Numer. Math. Theor. Meth. Appl. Vol. **2**, No. 4, pp. 377-402
doi: 10.4208/nmtma.2009.m9002s November 2009

A Coordinate Gradient Descent Method for Nonsmooth Nonseparable Minimization

Zheng-Jian Bai[1], Michael K. Ng[2,*] and Liqun Qi[3]

[1] *School of Mathematical Sciences, Xiamen University, Xiamen 361005, China.*
[2] *Centre for Mathematical Imaging and Vision and Department of Mathematics, Hong Kong Baptist University, Kowloon Tong, Hong Kong.*
[3] *Department of Applied Mathematics, The Hong Kong Polytechnic University, Hung Hom, Hong Kong.*

Received 18 February 2009; Accepted (in revised version) 9 June 2009

Abstract. This paper presents a coordinate gradient descent approach for minimizing the sum of a smooth function and a nonseparable convex function. We find a search direction by solving a subproblem obtained by a second-order approximation of the smooth function and adding a separable convex function. Under a local Lipschitzian error bound assumption, we show that the algorithm possesses global and local linear convergence properties. We also give some numerical tests (including image recovery examples) to illustrate the efficiency of the proposed method.

AMS subject classifications: 65F22, 65K05

Key words: Coordinate descent, global convergence, linear convergence rate.

1. Introduction

We consider a nonsmooth optimization problem of minimizing the sum of a smooth function and a convex nonseparable function as follows.

$$\min_x F_c(x) \stackrel{\text{def}}{=} f(x) + cP(x), \tag{1.1}$$

where $c > 0$, $P : \mathbb{R}^n \to (-\infty, \infty]$ is proper, convex, lower semicontinuous (lsc) function, and f is smooth (i.e., continuously differentiable) on an open subset of $\mathbb{R}^n$ containing $\text{dom}P = \{x | P(x) < \infty\}$. In this paper, we assume that P is a nonseparable function in the form $P(x) := \|Lx\|_1$, where $L \neq I$ is preferred to be a sparse matrix. In particular, we focus on a special case of (1.1) defined by

$$\min_x F_c(x) \stackrel{\text{def}}{=} f(x) + c\|Lx\|_1, \tag{1.2}$$

*Corresponding author. *Email addresses:* `zjbai@xmu.edu.cn` (Z.-J. Bai), `mng@math.hkbu.edu.hk` (M. K. Ng), `maqilq@polyu.edu.hk` (L. Qi)

where L is the first order or second order differentiation matrix. Problem (1.1) with $P(x) = \|x\|_1$ and Problem (1.2) arise in many applications, including compressed sensing [9, 13, 24], signal/image restoration [5, 19, 23], data mining/classification [3, 14, 21], and parameter estimation [8, 20].

There has been considerable discussion on the problem (1.1), see for instance [2, 6, 7, 11, 15]. If P is also smooth, then a coordinate gradient descent based on Armijo-type rule was well developed for the unconditional minimization problem (1.1) in Karmanov [10, pp. 190–196 and pp. 246–250], where the global convergence and geometrical convergence are provided if $F_c(x)$ is assumed to be strongly convex. Recently, Tseng and Yun [22] gave a coordinate gradient descent method with stepsize chosen by an Armijo-type rule for the problem (1.1) under the assumption that P is (block) separable, where the coordinates are updated in either the Gauss-Seidel rule or the Gauss-Southwell-r rule or the Gauss-Southwell-q rule. Moreover, the global convergence and linear convergence for this method were established. However, the method cannot be employed to solve (1.2) directly since $P(x) = \|Lx\|_1$ is no longer a (block) separable function.

Recently, various methods have been considered for image restoration / deblurring/ denoising problems with ℓ_1-regularization, see for instance [5,17,19,23,25]. In particular, Fu *et al.* [5] gave a primal-dual interior point method for solving the following optimization problem with $\ell 1$ regularization:

$$\min_x \|Ax - b\|_2^2 + c\|Lx\|_1, \tag{1.3}$$

where A is a linear blurring operator, x is the original true image, and b is the observed blurred image. In each interior point iteration, the linear system is solved by a preconditioned conjugate gradient method. However, the number of conjugate gradient iterations are still large since the linear system is ill-conditioned and the performance of the preconditioner depends on the support of the blurring function and on how fast such function decays in spatial domain. Osher *et al.* [17, 25] presented the Bregman iterative algorithm for solving (1.3) with L being the identity matrix or the first order differentiation matrix. In each Bregman iteration, we need to solve an unconstrained convex subproblem.

In this paper, we aim to provide a coordinate gradient descent method with stepsize chosen by an Armijo-type rule to solve the problem (1.2) and (1.3) efficiently, especially when the problem dimension is large. Our idea is to find a coordinate-wise search direction by finding a minimum in a subproblem, which is obtained by a strictly convex quadratic approximate of f and adding a separable function term (derived from $P(x) = \|Lx\|_1$). Then, we update the current iterate in the direction of the coordinate-wise minimizer. We will show that the coordinate-wise minimizer can be sufficient close to the coordinate-wise minimizer of the subproblem of the sum of the same strictly convex quadratic approximate of f and $P(x) = \|Lx\|_1$ if the parameter c is small enough. This approach can be implemented simply and is capable to solve large-scale problems. We show that our algorithm converges globally if the coordinates are chosen by either the Gauss-Seidel rule or the Gauss-Southwell-r rule or the Gauss-Southwell-q rule. Moreover, we prove that our approach with Gauss-Southwell-q rule converges at least linearly based on a local Lips-