Tight Convergence Rates in Gradient Mapping for the Difference-of-Convex Algorithm

Teodor Rotaru[1,2(✉)], Panagiotis Patrinos[1], and François Glineur[2]

[1] Department of Electrical Engineering (ESAT-STADIUS), KU Leuven, Kasteelpark Arenberg 10, Leuven 3001, Belgium
{teodor.rotaru,panos.patrinos}@esat.kuleuven.be

[2] Department of Mathematical Engineering (ICTEAM-INMA), UCLouvain, Av. Georges Lemaître 4, Louvain-la-Neuve 1348, Belgium
{teodor.rotaru,francois.glineur}@uclouvain.be

Abstract. We establish new theoretical convergence guarantees for the difference-of-convex algorithm (DCA), where the second function is allowed to be weakly-convex, measuring progress via composite gradient mapping. Based on a tight analysis of two iterations of DCA, we identify six parameter regimes leading to sublinear convergence rates toward critical points and establish those rates by proving adapted descent lemmas. We recover existing rates for the standard difference-of-convex decompositions of nonconvex-nonconcave functions, while for all other curvature settings our results are new, complementing recently obtained rates on the gradient residual. Three of our sublinear rates are tight for any number of DCA iterations, while for the other three regimes we conjecture exact rates, using insights from the tight analysis of gradient descent and numerical validation using the performance estimation methodology. Finally, we show how the equivalence between proximal gradient descent (PGD) and DCA allows the derivation of exact PGD rates for any constant stepsize.

Keywords: Difference-of-Convex Algorithm · Performance Estimation

1 Introduction

We investigate the difference-of-convex (DC) formulation

$$\underset{x\in\mathbb{R}^d}{\text{minimize}}\, F(x) := f_1(x) - f_2(x),$$

where $f_i \in \mathcal{F}_{\mu_i,L_i}$, $i = \{1,2\}$, denoting proper, lower semicontinuous (l.s.c.) functions $f_i : \mathbb{R}^d \to \mathbb{R}$ having curvature in the interval $[\mu_i, L_i]$, i.e., $\frac{L_i\|\cdot\|^2}{2} - f_i$ and $f_i - \frac{\mu_i\|\cdot\|^2}{2}$ are convex. If $f_i \in \mathcal{C}^2$, then μ_i and L_i correspond to the minimum and maximum eigenvalues of the Hessian. When $L_i < \infty$, then f_i is smooth.

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2026
H. A. Le Thi et al. (Eds.): MCO 2025, LNNS 1688, pp. 61–73, 2026.
https://doi.org/10.1007/978-3-032-08381-4_5

Algorithm 1: Difference-of-convex algorithm (DCA)

Data: $f_1 \in \mathcal{F}_{\mu_1,L_1}$, $f_2 \in \mathcal{F}_{\mu_2,L_2}$, with $\mu_1 \geq 0$, $\mu_1 + \mu_2 > 0$ or $\mu_1 = \mu_2 = 0$;
$N \geq 1$ iterations starting from $x^0 \in \operatorname{dom} \partial f_1$

for $k = 0, \ldots, N$ **do**
  Select $g_2^k \in \partial f_2(x^k)$;
  Select $x^{k+1} \in \operatorname{argmin}_{w \in \mathbb{R}^d} \{f_1(w) - \langle g_2^k, w \rangle\}$

Result: Best iterate x^k with $k = \operatorname{argmin}_{0 \leq k \leq N} \{\|x^k - x^{k+1}\|^2\}$

Moreover, f_i is μ_i-convex, including weakly-convexity under $\mu_i < 0$. We assume $\mu_1 \geq 0$ and allow $\mu_2 < 0$, which extends the standard DC case where both functions are convex. The difference-of-convex algorithm (DCA), stated in Algorithm 1, is a standard method to solve the DC problem. Various work provide extensive analysis of DCA [6,10,13,20], which is also referred as the convex-concave procedure (CCCP) [11,14,27]. In this work we focus on improving the theoretical guarantees. For the interested reader, a comprehensive list of DCA applications is found in the review paper [13].

