Hindawi
Mathematical Problems in Engineering
Volume 2021, Article ID 5629868, 5 pages
https://doi.org/10.1155/2021/5629868

Research Article

Convergence Rate Analysis of the Proximal Difference of the Convex Algorithm

Xueyong Wang,[1] Ying Zhang,[2] Haibin Chen ,[1] and Xipeng Kou[3]

[1]*School of Management Science, Qufu Normal University, Rizhao, Shandong 276800, China*
[2]*School Basic Teaching, Shandong Water Conservancy Vocat Coll, Rizhao, Shandong 276800, China*
[3]*College of Mathematics Physics and Data Science, Chongqing University of Science and Technology, Chongqing 401331, China*

Correspondence should be addressed to Haibin Chen; chenhaibin508@qfnu.edu.cn

Received 20 July 2020; Accepted 16 December 2020; Published 4 January 2021

Academic Editor: Chuanjun Chen

In this paper, we study the convergence rate of the proximal difference of the convex algorithm for the problem with a strong convex function and two convex functions. By making full use of the special structure of the difference of convex decomposition, we prove that the convergence rate of the proximal difference of the convex algorithm is linear, which is measured by the objective function value.

1. Introduction

Difference of convex programming (DCP) is a kind of important optimization problem that the objective function can be written as the difference of convex (DC) functions. The DCP problem has found many applications in assignment and power allocation [1], digital communication system [2], compressed sensing [3], and so on [4–6].

Up to now, one of the classical algorithms for DCP is the DC algorithm (DCA) [7] in which the nonconvex part of the objective function is replaced by a linear approximation. By DCA, only a convex optimization subproblem needs to be solved at each iteration. After that, the DCA has been attracted by a lot of researchers. Le Thi et al. [8] proved the linear convergence rate of DCA by employing the Kurdyka–Lojasiewicz inequality. Assuming that the subproblem of DCA can be easily solved [6], Gotoh et al. [4] proposed the proximal DC algorithm (PDCA) for solving the DCP, in which not only the nonconvex part in the objective function is replaced by the same technique as in DCA but also the convex part is replaced by a quadratic approximal. The PDCA reduces to the classical proximal gradient algorithm for convex programming if the nonconvex part of the objective function is void [9]. To accelerate the PDCA, Wen et al. [10] introduced a new type of proximal algorithm ($PDCA_e$) with the help of an extrapolation technique. Since the convergence rate of $PDCA_e$ heavily depends on the Kurdyka–Lojasiewicz inequality, $PDCA_e$ converges linearly in general [10].

In this paper, we study the linear convergence rate of PDCA by the structure, which is different from the techniques in [8, 10]. Under conditions that the objection function can be divided into difference of a strong convex function and two convex functions with Lipschitz continuous gradient, we prove the linear convergence rate of PDCA, which is measured by the objective function value.

The remainder of the paper is organized as follows. In Section 2, several useful preliminaries are recalled. In Section 3, more details about the DC optimization problem are given, and the PDCA proposed in [4] is listed for the sake of simplicity. The linear convergence rate of the PDCA is established in Section 4. Final remarks are given in Section 5.

2. Preliminaries

In this section, we recall some useful definitions and properties.

Let $f\colon R^n \longrightarrow [-\infty, +\infty]$ be an extended real function. The domain of f is denoted by

$$\text{dom } f = \{x \in R^n\colon\ f(x) < +\infty\}. \tag{1}$$

If $f(x)$ never equals $-\infty$ for all $x \in \text{dom } f$ and dom $f \neq \emptyset$, we say that f is a proper function. If the proper function is lower semicontinuous, then it is called a closed function. A proper closed function $f(x)$ is said to be level bounded if the lower level set of f (i.e., $\{x \in R^n | f t(nxq)h \leq_r x, 7 C\forall; r \in R\}$) is bounded.

