An Inexact Augmented Lagrangian Framework for Non-Convex Optimization with Nonlinear Constraints

Mehmet Fatih Sahin  Armin Eftekhari  Ahmet Alacaoglu  Fabian Latorre  Volkan Cevher  
Laboratory for Information and Inference Systems (LIONS)  
Institute of Electrical Engineering  
École Polytechnique Fédérale de Lausanne  
Lausanne, VD 1015, Switzerland  
Email: armin.eftekhari@epfl.ch

Abstract—We investigate the convergence rate analysis of the classical inexact augmented Lagrangian method (iALM) for nonlinear-constrained non-convex problems subject to a geometric condition involving the nonlinear operator. We show that when coupled with a first-order method, iALM finds first-order stationary points in $\tilde{O}(1/\epsilon^3)$ calls to the first-order oracle. In addition, when coupled with a second-order method, iALM finds second-order stationary points in $\tilde{O}(1/\epsilon^5)$ calls to the second-order oracle. We provide numerical evidence on large-scale signal processing and machine learning problems, modeled typically via the second-order oracle. We provide numerical evidence on large-scale

I. INTRODUCTION

We study the nonconvex optimization problem

$$\min_{x \in \mathbb{R}^d} \ f(x) + g(x)$$

$$A(x) = b,$$ (1)

where $f: \mathbb{R}^d \to \mathbb{R}$ is possibly nonconvex and $A: \mathbb{R}^d \to \mathbb{R}^m$ is a nonlinear operator and $b \in \mathbb{R}^m$. For clarity of notation, we take $b = 0$ in the sequel, the extension to any $b$ is trivial. We assume that $g: \mathbb{R}^d \to \mathbb{R}$ is a proximal-friendly (but possibly nonsmooth) convex function.

A host of problems in computer science [8], [11], machine learning [12], [18], and signal processing [16], [17] naturally fall under the template of (1), including max-cut, clustering, generalized eigenvalue, and community detection.

To address these applications, this paper builds up on the vast literature on the classical inexact augmented Lagrangian framework and proposes a simple, intuitive and easy-to-implement algorithm for solving (1) with total iteration complexity results, under an interpretable geometric condition detailed below.

To solve (1), the inexact Augmented Lagrangian Method (iALM) is widely used [3], [4], [9], due to its cheap per-iteration cost and also its empirical success in practice. Every (outer) iteration of iALM calls a solver to inexacty solve an intermediate augmented Lagrangian subproblem to near stationarity, and the user has freedom in choosing this solver, which could be a first-order algorithm (say, proximal gradient descent [15]) or a second-order algorithm, such as BFGS [14].

We argue that, unlike its convex counterpart [13], [10], [20], the convergence rate and the complexity of iALM for (1) are not well-understood. Indeed, addressing this important theoretical gap is one of the key contribution of the present work.

II. SUMMARY OF CONTRIBUTIONS

\(\triangleright\) Our framework is future-proof in the sense that we obtain the convergence rate of iALM for (1) with an arbitrary solver for finding first- and second-order stationary points of each intermediate subproblem.  
\(\triangleright\) We investigate the use of different solvers for augmented Lagrangian subproblems and provide overall iteration complexity bounds for finding first- and second-order stationary points of (1). Our complexity bounds match the best theoretical complexity results in optimization.  
\(\triangleright\) We propose a novel geometric condition that simplifies the algorithmic analysis of iALM. We verify the condition for a few key problems, including basis pursuit below.

III. EXAMPLE: BASIS PURSUIT

As one example of our framework, consider Basis Pursuit (BP), which finds sparse solutions of an under-determined system of linear equations, namely,

$$\min_{x \in \mathbb{R}^d} \ ||z||_1$$

$$Bz = b,$$ (2)

where $B \in \mathbb{R}^{n \times d}$ and $b \in \mathbb{R}^n$. BP has found many applications in machine learning, statistics and signal processing [7], [5], [1]. A huge number of primal-dual convex optimization algorithms are proposed to solve BP, including, but not limited to [19], [6]. There also exists many line of works [2] to handle sparse regression problem via regularization with $\ell_1$ norm.

Here, we take a different approach and cast (2) as an instance of (1) to find the equivalent problem

$$f(x) = ||x||_2^2, \ g(x) = 0$$

$$A(x) = Bx - z^2 - b.$$ (3)

We draw the entries of $B$ independently from a zero-mean and unit-variance Gaussian distribution. For a fixed sparsity level $k$, the support of $z \in \mathbb{R}^d$ and its nonzero amplitudes are also drawn from the standard Gaussian distribution. Then the measurement vector is created as $b = Bz + \epsilon$, where $\epsilon$ is the noise vector with entries drawn independently from the zero-mean Gaussian distribution with variance $\sigma^2 = 10^{-6}$.

Figure 1 compiles our results for the proposed relaxation. The true potential of our reformulation is in dealing with more structured norms rather than $\ell_1$, where computing the proximal operator is often intractable.

REFERENCES

Figure 1. Convergence with different subsolvers for the basis pursuit problem.