Iteration complexity of the Difference-of-Convex Algorithm for unconstrained optimization: a simple proof

S. Gratton* and Ph. L. Toint†

12 I 2026

Abstract

We propose a simple proof of the worst-case iteration complexity for the Difference of Convex functions Algorithm DCA for unconstrained minimization, showing that the global rate of convergence of the norm of the objective function's gradients at the iterates converge to zero like $o(1/k)$. A small example is also provided indicating that the rate cannot be improved.

Keywords: Difference of convex functions (DC), DCA algorithm, iteration complexity.

1 Introduction

Proposed by Pham Dinh and Souad in [16] following a first decomposition approach by Mine and Fukushima [11], the possibility to decompose a possibly nonconvex function into the difference of two convex functions (DC) has raised a considerable interest, both practically and theoretically, as is testified by the many references on this topic (see, for instance, [2, 5, 7, 8, 14, 15, 10, 13] or the survey [9]). This interest is often motivated by the fact that functions of this particular class exhibit some structure, which one hopes to exploit in order to improve computational efficiency. It is therefore natural to wonder if this hope of improved performance can be translated into an improved worst-case complexity. As it turns out, this question was solved in [1] and further refined in [17], using the complex machinery of performance analysis as proposed in [4]. The proofs in these papers are non-trivial as this framework entails the semi-definite programming relaxation of an infinite dimensional optimization problem with an infinite number of constraints. The purpose of this short note is to show that both lower and upper tight complexity bounds may be obtained by a much simpler, direct and explicit approach, which has the advantage of conciseness but also of improved interpretability.

Thus, in what follows, we (re-)analyze the global convergence rate of the iterations of the standard algorithm for unconstrained minimization of DC functions and show that it is in order identical to that of the steepest descent method applied on the unstructured problem.

More specifically, we consider solving the smooth unconstrained problem

$$\min_{x\in\mathbb{R}^n} f(x) = g(x) - h(x) \tag{1.1}$$

where g and h are continuously differentiable convex functions from $\mathbb{R}^n$ into $\mathbb{R}$, the gradient of h being Lipschitz continuous. To do so, we will use the standard DCA (DC algorithm) [11, 16], which is stated on the following page.

*Université de Toulouse, INP, IRIT, Toulouse, France. Email: serge.gratton@enseeiht.fr. Work partially supported by 3IA Artificial and Natural Intelligence Toulouse Institute (ANITI), French "Investing for the Future - PIA3" program under the Grant agreement ANR-19-PI3A-0004"

†NAXYS, University of Namur, Namur, Belgium. Email: philippe.toint@unamur.be

Algorithm 1.1: DCA

Step 0: Initialization. A starting point x_0 is given. Set $k = 0$.

Step 1: Step computation. For $k \geq 0$, set

$$x_{k+1} = \operatorname*{argmin}_{x \in \mathbb{R}^n} g(x) - \nabla_x h(x_k)^T (x - x_k) \tag{1.2}$$

Observe that the function $g(x) - \nabla_x h(x_k)^T (x - x_k)$ (used to define the next iterate in Step 1) inherits the convexity of g. As a consequence the solution of the global minimization in (1.2) can make use of the arsenal of methods for convex problems, whose efficiency often vastly exceeds that of methods for general nonconvex ones. While this does influence the total evaluation complexity of the DCA, as we briefly discuss below, we focus in this note on its iteration complexity, or, equivalently, on the global rate of convergence of its iterations.

We also immediately note that, because h is convex, $g(x) - h(x_k) - \nabla_x h(x_k)^T (x - x_k)$ is an overestimate of $f(x)$ and therefore

$$f(x_{k+1}) \leq f(x_k) \tag{1.3}$$

for all $k \geq 0$.

2 Global rate of convergence

We start our analysis of the global rate of convergence of the algorithm by establishing an upper bound on the global rate of decrease of the sequence $\{\|\nabla_x f(x_k)\|\}$ to zero (see [1, Theorem 3.1]).

