Solving a Class of Nonsmooth Nonconvex Optimization Problems Via Proximal Alternating Linearization Scheme with Inexact Information

Ming Huang [1, a], Yue He [1, b, *], Pingping Qiao [1, c], Siqi Zhang [1, d], and Yongxiu Feng [1, e]

[1] School of Science, Dalian Maritime University, Dalian, 116026, China.

[a] huangming0224@163.com, [b, *] yuehe_zstu@163.com, [c] wo820360592@163.com,

[d] zsq0216@dlmu.edu.cn, [e] fengyongxiu2001@163.com

Abstract. For optimization problems minimizing the sum of two nonconvex and nonsmooth functions, we propose an alternate linearization method with inexact data. In many practical optimization applications, only the inexact information of the function can be obtained. The core idea of this method is to add a quadratic function term to the nonconvex function(called local convexification of nonconvex function), and then to construct an approximate proximal point model. In each iteration, a series of iteration points are obtained by solving subproblems alternately. It can be proved that, in the sense of inexact oracles, these iteration points converge to the stable point of the original problem, and theoretically show that the algorithm has good convergent properties.

Keywords: proximal alternating linearization; nonsmooth optimization; nonconvex optimization; inexact oracles; lower-C^2 function.

1. Introduction

In this paper, we consider an unconstrained nonsmooth optimization problem in the following form:

$$(P) \qquad \min_{y\in\mathbb{R}^n} \Phi(y) := \phi(y) + \psi(y) \quad , \tag{1}$$

where the functions $\phi:\mathbb{R}^n \to \mathbb{R}$ and $\psi:\mathbb{R}^n \to \mathbb{R}$ are prox-regular Lipschitz, not necessarily convex or smooth. Suppose that we can compute the proximal point of ψ and ϕ more easily or less costly than to solve the problem (P) directly.

We utilize the new notion of computing the proximal point algorithm of nonconvex functions, especially the prox-bounded lower-C^2, and design the alternating linearization scheme with inexact information for working out the sum of nonconvex nonsmooth functions. The important step is to build up the approximate proximal point model, which is mainly split into the following three steps:

1. according to the property of the lower-C^2 function, add the corresponding proper proximal terms whose center is the current iteration point to convexify the function ψ and ϕ, respectively;

2. keep one convexified function fixed and make the other convexified one linearized at the iteration point;

3. the oracle outputs an approximative function value at each iteration point, and the error of function values changes with iteration.

On account of the property of the lower-C^2 function, combined with the feature of proximal point mapping, the executable inexact alternate linearization approach, which decomposes the original problem into two different convex approximate subproblems, is proposed. For the sake of getting over the difficulty of obtaining the exact information about the function values and gradients in the practical optimization problem, we use the values of the function with certain errors in the construction of the approximate model. In each iteration, the proximal point of the approximate model needs to be calculated alternately. Moreover, the parameters and errors are updated in the light of explicit measure. We also proved that the method owns great properties of convergence from a theoretical point of view.

In essence, we propose a new approximate proximal point approach that adds inexact data into the alternating linearization algorithm. This can be considered as extending the algorithm, initially only applicable to convex optimization, to nonconvex case. As a result, this illustrates that our proposed algorithm is more widespread. On the other hand, traditional linearization constructions or cutting plane models are straightly formed from the values of the function and subgradients at a given point, where the exact oracle can obtain. However, in the nonconvex case, the linearization error of these schemes may be negative and even cut off some regions containing local optimal solutions, so it is challenging to guarantee convergence of the approach. But we can eliminate the negative linearization error of the approximate model by locally convexifying the original problem, respectively, which is another strength of our method.

The framework of the paper is as follows. In Section 2, we states the core idea and concrete form of the approximate model construction in the nonconvex case and presents an executable inexact alternating linearization approach. In Section 3, the convergence analysis of the algorithm is provided to validate the effectiveness of the algorithm. In Section 4, we our present conclusions.

2. A proximal alternating linearization approach with inexact information

In this section, we give an implementable approach to solve the problem (P) and state the major idea of forming approximate proximal points in the nonconvex setting.

