View Reviews

Paper ID
2762

Paper Title
An inexact augmented Lagrangian framework for nonconvex optimization with nonlinear constraints

Reviewer #1

Questions

1. Please enter a detailed review describing the strengths and weaknesses of the submission.
Summary : This paper provides convergence rates for inexact augmented Lagrangian method, which take into account the complexity of the inner loop used to solve the subproblems in the algorithm.

Strengths : The paper analyses the case where the iterations of the augmented Lagrangian method are not exactly computed, which is often the case in practice. Hence, having a precise idea of how the error may affect the whole convergence is essential. Here, the author(s) consider two types of solvers, which are widely used.

Comparisons of their results with those found in recent literature are clearly written.

Generally speaking, the paper is well-written and easy to follow. It seems technically correct as well.

Weaknesses : The geometric condition is not sufficiently motivated, though it is claimed to be central to ensure the convergence of the scheme. In particular, the devoted paragraph "Regularity" means to give an intuition thanks to a simple example, but the links with the condition are not clear.

The convergence rate is given with respect to the optimality first-order conditions. Any clue for other type of convergence?

The influence of the sequence \beta_k choice is not clear to me. How should it be chosen in practice?

2. Please provide an overall score for the submission.
Weak Reject: Borderline, tending to reject

3. Please enter a 2-3 sentence summary of your review explaining your overall score.
The paper seems technically correct, and provide interesting results on convergence rates for a widely used scheme.

5. Please rate your confidence in the score assigned.
Low: Reviewer is making an educated guess.

8. I agree to keep the paper and code submissions confidential, and delete any submitted code at the end of the review cycle to comply with the confidentiality requirement.
Agreement accepted

Reviewer #2

Questions

1. Please enter a detailed review describing the strengths and weaknesses of the submission.
This paper studies an very interesting problem and claims a very strong result. However, there are quite a few issues in this paper that make the results questionable. Please see my comments below.

1. The authors should clear state in the abstract and the beginning of the paper that they can find a second-order stationary point only when g=0. I only realized this until page 3 because of the current writing. When g is non-smooth, defining a second-order stationary point is very tricky.

2. In algorithm 1, I know we need to choose rho to ensure (46) holds. However, I don't know how to set rho in the algorithm because (46) involve unknown quantity. Some explanation is needed. Moreover, I don't see this rho ever appears again anywhere in the algorithm. If so, why do we need it as an input?? I know rho is used in the proof in order to bound lambda_beta in (17). However, I don't know why rho is also mentioned in the first inequality in (41) in order to prove (40). I think you only need the second inequality in (41) to prove (40) so rho should not be introduced in (41).

3. In line 166 in the right column, y_max(...) equals what?

4. The stopping condition in algorithm 1 is problematic. Because sigma_k<1, you can only guarantee |Ax| <=tau/sigma_{k+1} which can be very large. Even if you can theoretically guarantee a small |Ax| for sufficiently large number of iterations, this stopping condition will terminate the algorithm early and do not return a small |Ax|.

5. Condition (20) is very interesting and deserves more study. However, I am not convinced by the justification to assumption (20) on the right column of page 4. I understand the example in Figure 1 but this example only shows that, if (20) fails, the algorithm converges slowly but still converges to a feasible solution. However, if I understand correctly, you are trying to claim that if (20) fails, algorithm cannot find a feasible solution. More exploitation is appreciated.

6. I don't understand why you need to introduce the interval [k0,k1] in theorem 4.1. You can just require (20) holds for all k because there are only finitely many k<=k0 so that you can always find a small enough v such that (20) holds for those k<=k0. Same for rho. You don't need k_0 to be large enough to ensure (42). Instead, you can just consider a bigger rho that can bound |A(x_k)| for k <=k0.

7. There is a typo on line 630 in the left column. It should be \partial g(x_k) not \partial g(x_{k-1})

8. This is my biggest concern. I realized the entire proof of Theorem 4.1 works when sigma_k=0 for all k. In other words, you can just use the same y_k=y0 in each iteration and the algorithm still find an epsilon first and second order stationary point. In fact, (39) holds with a fixed y_k and then y_max will be trivially finite so that you can achieve (46) easily. The rest of the proof of Thm 4.1 will holds with sigma_k=0. This is too good to be true. Dual variable y_k is the key for the ALM methods to ensure Ax=0. If what I said is correct, it means that, with (20), your iterate x_k can always converge to Ax=0 by using a same y_k in each iteration. This may suggest that (20) is too restricted.

