Applied Mathematics and Computation 253 (2015) 345–362

Contents lists available at ScienceDirect

Applied Mathematics and Computation

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

On some steplength approaches for proximal algorithms

Federica Porta [a,*], Ignace Loris [b]

[a] Dipartimento di Scienze Fisiche, Informatiche e Matematiche, Università di Modena e Reggio Emilia, Italy
[b] Département de Mathématique, Université Libre de Bruxelles, Belgium

ARTICLE INFO

Keywords:
Proximal algorithms
Steplength selection
Non-smooth optimization
Signal recovering

ABSTRACT

We discuss a number of novel steplength selection schemes for proximal-based convex optimization algorithms. In particular, we consider the problem where the Lipschitz constant of the gradient of the smooth part of the objective function is unknown. We generalize two optimization algorithms of Khobotov type and prove convergence. We also take into account possible inaccurate computation of the proximal operator of the non-smooth part of the objective function. Secondly, we show convergence of an iterative algorithm with Armijo-type steplength rule, and discuss its use with an approximate computation of the proximal operator. Numerical experiments show the efficiency of the methods in comparison to some existing schemes.

1. Introduction

In this paper we consider the problem of minimizing the sum of two given functions

$$\min_{\boldsymbol{x}\in\mathbb{R}^N} \mathcal{F}(\boldsymbol{x}) \equiv f(\boldsymbol{x}) + g(\boldsymbol{x}), \tag{1}$$

where $f : \mathbb{R}^N \longrightarrow \mathbb{R}$ is a convex, continuously differentiable function and $g : \mathbb{R}^N \longrightarrow \bar{\mathbb{R}}$ is an extended-value convex function, possibly including constraints on the unknown.

The minimization problem (1) has been handled by several algorithms especially tailored to deal with a non-differentiable function g. In particular, numerical schemes known in the literature as *proximal gradient methods* have earned a great popularity in the last years. They find a very general applicability in problems concerning with large or high-dimensional datasets from several scientific areas, like compressed sensing, machine learning and signal processing (see for example [1–4]).

In this paper we discuss new techniques to select the steplength in the proximal gradient methods, without the assumption of knowing the Lipschitz constant of the gradient of the smooth part of the objective function. We start our analysis from two approaches developed in the constrained differentiable optimization context: the Khobotov extra-gradient method [5,6] and the gradient projection method along the feasible directions [7]. Due to the non-smooth term g in the function (1), a possible generalization of these algorithms has to account for the presence of the proximal operator of this term instead of a projection onto a suitable set.

We study several extensions of constrained optimization algorithms of Khobotov type. In particular, we propose an extension of Khobotov's original scheme [5] to the more general proximal case. We prove convergence, even when the proximal operators cannot be computed exactly. We also consider the saddle-point formulation for the minimization problem (1). The

* Corresponding author.
E-mail addresses: federica.porta@unimore.it (F. Porta), igloris@ulb.ac.be (I. Loris).

http://dx.doi.org/10.1016/j.amc.2014.12.079
0096-3003/

so-called Alternating Extragradient Method (AEM) [8] is a variant of Khobotov's method for constrained smooth saddle-point problems. We propose a generalization of the AEM algorithm for a general (not necessarily smooth) saddle-point problem. This extension is again achieved through the use of the proximal operator of the non-smooth part of the objective function. Again, none of these algorithms require any knowledge of the Lipschitz constant of the gradient of the smooth part of the objective function. Such a problem is also recently discussed in [9].

Secondly, and following the basic idea behind the gradient projection methods, we suggest an iterative proximal algorithm that exploits an Armijo-type steplength selection rule similar to [10]. A proof of convergence of the algorithm is provided. We also explore its use in case only an approximation for the required proximal operator is available.

Finally, in order to evaluate the effectiveness of the presented methods, we conduct a numerical study on some signal recovering test problems that can be modeled by Eq. (1): the performance of the discussed schemes is assessed through a comparison with some algorithms already known in the literature and designed to solve this type of problems.