We assume that for an accuracy tolerance $\varepsilon_k > 0$, at each point $y^k \in \mathbb{R}^n$ the oracle delivers an approximate value $\hat{\psi}(y^k)$ of f. This ensures that the inexact and exact function values meet the following relationship:

$$\hat{\psi}(y^k) \in \left[\psi(y^k) - \varepsilon_k, \psi(y^k)\right]. \tag{2}$$

Similarly, approximate value $\hat{\phi}(y^k)$ satisfies:

$$\hat{\phi}(y^k) \in \left[\phi(y^k) - \varepsilon_k, \phi(y^k)\right]. \tag{3}$$

For simplicity, we denote

$$u_\phi^k = \arg\min_y \left\{ \pi_\phi^k(y) + \frac{1}{2}\mu_k \left\| y - y^k \right\|^2 \right\},\quad u_\psi^k = \arg\min_y \left\{ \pi_\psi^k(y) + \frac{1}{2}\mu_{k+1} \left\| y - y^{k+1} \right\|^2 \right\}.$$

Considering the k th iteration point as y_k and using the equivalent form of lower-C^2 functions, two lower approximations for the function $\Phi(y) + \frac{1}{2}\lambda \left\| y - y^k \right\|^2$ can be built as follows:

$$\pi_\phi^k(y) = \phi(y) + \frac{1}{2}\lambda_\phi^k \left\| y - y^k \right\|^2 + \hat{\psi}(u_\psi^{k-1}) + \frac{1}{2}\lambda_\psi^{k-1} \left\| u_\psi^{k-1} - y^k \right\|^2 + \left\langle g_\psi^{k-1} + \lambda_\psi^{k-1}(u_\psi^{k-1} - y^k), y - u_\psi^{k-1} \right\rangle,$$

$$\pi_\psi^k(y) = \hat{\phi}(u_\phi^k) + \frac{1}{2}\lambda_\phi^k \left\| u_\phi^k - y^{k+1} \right\|^2 + \left\langle g_\phi^k + \lambda_\phi^k(u_\phi^k - y^{k+1}), y - u_\phi^k \right\rangle + \psi(y) + \frac{1}{2}\lambda_\psi^k \left\| y - y^{k+1} \right\|^2,$$

where $g_\psi^{k-1} \in \partial\psi(u_\psi^{k-1})$ and $g_\phi^k \in \partial\phi(u_\phi^k)$.

Obviously, the two approximate models discussed above stand for the sum of a convex function and an affine function. Therefore, based on the property, we can prove that when the parameters meet conditions, the constructed model can satisfy the properties of (a)-(d) in [1], which plays an important role in our subsequent proof of the algorithms convergence.

Lemma 2.1. In the context of $\lambda_\psi^{k-1} \ge \bar{\lambda}_\psi > 0$, $\lambda_\phi^k \ge \bar{\lambda}_\phi > 0$ and $\mu_k \ge \mu_{\min}$, the proximal point of π_ϕ^k at y^k is indicated by the sequence $\left\{u_\phi^k\right\}$. Furthermore, the following item hold:

(i) π_ϕ^k is a convex function, with the Lipschitz continuity on $\bar{B}_\varepsilon(y^k)$;

(ii) $\pi_\phi^k\left(y^k\right) \le \Phi\left(y^k\right)$;

(iii) $\pi_\phi^k(y) \ge \pi_\phi^k(u_\phi^k) + \mu_k \left\langle y^k - u_\phi^k, y - u_\phi^k \right\rangle, \forall y$;

Proof. These results resemble Lemma 2.4 in [2] and are easily proved similar to it.

In the algorithm, we define the following linearization error:

$$e_k = \Phi(y^k) - \pi_\phi^k(u_\phi^k).$$

The linearization error is made sure of nonnegativity by the convexity of the approximate model π_ϕ^k, which is proved the convergence of our algorithm significantly. Besides, our elementary work is to prove that $e_k = 0$ is the criterion of implementable stopping. In addition, prior to the steps of the algorithm being listed, we make assumptions about the constants that can be achieved in the algorithm.

Assumption 2.1. The parameters $\mu_k > 0$, $\lambda_\psi^k > 0$ in the algorithm satisfy $\mu_k = \mu$, $\lambda_\psi^k = \bar{\lambda}_\psi$, $\bar{\lambda}_\phi^k = \bar{\lambda}_\phi$ when k is large enough. Let $\rho_k = \lambda_\psi^k + \lambda_\phi^k + \mu_k$ and denote $\rho_k = \bar{\rho}$ after a sufficient large k.