UPDATE:
I would like to comment on (20) in addition. If we apply the proposed method to the problem: min 0 subject to A(x)=0. The assumption (20) essentially says the stationary point of ||A(x)||^2 is a feasible point (i.e. A(x)=0). In fact, let h(x):=||A(x)||^2. With g=f=0, their assumption (20) becomes vh(x)<=|| \nabla h ||^2, which is the Polyak-Lojasiewicz Inequality (see (3) in https://arxiv.org/pdf/1608.04636.pdf). If the authors can explain it in this way, this will be an interesting insight.

Even if we accept Assumption (20), I still think this paper has some error in the technique proof. Let's take a look at page 11 in the supplementary file. The author prove (39) using the fact that the gradient of f is bounded by lambda_f. This is only true when x_k is in the ball with a radius rho by the definition of lambda_f. However, I don't think we can guarantee x_k is in the ball by making some assumptions. In Step 3 of the algorithm, the authors said "If necessary, project x_k to the ball". However, this does not help because x_k in (39) is solved from Step 2 using another first-order algorithm which does not guarantee x_k to be in the ball. When Step 3 projects x_k back to the ball, it is already too late. The x_k after the projection is no longer the x_k the authors

need in (39). Therefore, I think the proof is wrong unless you really assume the output x_k from Step 2 is bounded by rho, which I think is too strong.

I am satisfied with the explanation by the authors on why the convergence proof holds if I fix y_0=y_1=y_2,.... to be any constant. As beta increase to infinity, this will become penalty method.

2. Please provide an overall score for the submission.
Weak Reject: Borderline, tending to reject

3. Please enter a 2-3 sentence summary of your review explaining your overall score.
The main proof in this paper is not well organized which make their results hard to understand. The result in this paper is strong if it is correct but I found several places which seem counter intuitive. I appreciate the authors explaining more.

5. Please rate your confidence in the score assigned.
High: Reviewer has understood the main arguments in the paper, and has made high level checks of the proofs.

8. I agree to keep the paper and code submissions confidential, and delete any submitted code at the end of the review cycle to comply with the confidentiality requirement.
Agreement accepted

Reviewer #3

Questions

1. Please enter a detailed review describing the strengths and weaknesses of the submission.
This paper studies a variant of ADMM algorithm (iALM)for a family of nonconvex optimization problems, with nonlinear equality constraint.

This paper is clear written. The proposed algorithm iterates the following 2 parts:
a) inexact solution of the primal problem
b) dual ascent using an updated step size

The main differences with standard ADMM are
i) the dual step size is not equal to the Lagrangian augment parameter (beta), to ensure the dual variables are bounded
ii) the Lagrangian augment parameter (beta) is increasing, in order to enforce the constraints.

The authors did convergence analysis of the algorithm. I'm a bit confused by the results obtained in Section 4 & 5 though, and would like to discuss with the authors to help me understand them.

1. I'm confused by the assumption (20) in Thm 4.1. Are you denoting nu = min(dist / norm(A(xk))) for k in K, or you have a fixed nu and assume the nonlinear operator A and function g satisfy the condition? For the latter, note that {xk} are iterates generated by your algorithm, and I don't think you can have this type of assumption, except you make it a restricted condition -- for all x in a certain subset, you have this assumption and you keep the iterates contained in this set.

The results in Thm 4.1 states that if you can approximately solve the primal problem, up to certain accuracy, because of the assumption (20), you can obtain a good dual solution too and that's what we want. So, in my understanding that (20) is the key to your result so that I hope this can be clarified.

2. The constant nu depends on the interval K. This looks weird to me, can you set k1 = infty?

3. From the results you obtained, the rate of {beta_k} increases determines the convergence rate of your algorithm. Have you conducted synthetic experiments to show that they match? How do you set them in your experiments?

====
After reading the author rebuttal, I still think that the assumption defined by (20) needs further justification. I'll keep my rating.

2. Please provide an overall score for the submission.
Reject: Clearly below the acceptance threshold

3. Please enter a 2-3 sentence summary of your review explaining your overall score.
I am confused by the theoretical results presented in this paper and would like them to be clarified.

5. Please rate your confidence in the score assigned.
Medium: Reviewer has understood the main points in the paper, but skipped the proofs and technical details.

