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 $\tilde{\mathcal{O}}(1/\epsilon^3)$ calls to the first-order oracle. In addition, when coupled with a second-order method, iALM finds second-order stationary points in $\tilde{\mathcal{O}}(1/\epsilon^5)$ calls to the second-order oracle. We provide numerical evidence on large-scale signal processing and machine learning problems, modeled typically via semidefinite relaxations, for which we verify the geometric condition.**

We study the nonconvex optimization problem

$$\begin{cases} \min\limits_{x\in\mathbb{R}^d} f(x)+g(x) \\ A(x)=b, \end{cases} \tag{1}$$

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 inexactly 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.

Summary of contributions:

▷ 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.

▷ 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.

▷ We propose a novel geometric condition that simplifies the algorithmic analysis of iALM. We verify the condition for a few key problems.

I. Example: Basis Pursuit

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

$$\begin{cases} \min_z \|z\|_1 \\ Bz=b, \end{cases} \tag{2}$$

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 ℓ_1 norm.

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

$$\begin{aligned} f(x) =& \|x\|_2^2, \quad g(x)=0 \\ A(x) =& \overline{B}x^{\circ 2}-b. \end{aligned} \tag{3}$$

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 ϵ 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 ℓ_1, where computing the proximal operator is often intractable.

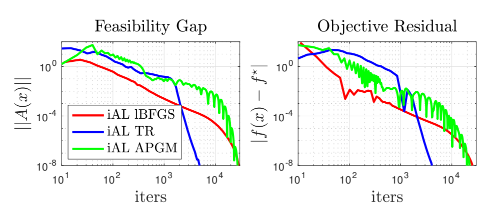

Figure 1. Convergence with different subsolvers for the basis pursuit problem.

REFERENCES

[1] Sanjeev Arora, Mikhail Khodak, Nikunj Saunshi, and Kiran Vodrahalli. A compressed sensing view of unsupervised text embeddings, bag-of-n-grams, and lstms. 2018.

[2] Amir Beck and Marc Teboulle. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM journal on imaging sciences*, 2(1):183–202, 2009.

[3] Samuel Burer and Renato DC Monteiro. A nonlinear programming algorithm for solving semidefinite programs via low-rank factorization. *Mathematical Programming*, 95(2):329–357, 2003.

[4] Samuel Burer and Renato DC Monteiro. Local minima and convergence in low-rank semidefinite programming. *Mathematical Programming*, 103(3):427–444, 2005.

[5] Emmanuel J Candès and Michael B Wakin. An introduction to compressive sampling. *IEEE signal processing magazine*, 25(2):21–30, 2008.

[6] Antonin Chambolle and Thomas Pock. A first-order primal-dual algorithm for convex problems with applications to imaging. *Journal of mathematical imaging and vision*, 40(1):120–145, 2011.

[7] Scott Shaobing Chen, David L Donoho, and Michael A Saunders. Atomic decomposition by basis pursuit. *SIAM review*, 43(1):129–159, 2001.

[8] Subhash Khot and Assaf Naor. Grothendieck-type inequalities in combinatorial optimization. *arXiv preprint arXiv:1108.2464*, 2011.

[9] Brian Kulis, Arun C Surendran, and John C Platt. Fast low-rank semidefinite programming for embedding and clustering. In *Artificial Intelligence and Statistics*, pages 235–242, 2007.

[10] Guanghui Lan and Renato DC Monteiro. Iteration-complexity of first-order augmented lagrangian methods for convex programming. *Mathematical Programming*, 155(1-2):511–547, 2016.

[11] László Lovász. Semidefinite programs and combinatorial optimization. In *Recent advances in algorithms and combinatorics*, pages 137–194. Springer, 2003.

[12] Elchanan Mossel, Joe Neeman, and Allan Sly. Consistency thresholds for the planted bisection model. In *Proceedings of the forty-seventh annual ACM symposium on Theory of computing*, pages 69–75. ACM, 2015.

[13] Valentin Nedelcu, Ion Necoara, and Quoc Tran-Dinh. Computational complexity of inexact gradient augmented lagrangian methods: application to constrained mpc. *SIAM Journal on Control and Optimization*, 52(5):3109–3134, 2014.

[14] J. Nocedal and S. Wright. *Numerical Optimization*. Springer Series in Operations Research and Financial Engineering. Springer New York, 2006.

[15] Neal Parikh, Stephen Boyd, et al. Proximal algorithms. *Foundations and Trends® in Optimization*, 1(3):127–239, 2014.

[16] Amit Singer. Angular synchronization by eigenvectors and semidefinite programming. *Applied and computational harmonic analysis*, 30(1):20, 2011.

[17] Amit Singer and Yoel Shkolnisky. Three-dimensional structure determination from common lines in cryo-em by eigenvectors and semidefinite programming. *SIAM journal on imaging sciences*, 4(2):543–572, 2011.

[18] Le Song, Alex Smola, Arthur Gretton, and Karsten M Borgwardt. A dependence maximization view of clustering. In *Proceedings of the 24th international conference on Machine learning*, pages 815–822. ACM, 2007.

[19] Quoc Tran-Dinh, Ahmet Alacaoglu, Olivier Fercoq, and Volkan Cevher. An adaptive primal-dual framework for nonsmooth convex minimization. *arXiv preprint arXiv:1808.04648*, 2018.

[20] Yangyang Xu. Iteration complexity of inexact augmented lagrangian methods for constrained convex programming. *arXiv preprint arXiv:1711.05812v2*, 2017.