}_\Omega : \mathbb{R}^{m\times n} \to \mathbb{R}^{m\times n}$ defined as

$$[\mathcal{P}_\Omega X]_{ij} = \begin{cases} X_{ij}, & \text{if}(i, j) \in \Omega, \\ 0, & \text{if}(i, j) \notin \Omega, \end{cases}$$

and the resulting matrix $M_\Omega = \mathcal{P}_\Omega M$, we can reformulate the problem (6) as

$$\begin{aligned} &\text{minimize} \quad \|X\|_* \\ &\text{subject to } \mathcal{P}_\Omega X = M_\Omega. \end{aligned} \tag{7}$$

Robust principal component analysis. Compared to classical PCA capturing the low-rank structure of the data matrix corrupted by random gaussian noises, robust principal component analysis (RPCA) introduced by Candès et al. in [19] aims to capture low-rank structure of the data matrix including sparse gross errors. In [19], the authors have claimed that RPCA can be cast as the following convex optimization

$$\begin{aligned} &\text{minimize } \|L\|_* + \nu\|S\|_1 \\ &\text{subject to} \quad D = L + S, \end{aligned} \tag{8}$$

where $\|S\|_1 = \sum_{i,j} |(S)_{ij}|$, $D \in \mathbb{R}^{m\times n}$ is a data matrix and the parameter $\nu > 0$ is used to balance the weights of rank and sparsity.

Due to wide applicability of the problem (1) in different scientific research fields, it is very key to develop fast and efficient algorithms for solving (1). Many different algorithms have been designed for (1) based on properties of f and g or the composite function $g \circ A$. Specifically, we distinguish three cases for these algorithms. In the first case where the function f is differentiable with a Lipschitz continuous gradient and the composite function $g \circ A$ has inexpensive proximity operator, the proximal forward-backward algorithm [20] and its fast version based on Nesterov's acceleration technique [21] are most