Theorem 2.1 Consider applying the DCA to problem (1.1). Suppose that g is continuously differentiable and strongly convex with constant $\mu > 0$, that is

$$g(y) \geq g(x) + \nabla_x g(x)^T (y - x) + \mu \|y - x\|^2 \quad \text{for all} \quad x, y \in \mathbb{R}^n, \tag{2.1}$$

and that h is convex, continuously differentiable and that its gradient is Lipschitz continuous with constant L_h, that is

$$\|\nabla_x h(x) - \nabla_x h(y)\| \leq L_h \|x - y\| \quad \text{for all} \quad x, y \in \mathbb{R}^n. \tag{2.2}$$

Suppose finally that there exists $f_{\rm low} \in \mathbb{R}$ such that $f(x) \geq f_{\rm low}$ for all $x \in \mathbb{R}^n$. Then, for $k \geq 2$,

$$\frac{1}{\lfloor k/2 \rfloor} \sum_{i=\lceil k/2 \rceil}^{k} \|\nabla_x f(x_i)\|^2 \leq \frac{2 L_h^2 (f(x_{\lceil k/2 \rceil}) - f(x_{k+1}))}{\mu\, k} = o\left(\frac{1}{k}\right) \tag{2.3}$$

Proof. The strong convexity of g implies that

$$g(x_k) \geq g(x_{k+1}) + \nabla_x g(x_{k+1})^T (x_k - x_{k+1}) + \mu \|x_{k+1} - x_k\|^2. \tag{2.4}$$

while the convexity of h gives that

$$h(x_{k+1}) \geq h(x_k) + \nabla_x h(x_k)^T (x_{k+1} - x_k). \tag{2.5}$$

Observe now that (1.2) implies that

$$\nabla_x g(x_{k+1}) = \nabla_x h(x_k). \tag{2.6}$$

Thus (2.4) gives that

$$g(x_k) \geq g(x_{k+1}) + \nabla_x h(x_k)^T (x_k - x_{k+1}) + \mu \|x_{k+1} - x_k\|^2. \tag{2.7}$$

Summing (2.5) and (2.7) then yields that

$$g(x_k) - h(x_k) \geq g(x_{k+1}) - h(x_{k+1}) + \mu \|x_{k+1} - x_k\|^2.$$

Summing now this inequality over the last $\lceil k/2 \rceil$ iterations gives that

$$f(x_{\lceil k/2 \rceil}) - f(x_{k+1}) = \sum_{i=\lceil k/2 \rceil}^{k} \big[\big(g(x_{i+1}) - h(x_{i+1})\big) - \big(g(x_i) - h(x_i)\big) \big] \geq \mu \sum_{i=\lceil k/2 \rceil}^{k} \|x_{i+1} - x_i\|^2. \tag{2.8}$$

Now, using (2.6) and (2.2), we see that, for all $i \geq 0$,

$$\|\nabla_x f(x_i)\| = \|\nabla_x g(x_i) - \nabla_x h(x_i)\| = \|\nabla_x h(x_{i+1}) - \nabla_x h(x_i)\| \leq L_h \|x_{i+1} - x_i\|.$$

This and (2.8) then imply that

$$\sum_{i=\lceil k/2 \rceil}^{k} \|\nabla_x f(x_i)\|^2 \leq L_h^2 \sum_{i=\lceil k/2 \rceil}^{k} \|x_{i+1} - x_i\|^2 \leq \frac{L_h^2 (f(x_{\lceil k/2 \rceil}) - f(x_{k+1}))}{\mu}.$$

Thus, dividing by $\lfloor k/2 \rfloor + 1$,

$$\frac{1}{\lfloor k/2 \rfloor + 1} \sum_{i=\lceil k/2 \rceil}^{k} \|\nabla_x f(x_i)\|^2 \leq \frac{L_h^2 (f(x_{\lceil k/2 \rceil}) - f(x_{k+1}))}{\mu (\lfloor k/2 \rfloor + 1)} \leq \frac{2 L_h^2 (f(x_{\lceil k/2 \rceil}) - f(x_{k+1}))}{\mu\, k}. \tag{2.9}$$

Since the sequence $\{f(x_k)\}$ is monotonically decreasing because of (1.3) and bounded below by assumption, it must be convergent and thus

$$\lim_{k \to \infty} (f(x_{\lceil k/2 \rceil}) - f(x_{k+1})) = 0.$$

Thus (2.9) yields (2.3). □

Note that inequality (2.8) recovers a variant of a known result on the sequence of iterates ([14, Proposition 2], [10, Corollary 1], [1, Theorem 1.2]).