Let $f\colon R^n \longrightarrow R \bigcup \{+\infty\}$ be a proper closed function. Then, the limit subdifferential of f at $x \in \text{dom} f$ is defined as follows:

$$\partial f(x) = \left\{ v \in R^n\colon \exists x^t \xrightarrow{f} x, v^t \longrightarrow v \text{ with } \liminf_{y \longrightarrow x^t} \frac{f(y) - f(x^t) - \langle v^t, y - x^t \rangle}{\|y - x^t\|} \geq 0, \quad \forall t \right\}, \tag{2}$$

where $z \xrightarrow{f} x$ denote $z \longrightarrow x$ and $f(z) \longrightarrow f(x)$. Note that dom $\partial f = \{x \in R^n\colon \partial f(x) \neq \emptyset\}$. It is well known that the limit subdifferential reduces to the classical subdifferential in convex analysis when $f(x)$ is a convex function, that is,

$$\partial f(x) = \{v \in R^n\colon\ f(u) - f(x) - \langle v, u - x \rangle \geq 0, \quad \forall u \in R^n\}. \tag{3}$$

Furthermore, if f is continuously differentiable, then the limit subdifferential reduces to the gradient of f denoted by ∇f.

3. DC Programming and PDCA

In this section, we begin to consider the DC programming problem:

$$\min_{x \in R^n} \{F(x) := f(x) + g(x) - h(x)\}, \tag{4}$$

where $f\colon R^n \longrightarrow R$ is a strong convex function with constant $a > 0$ and $g, h\colon R^n \longrightarrow R$ are convex functions, and their gradients are Lipschitz continuous with constants $L_g > 0$ and $L_h > 0$, respectively. Throughout the paper, we assumed that $F(x)$ is level bounded and $a > 1$. Apparently, (4) is a DC optimization problem and can be solved by the following DCA Algorithm 1.

Although the subproblem (Algorithm 2) is convex, it may not have closed solutions. To solve this drawback, Gotoh et al. proposed the following PDCA.

4. The Convergence Rate of PDCA

In this section, we give the linear convergence rate of PDCA. To continue, the following lemma is useful.

Lemma 1. *Let $f\colon R^n \longrightarrow R$ be a continuous differentiable function with Lipschitz continuous gradient with Lipschitz constant $L > 0$. Then, for any $L' > L$, it holds that*

$$f(x) \leq f(y) + \langle \nabla f(y), x - y \rangle + \frac{L'}{2}\|x - y\|^2, \quad \forall x, y \in R^n. \tag{5}$$

By Lemma 1, we have the following result.

Lemma 2. *Let $\{x_k\}$ be generated in Algorithm 2. Then,*

$$\mu(F(x) - F(x_{k+1})) \geq \|x - x_{k+1}\|^2 - \|x - x_k\|^2. \tag{6}$$

Proof. Since f is strongly convex with parameter $a > 0$, it holds that

$$f(x) \geq f(x_{k+1}) + \langle \xi_{k+1}, x - x_{k+1} \rangle + \frac{a}{2\mu}\|x - x_{k+1}\|^2, \tag{7}$$

where $\xi_{k+1} \in \partial f(x_{k+1})$.

Since $\nabla h(x)$ is Lipschitz continuous with constant $L_h > 0$, by (5), there exists $0 < \mu \leq 1/L_h$ such that

$$h(x) \leq h(x_x) + \langle \nabla h(x_k), x - x_k \rangle + \frac{1}{2\mu}\|x - x_k\|^2, \tag{8}$$

that is,

$$-h(x) \geq -h(x_k) - \langle \nabla h(x_k), x - x_k \rangle - \frac{1}{2\mu}\|x - x_k\|^2. \tag{9}$$

Since g is a convex function, we have

$$g(x) \geq g(x_k) + \langle \nabla g(x_k), x - x_k \rangle. \tag{10}$$

Summing (7), (9), and (10), we get

$$\begin{aligned} f(x) + g(x) - h(x) \geq f(x_{k+1}) + g(x_k) - h(x_k) + \langle \xi_{k+1}, x - x_{k+1} \rangle + \\ \langle \nabla g(x_k) - \nabla h(x_k), x - x_k \rangle + \frac{a}{2\mu}\|x - x_{k+1}\|^2 - \frac{1}{2\mu}\|x - x_k\|^2. \end{aligned} \tag{11}$$