If the entire sequence of iterates $\{x^k\}$ is bounded, then each accumulation point of it is a critical point of F [6, Theorem 3(iv)]. Inspired by this result, we aim to find the best iterate x^k providing the lowest value of $\{\|x^k - x^{k+1}\|\}$ over N iterations. This metric is referred as the *best (composite) gradient mapping* [15, 16] or the *prox-gradient mapping* [5], and is complementary to the *best residual gradient norm* $\{\|g_1^k - g_2^k\|\}$ over N iterations, where $g_i \in \partial f_i(x^k)$, $i = \{1, 2\}$, are any subgradients of f_i evaluated at x^k. On this metric, [3] provides the first tight rates for any convex functions f_i when F is nonconvex and nonconcave, using the performance estimation problem (PEP) methodology [7,21,22]. The analysis is later refined and extended in [19] to accommodate weakly-convex functions f_2. For the standard DC setup, [8] provides convergence rates in terms of Bregman distance, viewing DCA as a instance of the Bregman proximal point algorithm, while [12,26] provides linear rates under the Polyak-Łojasiewicz (PL) inequality. The Frank-Wolfe algorithm [28] and the proximal gradient descent (PGD) [13] are also equivalent to DCA.

Contributions. In Theorem 1 we establish new theoretical convergence guarantees for DCA and PGD, measuring progress via composite gradient mapping. When both f_1 and f_2 are convex and F is nonconvex-nonconcave, we refine the analysis from [3], providing standard descent lemmas in Sect. 2.1. Our results extend existing work by introducing new rates for all other curvature settings, complementing the best gradient residual rates from [3,19]. Building on insights from gradient descent (GD), we conjecture in Conjecture 1 the exact behavior of DCA across all curvature choices, with some regimes rigorously proven (Theorem 2 and 3) and others remaining open. As shown in Sect. 4, the equivalence between PGD and DCA enables us to formulate precise PGD rates for any constant stepsize, including the ones exceeding the inverse Lipschitz constant.

Finally, we propose a general adaptive curvature shifting procedure for DCA, leveraging PGD stepsize schedules aligned with established GD strategies.

Our derivations are aided by the PEP technique, allowing to establish the tightness of the regimes and validate our conjecture. The setup of PEP for DCA is described in [3,19] and can be solved using dedicated software packages [9,23].

Theoretical Setup and Running Assumptions. Let f be a proper, convex, l.s.c. function. Its subdifferential at $x \in \mathbb{R}^d$ is $\partial f(x) := \{g \in \mathbb{R}^d \,|\, f(y) \geq f(x) + \langle g, y - x\rangle, \forall y \in \mathbb{R}^d\}$. If f is weakly-convex with $\mu < 0$, let $\tilde{f}(x) := f(x) - \mu\frac{\|x\|^2}{2}$ be a convex function. Since its subdifferential $\partial\tilde{f}(x)$ is well-defined, we take $\partial f(x) := \partial\tilde{f}(x) + \mu x$ [4, Proposition 6.3], which is well-defined. For f differentiable at x we have $\partial f(x) = \{\nabla f(x)\}$. The domain and range of f are $\operatorname{dom} f := \{x \in \mathbb{R}^d : f(x) < \infty\}$ and $\operatorname{range} f := \{y \in \mathbb{R} : \exists\, x \in \operatorname{dom} f \text{ with } y = f(x)\}$. The domain and range of the subdifferential are $\operatorname{dom}\partial f = \{x \in \mathbb{R}^d : \partial f(x) \neq \emptyset\}$ and $\operatorname{range}\partial f = \cup\{\partial f(x) : x \in \operatorname{dom}\partial f\}$. The convex conjugate of f l.s.c. function is defined as $f^*(y) := \sup_{x\in\operatorname{dom} f}\{\langle y, x\rangle - f(x)\}$, where f^* is closed and convex.