Should one wish to stop the algorithm as soon as an iterate x_k is found such that

$$\|\nabla_x f(x_k)\| \leq \epsilon \tag{2.10}$$

for some user-prescribed tolerance $\epsilon > 0$, Theorem 2.1 ensures that at most $o(\epsilon^{-2})$ iterations are needed for termination. While this bound is theoretically better that $\mathcal{O}(\epsilon^{-2})$ of [1, Theorem 3.1], we now show that the number of iterations necessary to achieve (2.10) may be arbitrarily close to $\mathcal{O}(\epsilon^{-2})$, providing a lower bound on iteration complexity for the DCA. This is achieved by providing an explicit example of "slow" convergence, much as in [1, Section 3], but using simpler and more interpretable functions.

Theorem 2.2 Let the univariate function g be defined by $g(x) = \frac{1}{2}x^2$. Then, for any $\delta > 0$, there exists a convex continuously differentiable function h from $\mathbb{R}$ to $\mathbb{R}$ with Lipschitz continuous gradient such that the sequence of iterates of the DCA applied to problem (1.1) starting from $x_0 = 0$ is such that

$$\|\nabla_x f(x_k)\|^2 = \frac{1}{(k+1)^{1+2\delta}}. \tag{2.11}$$

Proof. Define the sequence $\{x_k\}$ by

$$x_{k+1} = x_k - \frac{1}{(k+1)^{\frac{1}{2}+\delta}} \quad (k \geq 0) \tag{2.12}$$

which implies that $\{x_k\}$ tends to $-\infty$. We now construct a function h by first imposing the conditions

$$\nabla_x h(x_k) = x_{k+1} \quad (k \geq 0) \tag{2.13}$$

and then defining h on $[x_{k+1}, 0]$ as the continuously differentiable piecewise quadratic interpolating the conditions (2.13) and such that $h(x_0) = h(0) = 0$. One easily checks that its second derivative on the interval $[x_{k+1}, x_k]$ is given by

$$\frac{\nabla_x h(x_{k+1}) - \nabla_x h(x_k)}{x_{k+1} - x_k} = \frac{x_{k+2} - x_{k+1}}{x_{k+1} - x_k} = \frac{(k+1)^{\frac{1}{2}+\delta}}{(k+2)^{\frac{1}{2}+\delta}} \in (0, 1]$$

and we then define

$$\begin{aligned} h(x_{k+1}) &= h(x_k) + \nabla_x h(x_k)(x_{k+1} - x_k) + \tfrac{1}{2}\nabla_x^2 h(x_k)(x_{k+1} - x_k)^2 \\ &= h(x_k) - \frac{x_{k+1}}{(k+1)^{\frac{1}{2}+\delta}} + \frac{1}{2(k+1)^{1+2\delta}} \left(\frac{k+1}{k+2}\right)^{\frac{1}{2}+\delta} \end{aligned} \tag{2.14}$$

Hence h is convex on $(-\infty, 0]$ and, using [3, Theorem A.8.5], its gradient is Lipschitz continuous with constant equal to one. It is then easy to prolongate $h(x)$ for $x > 0$ by setting

$$h(x) = h(0) + \nabla_x h(0)\, x = -x$$

for $x > 0$ without altering its convexity or the Lipschitz continuity of its gradient. Writing $g(x)$ as its second-order Taylor's series, we also have that, for $k \geq 0$,

$$g(x_{k+1}) = g(x_k) - \frac{x_k}{(k+1)^{\frac{1}{2}+\delta}} + \frac{1}{2(k+1)^{1+2\delta}},$$

and thus, combining this relation with (2.14), that

$$f(x_{k+1}) = f(x_k) - \frac{x_k - x_{k+1}}{(k+1)^{\frac{1}{2}+\delta}} + \frac{1}{2(k+1)^{1+2\delta}} \left[1 - \left(\frac{k+1}{k+2}\right)^{\frac{1}{2}+\delta}\right] \geq f(x_k) - \frac{1}{(k+1)^{1+2\delta}}.$$

