Journal of Computational and Applied Mathematics 298 (2016) 105–122

Contents lists available at ScienceDirect

Journal of Computational and Applied Mathematics

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

Descent gradient methods for nonsmooth minimization problems in ill-posed problems

Pham Quy Muoi [a,*], Dinh Nho Hào [b], Peter Maass [c], Michael Pidcock [d]

[a] Department of Mathematics, Danang University of Education, 459 Ton Duc Thang, Danang, Viet Nam

[b] Hanoi Institute of Mathematics, 18 Hoang Quoc Viet Road, 10307 Hanoi, Viet Nam

[c] Center for Industrial Mathematics, University of Bremen, Bibliothekstr. 1, 28359 Bremen, Germany

[d] Department of Mechanical Engineering and Mathematical Sciences, Oxford Brookes University, Wheatley Campus, Wheatley, Oxford OX33 1HX, United Kingdom

ARTICLE INFO

Article history:
Received 18 April 2015
Received in revised form 15 September 2015

Keywords:
Nonlinear inverse problems
Total variation regularization
Sparsity regularization
Nonnegative sparse regularization
Descent gradient method
Nesterov's accelerated algorithm

ABSTRACT

Descent gradient methods are the most frequently used algorithms for computing regularizers of inverse problems. They are either directly applied to the discrepancy term, which measures the difference between operator evaluation and data or to a regularized version incorporating suitable penalty terms. In its basic form, gradient descent methods converge slowly.

We aim at extending different optimization schemes, which have been recently introduced for accelerating these approaches, by addressing more general penalty terms. In particular we use a general setting in infinite Hilbert spaces and examine accelerated algorithms for regularization methods using total variation or sparsity constraints.

To illustrate the efficiency of these algorithms, we apply them to a parameter identification problem in an elliptic partial differential equation using total variation regularization.

1. Introduction

In this paper, we investigate algorithms for the approximate solution of minimization problems of the form

$$\min_{u\in\mathcal{H}} \Theta(u) \quad \text{where } \Theta(u) := F(u) + \Phi(u). \tag{1}$$

Here $\mathcal{H}$ is a Hilbert space, $F, \Phi : \mathcal{H} \to \mathbb{R}$, where F is smooth, but not necessarily convex, and Φ is convex but nonsmooth. Such problems arise in many applications including regularization methods for inverse problems using Tikhonov regularization [1], where e.g. $F(u) := \|K(u) - y^\delta\|^p$ denotes a discrepancy term measuring the difference between operator evaluation and data and Φ denotes a penalty term such as total variation [2] or a sparsity constraint [3–7]. Typical applications include parameter identification for partial differential equations, see e.g. [8], or pose estimation problems in computer vision as treated e.g. in the EU-SceneNet project.[1]

This is a well studied problem both in most general settings as well as in more detail for some special cases of $\mathcal{H}$, F and Φ. Many algorithms have been proposed for solving Problem (1), for an incomplete list concerning mostly quadratic discrepancy terms see e.g. [3–5,9–11]. When F is not a quadratic functional and Φ is a sparsity functional, there are

* Corresponding author.
E-mail addresses: phamquymuoi@gmail.com (P.Q. Muoi), hao@math.ac.vn (D.N. Hào), pmaass@math.uni-bremen.de (P. Maass), mkpidcock@brookes.ac.uk (M. Pidcock).

1 EU FET-Open SME project 'SceneNet' (GA 309169).

http://dx.doi.org/10.1016/j.cam.2015.11.039
0377-0427/

fewer algorithms available in practice e.g. [5,9,10,12–15]. Most of these methods are gradient-type algorithms and thus they converge quite slowly. Consequently, accelerated algorithms of gradient-type methods [12,13,15] and of semismooth Newton as well as quasi-Newton methods, see e.g. [9], have been studied intensively.

In order to clarify our contributions to this problem, we briefly describe some closely related work in [12,13,15], which serves as the motivation for the present paper. In [12,13], the authors propose accelerated algorithms for solving Problem (1) in finite dimensional spaces, $\mathcal{H} = \mathbb{R}^N$. Both these papers address general functionals F and Φ, but then all algorithms are applied only for the sparsity functional $\Phi(u) = \sum_{i=1}^{N} |u_i|$.