Assumption 1. The objective function $F = f_1 - f_2$ is lower bounded by $F_{lo} := \inf_x F$, where $f_1 \in \mathcal{F}_{\mu_1,L_1}$ and $f_2 \in \mathcal{F}_{\mu_2,L_2}$ with $0 \leq \mu_1 < L_1 \leq \infty$, $\mu_2 \in (-\infty, \infty)$ and $L_2 \in (\mu_2, \infty]$. Additionally, assume $\varnothing \neq \operatorname{dom}\partial f_1 \subseteq \operatorname{dom}\partial f_2$ and $\operatorname{range}\partial f_2 \subseteq \operatorname{range}\partial f_1$. Note that $F \in \mathcal{F}_{\mu_1-L_2\,,\,L_1-\mu_2}$. The second part of Assumption 1 implies that the DCA iterations are well-defined, meaning there exists a sequence $\{x^k\}$, starting from $x^0 \in \operatorname{dom}\partial f_1$, generated by Algorithm 1, such that $x^{k+1} \in \partial f_1^*(\partial f_2(x^k))$. By definition, this condition implies the existence of $g_1^{k+1} \in \partial f_1(x^{k+1})$ such that $g_1^{k+1} = g_2^k$, where $g_2^k \in \partial f_2(x^k)$. This is the only characterization of DCA iterations employed in our derivations.

DCA finds critical points x^* satisfying $\partial f_2(x^*) \cap \partial f_1(x^*) \neq \emptyset$. These are stationary only if f_2 is smooth [17, Exercise 10.10]. If only f_1 is smooth, then only the inclusion $\partial(-f_2)(x^*) \subseteq -\partial f_2(x^*)$ is guaranteed [17, Corollary 9.21], hence $\partial F(x^*) \subseteq \nabla f_1(x^*) - \partial f_2(x^*)$ and critical points are not necessary stationary.

We assume that at each iteration a subgradient from ∂f_1^* is available and only focus on the progress of the iterations $\min_{0\leq k\leq N}\{\|x^k - x^{k+1}\|^2\}$ subject to the initial condition $F(x^0) - F(x^{N+1}) \leq \Delta$, with $\Delta > 0$. To remove the dependence of the initial condition on the last iterate, in all our rates the initial objective gap can be further upper bounded by $F(x^0) - F_{lo}$.

2 Sublinear Convergence Rates

Theorem 1. *Let $f_1 \in \mathcal{F}_{\mu_1,L_1}$ and $f_2 \in \mathcal{F}_{\mu_2,L_2}$ satisfying Assumption 1 and $\mu_1 + \mu_2 > 0$. Then after $N + 1 \geq 2$ iterations of DCA starting from x^0 it holds*

$$\tfrac{1}{2}\min_{0\leq k\leq N}\{\|x^k - x^{k+1}\|^2\} \leq \frac{F(x^0) - F(x^{N+1})}{(\mu_1 + \mu_2) + p_i(\mu_1, L_1, \mu_2, L_2)N}, \tag{1}$$

where the expressions of the regimes $p_{i=\{1,\dots,6\}}$ are given in Table 1.

Table 1. Coefficients p_i from Theorem 1. Their domains cover all parameter space under the condition $\mu_1+\mu_2 > 0$. Regimes $p_{\{1,2,3\}}$ are tight for any number of iterations. (s.c.: strongly convex; n.c.: nonconvex; ccv.: concave; n.ccv.: nonconcave)

Regime	Domain		Description
$p_1 = \mu_1 + \mu_2 + \frac{(\mu_1-\mu_2)^2}{L_2-\mu_2}$	$\mu_1 \geq \mu_2 \geq 0$	$L_2 > \mu_1$	f_1, f_2 convex F n.c.-n.ccv.
$p_2 = \mu_1 + \mu_2 + \frac{(\mu_1-\mu_2)^2}{L_1-\mu_1}$	$\mu_2 \geq \mu_1 \geq 0$	$L_1 > \mu_2$	f_1, f_2 convex F n.c.-n.ccv.
$p_3 = \frac{(\mu_1+\mu_2)(L_2+\mu_1)}{L_2+\mu_2}$	$\mu_2 \in [\frac{-L_2\mu_1}{L_2+\mu_1}, 0)$	$L_2 > \mu_1 > 0$	f_1 s.c., f_2 n.c. F n.c.-n.ccv.
$p_4 = \frac{\mu_1^2(L_2+\mu_1)}{L_2^2}$	$\mu_2 \in \left[\frac{-L_2\mu_1}{L_2+\mu_1}, L_2\right)$	$L_2 \in [0, \mu_1]$	f_1 s.c., f_2 n.c. F convex
$p_5 = \frac{\mu_1^2(\mu_1+\mu_2)}{\mu_2^2}$	$\mu_2 < 0$, $\mu_2 \in \left(-\mu_1, \frac{-L_2\mu_1}{L_2+\mu_1}\right]$		f_1 s.c., f_2 n.c. F n.ccv.
$p_6 = \frac{\mu_2^2(L_1+\mu_2)}{L_1^2}$	$\mu_2 \geq \mu_1 \geq 0$	$L_1 \in (0, \mu_2]$	f_1 convex, f_2 s.c. F ccv.