Several problems arising from real-world applications [11–13] can be formalized through the mathematical model introduced in Eq. (1): the applications of this work will be focused on one-dimensional and two-dimensional signal restoration problems with data perturbed by Poisson noise [14–16]. Signal and image restoration consist in recovering an approximation of an object detected by an acquisition system, starting from the data provided by the instrument and a model representing the distortion occurring during the acquisition process itself. More precisely, the signal formation process is an inverse problem that can be formalized through a linear system $\boldsymbol{g} = H\boldsymbol{x} + \boldsymbol{b} + \eta$ where $\boldsymbol{g} \in \mathbb{R}^M$ is the observed data, $\boldsymbol{x} \in \mathbb{R}^N$ represents an ideal, undistorted object to be recovered, $H \in \mathbb{R}^{M \times N}$ is a typically ill-conditioned matrix describing the acquisition instrument effect, $\boldsymbol{b} \in \mathbb{R}^M$ expresses a non-negative constant background radiation and $\eta \in \mathbb{R}^M$ is the noise corrupting the data. In this paper we will work under the hypothesis of having non-negative signals, therefore we will take into account this type of constraint in the problem formulation. In the Bayesian approach [17,18], the approximated restored signal is found by solving the following optimization problem

$$\min_{\boldsymbol{x} \geqslant \boldsymbol{0}} J_0(\boldsymbol{x}) + \mu J_R(\boldsymbol{x}), \tag{2}$$

where $J_0 : \mathbb{R}^N \longrightarrow \mathbb{R}$ is a continuously differentiable function measuring the distance between the model and the data, $J_R : \mathbb{R}^N \longrightarrow \mathbb{R}$ is a regularization term adding a priori information on the solution and μ is a positive parameter balancing the role of the two objective function components J_0 and J_R. When the data are affected by Poisson noise, the so-called Kullback–Leibler divergence is used to describe J_0:

$$J_0(\boldsymbol{x}) = \mathrm{KL}(\boldsymbol{x}) = \sum_{i=1}^{N} \left\{ \boldsymbol{g}_i \ln \frac{\boldsymbol{g}_i}{(H\boldsymbol{x} + \boldsymbol{b})_i} + (H\boldsymbol{x} + \boldsymbol{b})_i - \boldsymbol{g}_i \right\} \tag{3}$$

with $\boldsymbol{g}_i \ln(\boldsymbol{g}_i) = 0$ if $\boldsymbol{g}_i = 0$. As for the regularization term, we will consider properly chosen functionals that enforce a priori information depending on the features of the problem.

2. Mathematical tools

This section recalls some useful definitions and properties on proximal operators and describes a well-known proximal gradient method. For a more complete discussion of proximal operator methods we refer the reader to [4,3,19,20]. In the following we consider convex function that are proper (nowhere equal to $-\infty$ and not identically equal to $+\infty$) and lower semi-continuous.

2.1. Proximal operators

The *proximal operator* $\mathbf{prox}_h : \mathbb{R}^N \longrightarrow \mathbb{R}^N$ of a convex function $h : \mathbb{R}^N \longrightarrow \bar{\mathbb{R}}$ is defined as:

$$\mathbf{prox}_h(\boldsymbol{u}) = \underset{\boldsymbol{x} \in \mathbb{R}^N}{\operatorname{argmin}} \ \frac{1}{2} \|\boldsymbol{x} - \boldsymbol{u}\|^2 + h(\boldsymbol{x}).$$

We remark that if h is convex and closed then $\mathbf{prox}_h(\boldsymbol{u})$ exists and is unique for all $\boldsymbol{u} \in \mathbb{R}^N$.

Lemma 1. [Subgradient characterization] *Let $h : \mathbb{R}^N \longrightarrow \bar{\mathbb{R}}$ be an extended-value function. The following characterization for the proximal operator of h holds true: $\boldsymbol{x} = \mathbf{prox}_h(\boldsymbol{u})$ if and only if $\boldsymbol{u} - \boldsymbol{x} \in \partial h(\boldsymbol{x})$ if and only if $h(\boldsymbol{z}) \geqslant h(\boldsymbol{x}) + \langle \boldsymbol{u} - \boldsymbol{x}, \boldsymbol{z} - \boldsymbol{x} \rangle, \ \forall \boldsymbol{z} \in \mathbb{R}^N$.*

Proof. See [3]. □

Remark 1. From Lemma 1 and by setting $\boldsymbol{w} = \boldsymbol{u} - \boldsymbol{x}$, it follows that $\boldsymbol{w} \in \partial h(\boldsymbol{x})$ iff $\boldsymbol{x} = \mathbf{prox}_h(\boldsymbol{x} + \boldsymbol{w})$.

Remark 2. The minimizer $\hat{\boldsymbol{x}}$ of problem (1) is characterized by the inclusion $\boldsymbol{0} \in \nabla f(\hat{\boldsymbol{x}}) + \partial g(\hat{\boldsymbol{x}})$, or equivalently by the relations $\alpha \nabla f(\hat{\boldsymbol{x}}) + \boldsymbol{w} = 0$ and $\boldsymbol{w} \in \partial \alpha g(\hat{\boldsymbol{x}})$ with $\alpha > 0$. Using Remark 1, these can be rewritten as the single condition