Therefore, for all $k \geq 0$,

$$f(x_{k+1}) \geq f(x_0) - \sum_{i=0}^{k} \frac{1}{(i+1)^{1+2\delta}} \geq -\zeta(1+2\delta),$$

where $\zeta(\cdot)$ is the Riemann zeta function. Moreover, since both $g(x)$ and $h(x)$ have Lipschitz continuous gradients with unit constant, we have that, for $x \in [x_{k+1}, x_k]$,

$$f(x) \geq f(x_k) - \frac{1}{(k+1)^{1+2\delta}} - \|x - x_k\|^2 \geq f(x_k) - \frac{2}{(k+1)^{1+2\delta}} \geq f(x_k) - 2,$$

where we have used (2.12) to deduce the last inequality. This in turn implies that, for all $x \in (-\infty, 0]$,

$$f(x) \geq -\zeta(1+2\delta) - 2 \stackrel{\rm def}{=} f_{\rm low}. \tag{2.15}$$

Since $f(x) = \frac{1}{2}x^2 + x > 0$ for $x > 0$, we see that g and h satisfy all the conditions required by Theorem 2.1. We may then apply the DCA to minimize $f = g - h$. The definition of its iteration in (1.2) gives that

$$\operatorname*{argmin}_{x \in \mathbb{R}} \tfrac{1}{2}x^2 - x_{k+1}x = x_{k+1}. \tag{2.16}$$

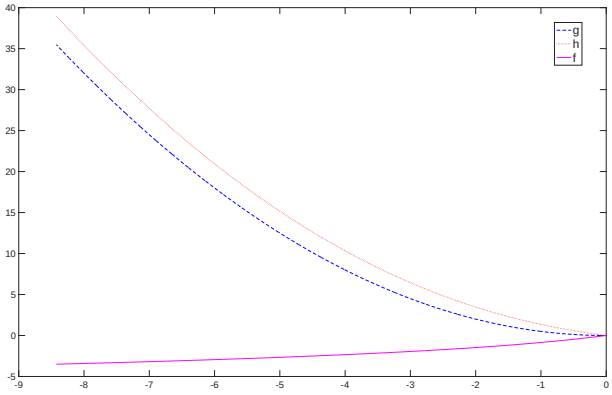

Figure 1: The shapes of f, g and h on $[x_{25}, x_0]$

and the iterates generated by the algorithm therefore coincide with the sequence $\{x_k\}$ defined in (2.12). Moreover, we deduce from (2.12) and (2.13) that

$$\|\nabla_x f(x_k)\| = \|x_k - x_{k+1}\| = \frac{1}{(k+1)^{\frac{1}{2}+\delta}},$$

which is (2.11). □

The shapes of f, g and h on $[x_{25}, x_0]$ are shown in Figure 1.

3 Discussion

We have provided proofs for the lower an upper bounds on iteration complexity for the DCA, simplifying results of [1]. The main conclusion that can be drawn from these results is that the minimization of DC functions using the DCA does not result in an better iteration complexity than that obtained for the steepest-descent method applied on f and ignoring the DC structure [6]. Thus distinguishing DC functions from general convex ones does not bring any benefit from the worst-case complexity point of view, should this algorithm be used. It is however important to remember that this does not mean that the practical perfoemance of the DCA is as slow as that of steepest descent (after all, this is also the case of several efficient variants of Newton's method, as shown in [3, Theorem 3.1.1]).

Our example of Theorem 2.2 also shows that using the DCA to exploit structure might be detrimental from the point of view of complexity. Indeed, it is easy to verify that the Hessian of f is strictly positive on each interval $[x_{k+1}, x_k]$, and the function f is therefore convex (albeit not strongly convex). The performance of the DCA on that convex function is therefore worse than what would be achieved by a good algorithm for general convex function (see [12, Chapter 2] for an extensive discussion).

We also note that evaluation complexity (that is the maximum number of oracle[1] evaluations needed to achieve (2.10)) is bounded below by iteration complexity, because the solution of the inner minimization subproblem of (1.2) does require at least one oracle evaluation.