The proof is deferred to Sect. 2.2. All regimes are exact for $N = \{0, 1\}$ and their domains span the entire range of parameters satisfying the condition $\mu_1 + \mu_2 > 0$, which is sufficient for decreasing the objective (see Proposition 1).

Based on extensive numerical experiments using PEP, we determine that regimes $p_{\{1,2,3\}}$ are tight for any number of iterations. Regimes p_1 and p_2 are derived in [3, Proposition 3.1], by applying Toland duality [25] on the rates measuring the best gradient residual criterion. Theorem 1 gives more explicit expressions and includes in the proofs a simpler derivation based on Lemmas 2 and 3. These regimes assume F is nonconvex-nonconcave ($L_1 > \mu_2$ and $L_2 > \mu_1$), with the active one indexed by $\operatorname{argmax}_i\{\mu_1, \mu_2\}$. All other regimes are new.

When F is nonconvex-nonconcave, regime p_3 is active when f_2 is weakly-convex with $\mu_2 \in [\frac{-L_2\mu_1}{L_2+\mu_1}, 0)$. When μ_2 is even lower, then regime p_5 is active. This is however not tight for $N \geq 2$, where a more complicated analysis is required. The exact rates expressions in this case are given in Conjecture 1.

When F is (strongly) convex ($\mu_1 \geq L_2$), the sublinear rates of regimes p_4 and p_5 are tight only for $N = 1$. These regimes are separated in this case by the threshold $\mu_2 = \frac{-L_2\mu_1}{L_2+\mu_1}$. When $N \geq 2$, we prove a linear rate for regime p_4 in Theorem 2, which is confirmed to be exact using PEP. Within the particular range of p_5 satisfying the condition $L_2 + \mu_2 \leq 0$, we demonstrate a linear rate in Theorem 3. For the rest domain of p_5, we conjecture the tight rates in Conjecture 1. Finally, regime p_6 characterizes the iterates over *concave* objectives F and is included for a complete covering of the parameter space.

2.1 Lemmas to Prove Sublinear Rates

For compactness of the proofs, at each iteration k we denote $\Delta x^k := x^k - x^{k+1}$, $f_1^k = f_1(x^k)$, $f_2^k = f_2(x^k)$, $\Delta F(x^k) := F(x^k) - F(x^{k+1})$, $g_1^k \in \partial f_1(x^k)$, $g_2^k \in \partial f_2(x^k)$, $G^k := g_1^k - g_2^k$. Recall that $g_1^{k+1} = g_2^k$ for all DCA iterations.

Lemma 1. *Let $f_1 \in \mathcal{F}_{\mu_1,L_1}$ and $f_2 \in \mathcal{F}_{\mu_2,L_2}$, with $L_2 > \mu_2$ and $L_1 > \mu_1 \geq 0$. Consider one DCA iteration connecting x^k and x^{k+1}. Let $g_1^j \in \partial f_1(x^j)$, $g_2^j \in \partial f_2(x^j)$ and $G^j = g_1^j - g_2^j$, where $j = \{k, k+1\}$. The following inequalities hold:*

$$\Delta F(x^k) \geq \tfrac{\mu_1+\mu_2}{2}\|\Delta x^k\|^2 + \tfrac{\|G^{k+1}-\mu_2\Delta x^k\|^2}{2(L_2-\mu_2)} + \tfrac{\|G^k-\mu_1\Delta x^k\|^2}{2(L_1-\mu_1)}; \tag{B^k}$$

$$\langle G^k, \Delta x^k\rangle \geq \mu_1\|\Delta x^k\|^2 + \tfrac{\|G^k-\mu_1\Delta x^k\|^2}{L_1-\mu_1}; \tag{$C_{f_1}^k$}$$

$$\langle G^{k+1}, \Delta x^k\rangle \geq \mu_2\|\Delta x^k\|^2 + \tfrac{\|G^{k+1}-\mu_2\Delta x^k\|^2}{L_2-\mu_2}. \tag{$C_{f_2}^k$}$$