On the contrary, since h is a convex function, we have

$$h(x) \geq h(x_k) + \langle \nabla h(x_k), x - x_k \rangle, \tag{12}$$

which is equivalent to the following form:

$$-h(x) \leq -h(x_k) - \langle \nabla h(x_k), x - x_k \rangle. \tag{13}$$

Since $\nabla g(x)$ is Lipschitz continuous with constant $L_g > 0$, by (5), there exists $0 < \mu \leq 1/L_g$ such that

$$g(x) \leq g(x_k) + \langle \nabla g(x_k), x - x_k \rangle + \frac{1}{2\mu}\|x - x_k\|^2. \tag{14}$$

(1) Initial step: choose $\varepsilon > 0$ and $x_0 \in R^n$, and set $k = 0$.
(2) Iterative step: compute the new point by the following formula:
(3) $x_{k+1} = \arg\min_{x\in R^n}\{f(x) + g(x) - h(x_k) - \langle \nabla h(x_k), x - x_k\rangle\}$,
(4) **until** $\|x_{k+1} - x_k\| \leq \varepsilon$ is satisfied.

Algorithm 1: DCA for problem (4).

Summing (13) and (14), we have

$$g(x) - h(x) \leq g(x_k) - h(x_k) + \langle \nabla g(x_k) - \nabla h(x_k), x - x_k\rangle + \frac{1}{2\mu}\|x - x_k\|^2. \tag{15}$$

Adding $f(x)$ on both sides of (15), we get

$$f(x) + g(x) - h(x) \leq f(x) + g(x_k) - h(x_k) + \langle \nabla g(x_k) - \nabla h(x_k), x - x_k\rangle + \frac{1}{2\mu}\|x - x_k\|^2. \tag{16}$$

Taking $x = x_{k+1}$, it follows that

$$f(x_{k+1}) + g(x_{k+1}) - h(x_{k+1}) \leq f(x_{k+1}) + g(x_k) - h(x_k) + \langle \nabla g(x_k) - \nabla h(x_k), x_{k+1} - x_k\rangle + \frac{1}{2\mu}\|x_{k+1} - x_k\|^2. \tag{17}$$

By optimality conditions of Algorithm 2, we know that

$$\xi_{k+1} + \nabla g(x_k) - \nabla h(x_k) + \frac{1}{\mu}(x_{k+1} - x_k) = 0, \tag{18}$$

where $\xi_{k+1} \in \partial f(x_{k+1})$, which means that

$$-\frac{1}{\mu}(x_{k+1} - x_k) = \xi_{k+1} + \nabla g(x_k) - \nabla h(x_k). \tag{19}$$

By (11) and (17), it holds that

$$\begin{aligned}
&F(x) - F(x_{k+1})\\
&\geq \langle \xi_{k+1}, x - x_{k+1}\rangle + \langle \nabla g(x_k) - \nabla h(x_k), x - x_{k+1}\rangle\\
&\quad + \frac{a}{2\mu}\|x - x_{k+1}\|^2 - \frac{1}{2\mu}\|x - x_k\|^2 - \frac{1}{2\mu}\|x_k - x_{k+1}\|^2\\
&= -\frac{1}{\mu}\langle x_{k+1} - x_k, x - x_{k+1}\rangle + \frac{a}{2\mu}\|x - x_{k+1}\|^2 - \frac{1}{2\mu}\|x - x_k\|^2 - \frac{1}{2\mu}\|x_k - x_{k+1}\|^2\\
&= \frac{1}{2\mu}\left(\|x_k - x_{k+1}\|^2 + \|x - x_{k+1}\|^2 - \|x - x_k\|^2\right) + \frac{a}{2\mu}\|x - x_{k+1}\|^2 - \frac{1}{2\mu}\|x - x_k\|^2 - \frac{1}{2\mu}\|x_k - x_{k+1}\|^2\\
&= \frac{1}{2\mu}\left((1 + a)\|x - x_{k+1}\|^2 - 2\|x - x_k\|^2\right)\\
&\geq \frac{1}{\mu}\left(\|x - x_{k+1}\|^2 - \|x - x_k\|^2\right),
\end{aligned} \tag{20}$$

where the first equality follows from (19) and the last inequality follows from $a > 1$. The desired result follows.