8. I agree to keep the paper and code submissions confidential, and delete any submitted code at the end of the review cycle to comply with the confidentiality requirement.
Agreement accepted

Reviewer #4

Questions

1. Please enter a detailed review describing the strengths and weaknesses of the submission.
Summary:
This paper analyzes convergence of the inexact augmented Lagrangian algorithm (iALM) on non-convex non-smooth objective functions with non-linear non-convex equality constraints. Even though ALM was proposed many decades ago, I believe we still don't have full understanding of its non-asymptotic convergence properties even for convex problem. Thus this paper tackles an important problem: convergence analysis of iALM for non-convex problems. It analyzes the minimization of f(x) + g(x) subject to A(x) = 0. All functions are non-convex and g(x) is non-smooth.

Strengths:
1. Authors identified a sufficient condition [Eq. (20)] on A(x) and g(x), which allows for the similar analysis as the convex smooth setting of iALM goes through for this non-convex case as well.
3. The paper gives convergence rate of O(1/\epsilon^3) to first order-stationary point when g(x) is an indicator function of convex compact set.
4. Paper gives convergence of rate of O(1/\epsilon^5) to second order-stationary point when g(x) = 0.
5. The experimental settings are well thought of and useful. For some problems, the non-convex algorithm performs much better than the convex baselines.
6. The related works section covers many relevant works and is comprehensive.
7. This is one of first work which analyze the non-asymptotic convergence rate of iALM on any non-convex problem

Questions and suggestions:
1. It would better if the authors can explain through simple examples what the sufficient condition Eqn. (20) means. I appreciate the author comparing the condition to the Slater's condition (page 4), but I think it would be better if they can show the sufficient condition in action in simpler problems at the same place on the manuscript.

2. The sufficient condition Eqn. (20) is fails even for very simple constraints. For example, if g(x) is the indicator function for convex set C = {(x, y) : x^2 + y^2 \leq 1} and A(x,y) = x-y = 0. The condition fails at the point

(1/\sqrt{2}, -1/\sqrt{2}). Note, that Slater's condition is satisfied here [for \min_{x \in C} f(x), subject to A(x,y) = 0], but analyses fails if ALM goes near the point (1/\sqrt{2}, -1/\sqrt{2}). Thus the condition is much stronger than the Slater's condition. Note that enforcing that ||A(x)|| is small does not seem like a good solution for this issue and the algorithm cannot explicitly ensure this.

In the various analysis of convex iALM, are there similar conditions as Eqn (20)? Is it possible to prove, that this sufficient condition is also a necessary condition?

3. The choice of K = \log_b(Q/\epsilon) is very similar to that of penalty method. Would it be possible to compare this iALM to penalty method in the experiments?

4. In page 5 (below eqn (26)), homotopy approach of Algorithm 1 is metioned. But the term "homotopy" is never defined in the paper. Could the authors clarify what is homotopy w.r.t. to Algorithm 1?

5. In the experimental sections, it if mentioned that the sufficient condition is met y ensuring
a) x_k is not too small (Clustering expt.). I couldn't understand why this is reasonable approach, especially since \nu in Eqn (20) is considered as constant.
b) that x_k doesn't go near the boundary of convex set represented by g(x) (Basis pursuit expt.), which seems like a very strong modification of the algorithm.

Could the author discuss the performance vs convergence rate trade-off when we enforce these conditions? In the experiments, how are these conditions enforced? Further could the authors discuss how their guarantees change when we enforce this conditions?

6. In the experiments of Basis Pursuit (6.2), would it be possible to compare iALM to the standard baselines of this problem? In Figure 3, I can only see different variations of Algorithm 1.

Minor errors and typos:
Paper is well written and does not have many typos or notational inconsistencies.

==========
UPDATE
I thank the authors for their feedback and answers.
2. I am not convinced that the sufficient condition is reasonable.
6. If it is already done, is it provided? I don't see it. If not, please provide those experimental results too for fair comparison.

Further as mentioned by another reviewer, I am not sure how we can get (39 of supplementary), if x_k need not be inside the ball..

2. Please provide an overall score for the submission.
Weak Reject: Borderline, tending to reject

3. Please enter a 2-3 sentence summary of your review explaining your overall score.
Author analyze an important algorithm for solving non-convex optimization problems and give convergence guarantees for simple variations of it. The experimental results are promising. But I could not fully grasp the implications of the sufficient condition Eqn (20) [which seem to fail for even simple cases] and how the experiments ensure it.

5. Please rate your confidence in the score assigned.
High: Reviewer has understood the main arguments in the paper, and has made high level checks of the proofs.

8. I agree to keep the paper and code submissions confidential, and delete any submitted code at the end of the review cycle to comply with the confidentiality requirement.
Agreement accepted