Proof. Given $\mu < L$ and $f \in \mathcal{F}_{\mu,L}$, from [21, Theorem 4], [22, Theorem 3.10], [18, Theorem 3.2] it holds $\forall x, y \in \mathbb{R}^d$, with $g^x \in \partial f(x)$, $g^y \in \partial f(y)$, that

$$f(x) - f(y) - \langle g^y, x-y\rangle \geq \tfrac{\mu}{2}\|x-y\|^2 + \tfrac{1}{2(L-\mu)}\|g^x - g^y - \mu(x-y)\|^2. \tag{2}$$

We exploit the property $g_1^{k+1} = g_2^k$, implying $G^k = g_1^k - g_2^k = g_1^k - g_1^{k+1}$ and $G^{k+1} = g_1^{k+1} - g_2^{k+1} = g_2^k - g_2^{k+1}$. By writing inequality (2) for: f_1 with the iterates (x^k, x^{k+1}) and (x^{k+1}, x^k) we get (3) and (4), respectively; f_2 with the iterates (x^k, x^{k+1}) and (x^{k+1}, x^k) we obtain (5) and (6), respectively:

$$f_1^k - f_1^{k+1} - \langle g_1^{k+1}, \Delta x^k\rangle \geq \tfrac{\mu_1}{2}\|\Delta x^k\|^2 + \tfrac{\|G^k-\mu_1\Delta x^k\|^2}{2(L_1-\mu_1)}; \tag{3}$$

$$f_1^{k+1} - f_1^k + \langle g_1^k, \Delta x^k\rangle \geq \tfrac{\mu_1}{2}\|\Delta x^k\|^2 + \tfrac{\|G^k-\mu_1\Delta x^k\|^2}{2(L_1-\mu_1)}; \tag{4}$$

$$f_2^k - f_2^{k+1} - \langle g_2^{k+1}, \Delta x^k\rangle \geq \tfrac{\mu_2}{2}\|\Delta x^k\|^2 + \tfrac{\|G^{k+1}-\mu_2\Delta x^k\|^2}{2(L_2-\mu_2)}; \tag{5}$$

$$f_2^{k+1} - f_2^k + \langle g_2^k, \Delta x^k\rangle \geq \tfrac{\mu_2}{2}\|\Delta x^k\|^2 + \tfrac{\|G^{k+1}-\mu_2\Delta x^k\|^2}{2(L_2-\mu_2)}. \tag{6}$$

Then summing (3) and (6) gives (B^k); inequality ($C_{f_1}^k$) results by summing (3) and (4); inequality ($C_{f_2}^k$) is obtained by summing (5) and (6). □

Proposition 1. (*Decrease after one iteration*) *Let $f_1 \in \mathcal{F}_{\mu_1,L_1}$ and $f_2 \in \mathcal{F}_{\mu_2,L_2}$ satisfying Assumption 1. Consider one DCA iteration connecting x^k and x^{k+1}. Then it holds*

$$F(x^k) - F(x^{k+1}) \geq \tfrac{\mu_1+\mu_2}{2}\|x^k - x^{k+1}\|^2. \tag{7}$$

Consequently, the objective function F decreases after each DCA iteration if $\mu_1 + \mu_2 \geq 0$ and the decrease is strict if $\mu_1 + \mu_2 > 0$, unless $x^{k+1} = x^k$.

Proof. Inequality (7) results by ignoring the mixed squares in (B^k). □

Proposition 1 was first proved in [6, Theorem 3, Proposition 2].

Remark 1. (*Proofs structure*) The proofs are based on multiplying the inequalities from Lemma 1 with nonnegative weights. For regimes $p_{1,3,4,5}$ we sum:

$$B^k + \beta_1 \times C_{f_1}^{k+1} + \beta_2 \times C_{f_2}^{k}, \tag{8}$$

while for the proofs of $p_{2,6}$ we sum: $B^{k+1} + \beta_1 \times C_{f_1}^{k} + \beta_2 \times C_{f_2}^{k+1}$, where $B^{\{k,k+1\}}$, $C_{f_1}^{\{k,k+1\}}$, $C_{f_2}^{\{k,k+1\}}$ are instances of inequalities (B^k), ($C_{f_1}^k$) and ($C_{f_2}^k$), respectively. The weights β_1 and β_2 are functions of the curvature parameters.