Now, we are at a position to prove the main theorem as follows. □

Theorem 1. *Let $\{x_k\}$ be generated in Algorithm 2. Then,*

$$F(x_k) - F(x^*) \leq \frac{\|x_0 - x^*\|^2}{\mu k}, \tag{21}$$

where x^* is the stationary point of (4).

Proof. By Lemma 2, let $x = x_k$, and we have that

(1) Initial step: choose $0 < \mu < 1/\max\{L_g, L_h\}$, $\varepsilon > 0$, and $x_0 \in R^n$, and set $k = 0$.
(2) Iterative step: compute the new point by the following formula:
(3) $x_{k+1} = \arg\min_{x\in R^n}\{f(x) + g(x_k) - h(x_k) - \langle \nabla g(x_k) - \nabla h(x_k), x - x_k\rangle + 1/2\mu\|x - x_k\|^2\}$,
(4) **until** $\|x_{k+1} - x_k\| \le \varepsilon$ is satisfied.

Algorithm 2:PDCA for problem (4).

$$\mu\left(F(x_k) - F(x_{k+1})\right) \ge \left\|x_{k+1} - x_k\right\|^2 \ge 0. \tag{22}$$

Then, it follows from $\mu > 0$ that $F(x_k) \ge F(x_{k+1})$, which means that the sequence $\{F(x_k)\}$ is nonincreasing. Then, for any $k_0 \in N$, it follows that

$$\sum_{k=0}^{k_0-1} F(x_{k+1}) \ge \sum_{k=0}^{k_0-1} F\left(x_{k_0}\right) = k_0 F\left(x_{k_0}\right). \tag{23}$$

By Lemma 2 again, let $x = x^*$, and we have that

$$\mu\left(F(x^*) - F(x_{k+1})\right) \ge \left\|x_{k+1} - x^*\right\|^2 - \left\|x_k - x^*\right\|^2, \tag{24}$$

which implies that

$$\begin{aligned}\mu\left(k_0 F(x^*) - \sum_{k=0}^{k_0-1} F(x_{k+1})\right) &\ge \left\|x_{k_0} - x^*\right\|^2 - \left\|x_0 - x^*\right\|^2 \\ &\ge -\left\|x_0 - x^*\right\|^2.\end{aligned} \tag{25}$$

By (23) and (25), it yields that

$$\mu k_0\left(F(x^*) - F\left(x_{k_0}\right)\right) \ge -\left\|x^* - x^0\right\|^2, \tag{26}$$

and the desired result follows. □

5. Conclusions

In this paper, we give the linear convergence rate of PDCA for the case that the objective function is divided into a strong convex function and two convex functions. Different from the method in [8, 10], which depends heavily on the Kurdyka–Lojasiewicz inequality, we give a simple proof by the special structure of the optimization problem. Actually, there may be some other potential applications about the proposed PDCA. We leave this work in the future. For example, we will study further applications of the PDCA algorithm to some nonconvex problems [11, 12], tensor optimization problems [13, 14], and so on [15–18].

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Authors' Contributions

Each author contributed equally to this paper and read and approved the final manuscript.

Acknowledgments

This project was supported by the National Natural Science Foundation of China (Grant nos. 11801309 and 12071249) and the Science and Technology Research Program of Chongqing Municipal Education Commission (Grant no. KJ1713334).