Algorithm 1: Proximal Alternating Linearization Algorithm with Inexact Oracles

Step 0 (Initialization) Choose a starting point $u_\psi^0 \in \mathbb{R}^n$, $g_\psi^0 \in \partial f(u_\psi^0)$, an initial error $\varepsilon_0 \ge 0$. Choose parameters $\mu_1 \ge \mu_{\min} > 0$, $\beta_1 \in (0,1)$, $\kappa > 1$, $\lambda_\psi^0 > 0$, $\lambda_\phi^1 > 0$ and let $y^1 = u_\psi^0$. Set $k = 1$.

Step 1 (Solving the ϕ-subproblem) Call the oracle (2) at u_ψ^{k-1} to obtain $\hat{\psi}(u_\psi^{k-1})$ and compute

$$u_\phi^k = \arg\min_y \left\{ \pi_\phi^k(y) + \frac{1}{2}\mu_k \left\| y - y^k \right\|^2 \right\},$$

then let $g_\phi^k = -g_\psi^{k-1} - \lambda_\psi^{k-1}(u_\psi^{k-1} - y^k) - (\lambda_\phi^k + \mu_k)(u_\phi^k - y^k)$.

Step 2 (Stopping criterion) If $e_k = 0$, then STOP!

Step 3 (Descent test and constants update) If

$$\Phi(u_\phi^k) \le \Phi(y^k) - \beta_1 e_k,$$

set $y^{k+1} = u_\phi^k$ (descent step) and choose $\mu_{k+1} \in \left[\max\left\{\mu_{\min}, \frac{\mu_k}{\kappa}\right\}, \mu_k\right]$; otherwise set $y^{k+1} = y^k$ (null step) and choose $\mu_{k+1} = \mu_k$. If $\lambda_\psi^{k-1} = \lambda_\psi^{k-1}$, otherwise $\lambda_\psi^k = \kappa \bar{\lambda}_\psi^k$.

Step 4 (Error updating) Let $\varepsilon_k = \gamma \varepsilon_{k-1}$, $\gamma \in (0,1)$.

Step 5 (Solving the ψ-subproblem) Call the oracle (3) at u_ϕ^k to obtain $\hat{\phi}(u_\phi^k)$ and compute

$$u_\psi^k = \arg\min_y \left\{ \pi_\psi^k(y) + \frac{1}{2}\mu_{k+1} \left\| y - y^{k+1} \right\|^2 \right\},$$

then let $g_\psi^k = -g_\phi^k - \lambda_\phi^k(u_\phi^k - y^{k+1}) - (\lambda_\psi^k + \mu_{k+1})(u_\psi^k - y^{k+1})$. If $\lambda_\phi^k \ge \tilde{\lambda}_\phi^k$, set $\lambda_\phi^{k+1} = \lambda_\phi^k$, otherwise $\lambda_\phi^k = \kappa \tilde{\lambda}_\phi^k$.

Step 6 (Loop) Set $k = k + 1$ and go to **Step 1**.

Remark 1. (i) According to the first-order optimality condition of ψ -subproblem, we can get the following result straightly

$$0 \in \partial\left(\pi_{\psi}^{k}(y)+\frac{1}{2}\mu_{k+1}\left\|y-y^{k+1}\right\|^{2}\right)\left(u_{\psi}^{k}\right),$$

and then $g_{\psi}^{k} \in \partial\psi(u_{\psi}^{k})$ can be directly verified. Similarly, $g_{\phi}^{k} \in \partial\phi(u_{\phi}^{k})$.

(ii) Via the update rule in Step 3, the proximal parameter $\{\mu_k\}$ does not increase whenever the null step or descent step occurs and satisfies that $\mu_{\min} \le \mu_{k+1} \le \mu_k$. Moreover, we can obtain μ_k is unchanged as $k \to \infty$.

3. Convergence analysis

In every iteration of Algorithm 1, we need to solve two subproblems on approximate proximal point mapping. Based on the items (i)-(iv) of Lemma 2.1, it could be proved that the series of the approximate point $\{u_{\phi}^{k}\}$ must be bounded and convergent under the condition of finitely many descent steps.