The demonstrations rewrite the inequalities from Remark 1 by building specific mixed squared norms of G and Δx at specific indices k, using Property 1. For space reasons, we only give the final inequalities containing all the squares.

Property 1. The following identities hold, where $i, j \in \{1, 2\}$ and G and Δx are specialized when building the squares within the proofs:
$-\langle G, \Delta x\rangle = \frac{1}{2\mu_j}\left[\|G - \mu_j \Delta x\|^2 - \|G\|^2 - \|\mu_j \Delta x\|^2\right]$ and
$\|G - \mu_i \Delta x\|^2 = \frac{\mu_i}{\mu_j}\|G - \mu_j \Delta x\|^2 - \frac{\mu_i}{\mu_j}(1 - \frac{\mu_i}{\mu_j})\|\mu_j \Delta x\|^2 + (1 - \frac{\mu_i}{\mu_j})\|G\|^2$.

Lemma 2. (Regime p_1). *Let $f_1 \in \mathcal{F}_{\mu_1,L_1}$ and $f_2 \in \mathcal{F}_{\mu_2,L_2}$, with $L_2 > \mu_1 \geq \mu_2 \geq 0$. Then after two DCA iterations connecting x^k, x^{k+1} and x^{k+2} it holds* $\Delta F(x^k) \geq \left(\mu_1 + \mu_2 \frac{L_2 - \mu_1}{L_2 - \mu_2}\right)\frac{1}{2}\|\Delta x^k\|^2 + \mu_1 \frac{\mu_1 - \mu_2}{L_2 - \mu_2}\frac{1}{2}\|\Delta x^{k+1}\|^2$.

Proof. Let $\beta_2 = 0$ and $\beta_1 := \frac{\mu_1 - \mu_2}{L_2 - \mu_2}$, which is nonnegative since $\mu_1 \geq \mu_2$. After replacing in (8), building squares and performing simplifications we get:

$$\Delta F(x^k) \geq \left(\mu_1 + \mu_2 \tfrac{L_2 - \mu_1}{L_2 - \mu_2}\right)\tfrac{1}{2}\|\Delta x^k\|^2 + \mu_1 \tfrac{\mu_1 - \mu_2}{L_2 - \mu_2}\tfrac{1}{2}\|\Delta x^{k+1}\|^2 + \frac{\|G^k - \mu_1 \Delta x^k\|^2}{2(L_1 - \mu_1)} + \frac{\mu_2 \|G^{k+1} - \mu_1 \Delta x^k\|^2}{2\mu_1 (L_2 - \mu_2)} + \frac{(L_1 + \mu_1)(\mu_1 - \mu_2)\|G^{k+1} - \mu_1 \Delta x^{k+1}\|^2}{2\mu_1 (L_1 - \mu_1)(L_2 - \mu_2)}.$$

Since $\mu_1 \geq \mu_2 \geq 0$, all coefficients of the squares are positive and the conclusion follows by dropping the mixed terms. □

Lemma 3. (Regime p_2). *Let $f_1 \in \mathcal{F}_{\mu_1,L_1}$ and $f_2 \in \mathcal{F}_{\mu_2,L_2}$, with $L_1 > \mu_2 \geq \mu_1 \geq 0$. Then after two DCA iterations connecting x^k, x^{k+1} and x^{k+2} it holds* $\Delta F(x^{k+1}) \geq \mu_2 \frac{\mu_2 - \mu_1}{L_1 - \mu_1}\frac{1}{2}\|\Delta x^k\|^2 + \left(\mu_2 + \mu_1 \frac{L_1 - \mu_2}{L_1 - \mu_1}\right)\frac{1}{2}\|\Delta x^{k+1}\|^2$.