References

[1] M. Sanjabi, M. Razaviyayn, and Z.-Q. Luo, "Optimal joint base station assignment and beamforming for heterogeneous networks," *IEEE Transactions on Signal Processing*, vol. 62, no. 8, pp. 1950–1961, 2014.
[2] A. Alvarado, G. Scutari, and J.-S. Pang, "A new decomposition method for multiuser DC-programming and its applications," *IEEE Transactions on Signal Processing*, vol. 62, no. 11, pp. 2984–2998, 2014.
[3] P. Yin, Y. Lou, Q. He, and J. Xin, "Minimization of $l_1 - l_2$ for compressed sensing," *SIAM Journal on Scientific Computing*, vol. 37, pp. 536–563, 2015.
[4] J.-y. Gotoh, A. Takeda, and K. Tono, "DC formulations and algorithms for sparse optimization problems," *Mathematical Programming*, vol. 169, no. 1, pp. 141–176, 2018.
[5] W. D. Oliveira, "Proximal bundle methods for nonsmooth DC programming," *J Global Optim*, vol. 75, pp. 523–563, 2019.
[6] D. T. Pham and H. A. Le Thi, "ADC Optimization algorithm for solving the trust-region subproblem," *SIAM Journal on Optimization*, vol. 8, pp. 476–505, 1998.
[7] D. T. Pham and H. A. Le Thi, "Convex analysis approach to DC programming:theory, algorithms and applications," *Acta Math Vietnam*, vol. 22, pp. 289–355, 1997.
[8] H. A. Le Thi, V. N. Huynh, and T. Pham Dinh, "Convergence analysis of difference-of-convex algorithm with subanalytic data," *Journal of Optimization Theory and Applications*, vol. 179, no. 1, pp. 103–126, 2018.
[9] B. Donoghue and E. J. Candes, "Adaptive restart for accelerated gradient schemes," *Foundations of Computational Mathematics*, vol. 15, pp. 715–732, 2015.
[10] B. Wen, X. Chen, and T. K. Pong, "A proximal difference-of-convex algorithm with extrapolation," *Computational Optimization and Applications*, vol. 69, no. 2, pp. 297–324, 2018.
[11] H. Chen, L. Qi, Y. Wang, and G. Zhou, "Further results on sum-of-squares tensors," *Optimization Methods and Software*, vol. 1, 2020.
[12] H. Chen, Y. Wang, and G. Zhou, "High-order sum-of-squares structured tensors: theory and applications," *Frontiers of Mathematics in China*, vol. 15, no. 2, pp. 255–284, 2020.

[13] C. Wang, H. Chen, Y. Wang, and G. Zhou, "On copositiveness identification of partially symmetric rectangular tensors," *Journal of Computational and Applied Mathematics*, vol. 372, p. 112678, 2020.

[14] W. Wang, H. Chen, and Y. Wang, "A new C-eigenvalue interval for piezoelectric-type tensors," *Applied Mathematics Letters*, vol. 100, p. 106035, 2020.

[15] X. Wang, Y. Wang, Y. Wang, and G. Wang, "An accelerated augmented Lagrangian method for multi-criteria optimization problem," *Journal of Industrial & Management Optimization*, vol. 16, no. 1, pp. 1–9, 2020.

[16] X. Zhang, L. Liu, Y. Wu, and Y. Cui, "A sufficient and necessary condition of existence of blow-up radial solutions for a k-Hessian equation with a nonlinear operator," *Nonlinear Analysis Modelling Control*, vol. 25, pp. 126–143, 2020.

[17] X. Zhang, J. Xu, J. Jiang, Y. Wu, and Y. Cui, "The convergence analysis and uniqueness of blow-up solutions for a Dirichlet problem of the general k-Hessian equations," *Applied Mathematics Letters*, vol. 102, p. 106124, 2020.

[18] X. Zhang, J. Jiang, Y. Wu, and Y. Cui, "The existence and nonexistence of entire large solutions for a quasilinear Schrödinger elliptic system by dual approach," *Applied Mathematics Letters*, vol. 100, p. 106018, 2020.

[19] M.-M. Dong and H.-B. Chen, "Geometry of the copositive tensor cone and its dual," *Asia-Pacific Journal of Operational Research*, vol. 37, no. 4, p. 2040008, 2020.