Recently, the algorithms given in [12,13] have been extended to Problem (1) in an infinite Hilbert space $\mathcal{H}$ for nonconvex F, but only for sparsity constraints i.e. Φ given by

$$\Phi_p(u) = \sum_{i\in\Lambda} \omega_i |\langle u, \varphi_i\rangle|^p, \tag{2}$$

where $\omega_i \geq \omega_0 > 0$ for all i and $\{\varphi_i\}_{i\in\Lambda}$ is an orthonormal basis (or frame) of $\mathcal{H}$ [15].

In summary, in [12,13,15], the authors have proposed a gradient-type method and two accelerated versions, they include numerical examples for the case of a sparsity functional Φ. When F is convex, the convergence of two accelerated algorithms is proven there and the objective functional decreases with rate $O(\frac{1}{n^2})$, where n is the iteration counter. This convergence rate is known to be the best possible for algorithms that use only the values of the objective functional and its gradient.

Our contributions in this paper are as follows:

- We will extend the algorithms in [12,13,15] to Problem (1) considered in a more general setting, i.e. in a general Hilbert space with general functionals F and Φ. We emphasize that the result on convergence of the gradient-type method given in Theorem 2.1 is new, and cannot be deduced from previous ones in [12,13,15]. This is the main contribution of the paper. Further contributions are an analysis of the effects of stepsize selection on the convergence rate and efficiency of the descent gradient iteration. We implement two accelerated versions and present results with different numerical examples.
- We will apply these algorithms to two popular regularization methods: sparsity regularization, nonnegative sparse regularization and total variation regularization. We emphasize that our algorithms (convergence analysis, numerical examples) applied to the total variation regularization has – to the best of our knowledge – not been addressed in the literature before. The efficiency of this approach will be illustrated in the section on numerical examples. For *nonnegative sparse regularization*, a thorough investigation on well-posedness, convergence rates as well as numerical algorithms is given in [16].

There are many existing nonsmooth optimization numerical methods in the literature. Such methods are divided into three classes: (1) nonsmooth black-box algorithms; (2) proximal mapping algorithms; and (3) smoothing algorithms. For the first class, the algorithms are developed based on the notion of the subgradient and its generalization. For more details on such algorithms, we refer to the books and papers [17–19] and the references therein. The algorithms analyzed in this paper falls into the second class and a good overview on algorithms and numerical comparisons are given in [9,15] for sparse regularization. For algorithms in the final class, we refer to the literature, e.g., [20,21] and the references therein.

The paper is organized as follows: in Section 2, we describe the descent gradient method and prove its convergence as well as discussing stepsize choices. In Section 3 we present two accelerated versions of the descent gradient method for Problem (1) with convex F. Section 4 is devoted to applications of our algorithms to the minimization problem in sparse regularization, nonnegative sparse regularization and total variation regularization. Finally, in Section 5 the algorithms are implemented and analyzed for the identification of the diffusion coefficient problem in an elliptic equation.

2. A descent gradient method

2.1. Problem setting and descent gradient iteration

Throughout this paper we make the following assumptions about Φ and F in order to be able to establish the convergence properties of the algorithms.

Assumption 2.1. (1) Φ is a positive, proper, convex, weakly lower semicontinuous and weakly coercive functional with $\mathrm{Dom}(\Phi) \neq \emptyset$.

(2) F is bounded from below and weakly lower semicontinuous. Without loss of generality, we assume $F(u) \geq 0,\ \forall u \in \mathcal{H}$.

(3) F is Lipschitz continuously Fréchet differentiable, i.e. it is Fréchet differentiable and there exists a constant L such that

$$\|F'(u) - F'(u')\| \leqslant L\|u - u'\|, \quad \forall u, u' \in \mathcal{H}.$$

(4) If $\{u^n\}$ converges weakly to u such that $\{\Theta(u^n)\}$ is monotonically decreasing, then there exists a subsequence $\{u^{n_j}\}$ such that

$$\{F'(u^{n_j})\} \to F'(u).$$