In light of the property of prox-bounded function with the form of lower- C^2 function, we prove that the sequence of iterates $\{y^k\}$ converges towards the proximal point of Φ in the nonconvex background with finite serious steps.

Theorem 3.1. Let $q = \arg\min_{y}\left\{\Phi(y)+\frac{1}{2}\bar{\rho}\left\|y-\bar{y}\right\|^{2}\right\}$. Suppose $\{u_{\phi}^{k}\}$ and $\{y^k\}$ are produced by Algorithm 1, and the Assumption 3.1 is satisfied for the parameters $\mu_k > 0$, $\lambda_{\psi}^{k} > 0$, $\lambda_{\phi}^{k} > 0$. If there just exists finitely many serious steps and the last serious step is denoted as $\bar{y}$, then y^k converge to $\bar{y}$ as $k \to \infty$ and $\bar{y}$ is just a stationary point of Φ.

Next, in the case of serious steps occurring infinitely, it could be showed that every cluster point of the series $\{y^k\}$ is stationary for Φ.

Theorem 3.2. Assume the sequences $\{u_k\}$ and $\{y^k\}$ are respectively generated by Algorithm 1 and Assumption 2.1 is satisfied for the parameters $\mu_k > 0$, $\lambda_{\psi}^{k} > 0$, $\lambda_{\phi}^{k} > 0$. Denote $L = \{k \mid y^{k+1} \ne y^k\}$ and when L is infinite, then every cluster point of $\{y^k\}$ is stationary for Φ.

4. Conclusions and future work

For the optimization problem of the sum of two non-convex and non-smooth functions, we propose an executable alternate linearization algorithm with inexact oracles, which decomposes the original problem into two different convex approximate subproblems. In each iteration, the proximal point of the approximate model needs to be calculated alternately. At the same time, the parameters and errors are updated according to specific criteria. Under appropriate conditions, it is also showed that the algorithm has good convergence properties in theory. In essence, we propose a new approximate proximal point method, which is suitable for solving more general non-smooth non-convex optimization problems. Hence, our proposed algorithm is more universal.

The convergence analysis shows that this algorithm effectively solves non-convex and nonsmooth minimization problems. However, there are still some problems that need further study:

1. Can alternate linearization algorithms with inexact data be applied to other decision problems?

2. How to select appropriate parameters to improve the efficiency of our algorithm?

3. We only consider the class of functions as lower- C^2 are other non-convex functions applied to our scheme?

These questions are still worth exploring and researching further.

References

[1] W. Hare, C. Sagastizbal. A redistributed proximal bundle method for nonconvex optimization. SIAM Journal on Optimization, 2010, 20(5): 2442-2473.

[2] D. Li, L. P. Pang, S. Chen. A proximal alternating linearization method for nonconvex optimization problem. Optimization Methods and Software, 2014, 29(4): 771-785.

[3] W. Hare, C. Sagastizbal. Computing proximal points of nonconvex functions. Mathematical Programming. 2009, 116(2): 221-258.

[4] H. Mine, M. Fukushima. A minimization method for the sum of a convex function and a continuously differentiable function. Journal of Optimization Theory and Applications, 1981, 33(1): 9-23.

[5] S. G. Mallat, Z. Zhang. Matching pursuits with time-frequency dictionarie. IEEE Transactions, 1993, 41(12): 3397-3415.

[6] B. K. Natarajan. Sparse approximate solutions to linear systems. SIAM Journal on Computing, 1995, 24(2): 227-234.

[7] H. Zou. The adaptive lasso and its oracle properties. Journal of the American Statistical Association, 2006, 101(476): 1418-1429.

[8] T. Paul. Applications of a splitting algorithm to decomposition in convex programming and variational inequalities. SIAM Journal on Control and Optimization, 2006, 29(1): 119-138.

[9] E. Jonathan. Some Saddle-function splitting methods for convex programming. Optimization Methods and Software, 1994, 4(1): 75-83.

[10] L. F. Yang, J. Y. Luo, Y. Xu, Z. R. Zhang and Z.Y. Dong. A distributed dual consensus ADMM based on partition for DC-DOPF with carbon emission trading. IEEE Transactions on Industrial Informatics, 2020, 16(3): 1858-1872.