Proof. Let $\beta_1 = 0$ and $\beta_2 := \frac{\mu_2 - \mu_1}{L_1 - \mu_1}$, nonnegative since $\mu_2 \geq \mu_1$, be the specific multipliers from Remark 1. Building the mixed squares we obtain:

$$\Delta F(x^{k+1}) \geq \left(\mu_2 + \mu_1 \tfrac{L_1 - \mu_2}{L_1 - \mu_1}\right)\tfrac{1}{2}\|\Delta x^{k+1}\|^2 + \mu_2 \tfrac{\mu_2 - \mu_1}{2(L_1 - \mu_1)}\|\Delta x^k\|^2 + \frac{\|G^{k+2} - \mu_2 \Delta x^{k+1}\|^2}{2(L_2 - \mu_2)} + \frac{\mu_1 \|G^{k+1} - \mu_2 \Delta x^{k+1}\|^2}{2\mu_2 (L_1 - \mu_1)} + \frac{(L_2 + \mu_2)(\mu_2 - \mu_1)\|G^{k+1} - \mu_2 \Delta x^k\|^2}{2\mu_2 (L_2 - \mu_2)(L_1 - \mu_1)}.$$

Since $\mu_2 \geq \mu_1 \geq 0$, all coefficients of the squares are positive and the conclusion follows by dropping the mixed terms. □

Property 2. (T_1) When $\mu_2 < 0$, we define $T_1 := \mu_1\left(\frac{L_2 + \mu_1}{L_2^2} - \frac{\mu_1 + \mu_2}{\mu_2^2}\right)$. One can check that $T_1 \geq 0$ if $\mu_2 \in (-\mu_1, \frac{-L_2\mu_1}{L_2 + \mu_1}]$ and $T_1 \leq 0$ if $\mu_2 \in [\frac{-L_2\mu_1}{L_2 + \mu_1}, L_2)$.

Lemma 4. (Regime p_3). *Let $f_1 \in \mathcal{F}_{\mu_1,L_1}$ and $f_2 \in \mathcal{F}_{\mu_2,L_2}$, with $L_2 > \mu_1 > 0$, $\mu_2 \in [\frac{-L_2\mu_1}{L_2+\mu_1}, 0)$ and $\mu_1 + \mu_2 > 0$. Consider two DCA iterations connecting x^k, x^{k+1} and x^{k+2}. Then $\Delta F(x^k) \geq \left(\mu_1 + \frac{L_2\mu_2}{L_2+\mu_2}\right)\frac{1}{2}\|\Delta x^k\|^2 + \frac{\mu_1^2}{2(L_2+\mu_2)}\|\Delta x^{k+1}\|^2$.*

Proof. Let $\beta_1 := \frac{\mu_1}{L_2+\mu_2}$ and $\beta_2 := \frac{-\mu_2}{L_2+\mu_2}$, which are nonnegative since $\mu_1 > -\mu_2 > 0$ and $L_2 + \mu_2 \geq \frac{-\mu_2 L_2}{\mu_1} > 0$. After replacing in (8), building squares and performing simplifications we obtain:

$$\Delta F(x^k) \geq \left(\mu_1 + \frac{L_2\mu_2}{L_2+\mu_2}\right)\frac{1}{2}\|\Delta x^k\|^2 + \frac{\mu_1^2}{2(L_2+\mu_2)}\|\Delta x^{k+1}\|^2 + \frac{\|G^k-\mu_1\Delta x^k\|^2}{2(L_1-\mu_1)} + \frac{(L_1+\mu_1)\|G^{k+1}-\mu_1\Delta x^{k+1}\|^2}{2(L_1-\mu_1)(L_2+\mu_2)}.$$

Since $L_2 > -\mu_2$, the coefficients of the mixed terms are positive and the conclusion follows by dropping them. The coefficient of $\|\Delta x^k\|^2$ is positive if $T_1 \leq 0$ as it can be factorized as $\frac{L_2^2\mu_2^2}{\mu_1(L_2^2-\mu_2^2)}(-T_1)$. The inequality holds even for $T_1 > 0$. □

Lemma 5. (Regime p_4). *Let $f_1 \in \mathcal{F}_{\mu_1,L_1}$ and $f_2 \in \mathcal{F}_{\mu_2,L_2}$, with $L_2 \in [0, \mu_1]$, $\mu_2 \in \left[\frac{-L_2\mu_1}{L_2+\mu_1}, L_2\right)$ and $\mu_1 + \mu_2 > 0$. Consider two DCA iterations connecting x^k, x^{k+1} and x^{k+2}. Then it holds $\Delta F(x^k) \geq \frac{\mu_1^2(L_2+\mu_1)}{L_2^2}\frac{1}{2}\|\Delta x^{k+1}\|^2$.*