We finally observe that our unidimensional example of slow convergence presented in Theorem 2.2 can also straightforwardly be extended to any dimension by considering functions of an arbitrary number of variables but whose value and gradient only depend on one. It is also remarkable that the example is independent of any termination rule, at variance with examples of slow convergence in [3]. This is due to the exploitation of the proof technique of [6] in Theorem 2.2.

[1] objective function's value and/or gradient.

Indeed, our example does not extend to the case where $\delta = 0$ unless a termination rule of the type (2.10) is introduced, because the constructed function f is no longer bounded below in that case.

Acknowledgement

Philippe Toint is grateful for the continued and friendly support of the APO team at Toulouse IRIT. Both authors thank M. Al-Baali (Sultan Qaboos University, Oman) for organizing the NAOVI-2026 conference in Muscat, during which the question of the complexity of DC optimization emerged after a talk by A. Bagirov.

References

[1] H. Abbaszadehpeivasti, E. de Klerk, and M. Zamani. On the rate of convergence of the difference-of-convex algorithm (dca). *Journal of Optimization Theory and Applications*, 202:475–496, 2024.

[2] F. J. Aragón Artacho, R. M. T. Fleming, and P. T. Vuong. Accelerating the DC algorithm for smooth functions. *Mathematical Programming, Series A*, 169:95–118, 2018.

[3] C. Cartis, N. I. M. Gould, and Ph. L. Toint. *Evaluation complexity of algorithms for nonconvex optimization*. Number 30 in MOS-SIAM Series on Optimization. SIAM, Philadelphia, USA, June 2022.

[4] Y. Drori and M. Teboulle. Performance of first-order methods for smooth convex minimization: a novel approach. *Mathematical Programming, Series A*, 145(1):451—-482, 2014.

[5] M. Gaudioso, G. Giallombardo, G. Miglionico, and A. Bagirov. Minimizing nonsmooth DC functions via successive DC piecewise-affine approximations. *Journal of Global Optimization*, 71(1):37–55, 2028.

[6] S. Gratton, C.-K. Sim, and Ph. L. Toint. Refining asymptotic complexity bounds for nonconvex optimization methods, including why steepest descent is $o(\epsilon^{-2})$ rather than $\mathcal{O}(\epsilon^{-2})$. *Computational Optimization and Applications*, 92:515–527, 2025.

[7] H. A. Le Thi. An efficient algorithm for globally minimizing a quadratic function under convex quadratic constraints. *Mathematical Programming*, 87(3):401–426, 2000.

[8] H. A. Le Thi and T. Pham Dinh. The DC (difference of convex functions) programming and DCA revisited with DC models of real world nonconvex optimization problems. *Annals of Operations Research*, 133(1-4):23–46, 2005.

[9] H. A. Le Thi and T. Pham Dinh. Open issues and recent advances in DC programming and DCA. *Journal of Global Optimization*, 88:533–590, 2023.

[10] H. A. Le Thi, D. N. Phan, and T. Pham Dinh. DCA based approaches for bi-level variable selection and application for estimate multiple sparse covariance matrices. *Neurocomputing*, 466:162—177, 2021.

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

[12] Yu. Nesterov. *Lectures on Convex Optimization*. Springer Verlag, Heidelberg, Berlin, New York, 2018.

[13] Y.-S. Niu. On the convergence of the DCA. arXiv:211.10942, 2011.

[14] T. Pham Dinh and H. A. Le Thi. Convex analysis approach to dc programming: theory, algorithms and applications. *Acta Math. Vietnam*, 22(1):289—355, 1997.

[15] T. Pham Dinh and H. A. Le Thi. D.C. optimization algorithm for solving the trust-region subproblem. *SIAM Journal on Optimization*, 8(2):476–505, 1998.

[16] T. Pham Dinh and E. B. Souad. Algorithms for solving a class of nonconvex optimization problems. methods of subgradients. In *FERMAT Days 85: Mathematics for Optimization*, volume 129 of *North-Holland Mathematics Studies*, page 249–271, Amsterdam, The Netherlands, 1986. Elsevier.

[17] T. Rotaru, P. Panagiotis, and F. Glineur. Tight analysis of difference-of-convex algorithm (DCA) improves convergence rates for proximal gradient descent. *Proceedings of Machine Learning Research*, 258:4114–4122, 2025.