Proof. Let $\beta_1 := \frac{\mu_1(\mu_1+L_2)}{L_2^2}$ and $\beta_2 := \frac{\mu_1}{L_2}$, both positive. After replacing in (8), building the mixed squares and performing simplifications we get:

$$\Delta F(x^k) \geq \frac{\mu_1^2(L_2+\mu_1)}{L_2^2}\frac{1}{2}\|\Delta x^{k+1}\|^2 + \frac{\|G^k-\mu_1\Delta x^k\|^2}{2(L_1-\mu_1)} + \frac{(L_1+\mu_1)(L_2+\mu_1)\|G^{k+1}-\mu_1\Delta x^{k+1}\|^2}{L_2^2(L_1-\mu_1)} + \frac{\mu_2^2\,(-T_1)\|G^{k+1}-L_2\Delta x^k\|^2}{\mu_1(L_2-\mu_2)^2}.$$

Since $T_1 \leq 0$ (Property 2), the conclusion follows. □

Lemma 6. (Regime p_5). *Let $f_1 \in \mathcal{F}_{\mu_1,L_1}$ and $f_2 \in \mathcal{F}_{\mu_2,L_2}$, with $\mu_1 > -\mu_2 > 0$ such that $\mu_2 \in \left(-\mu_1, \frac{-L_2\mu_1}{L_2+\mu_1}\right]$. Consider two DCA iterations connecting x^k, x^{k+1} and x^{k+2}. Then it holds $\Delta F(x^k) \geq \frac{\mu_1^2(\mu_1+\mu_2)}{\mu_2^2}\frac{1}{2}\|\Delta x^{k+1}\|^2$.*

Proof. Let $\beta_1 := \frac{\mu_1(\mu_1+\mu_2)}{\mu_2^2}$ and $\beta_2 := \frac{\mu_1+\mu_2}{-\mu_2}$, both positive. After replacing in (8), building the mixed squares and performing simplifications we get:

$$\Delta F(x^k) \geq \frac{\mu_1^2(\mu_1+\mu_2)}{2\mu_2^2}\|\Delta x^{k+1}\|^2 + \frac{\|G^k-\mu_1\Delta x^k\|^2}{2(L_1-\mu_1)} + \frac{(L_1+\mu_1)(\mu_1+\mu_2)\|G^{k+1}-\mu_1\Delta x^{k+1}\|^2}{\mu_2^2(L_1-\mu_1)} + \frac{L_2^2\,T_1\|G^{k+1}-\mu_2\Delta x^k\|^2}{\mu_1(L_2-\mu_2)^2}.$$

Since $T_1 \geq 0$ (property 2), the conclusion follows. □

Lemma 7. (Regime p_6). *Let $f_1 \in \mathcal{F}_{\mu_1,L_1}$ and $f_2 \in \mathcal{F}_{\mu_2,L_2}$, with $\mu_2 \geq \mu_1 \geq 0$ and $L_1 \in (0, \mu_2]$. Then after two DCA iterations connecting x^k, x^{k+1} and x^{k+2} we have $\Delta F(x^{k+1}) \geq \frac{\mu_2^2(L_1+\mu_2)}{L_1^2}\frac{1}{2}\|\Delta x^k\|^2$.*

Proof. Let $\beta_2 := \frac{\mu_2(\mu_2+L_1)}{L_1^2}$ and $\beta_1 := \frac{\mu_2}{L_1}$, both positive. After replacing in (8), building the mixed squares and performing simplifications we get:

$$\Delta F(x^{k+1}) \geq \frac{\mu_2^2(L_1+\mu_2)}{L_1^2}\frac{1}{2}\|\Delta x^k\|^2 + \frac{\|G^{k+2}-\mu_2\Delta x^{k+1}\|^2}{2(L_2-\mu_2)} + \frac{(L_1+\mu_2)(L_2+\mu_2)\|G^{k+1}-\mu_2\Delta x^k\|^2}{L_1^2(L_2-\mu_2)} + \frac{\mu_1\mu_2+L_1(\mu_1+\mu_2)\|G^{k+1}-L_1\Delta x^{k+1}\|^2}{L_1^2(L_1-\mu_1)}.$$

The conclusion follows by dropping all mixed squares terms. □

2.2 Proof of Theorem 1

Proof (*of Theorem* 1). For regimes p_1, p_3, p_4 and p_5, by telescoping inequalities from Lemmas 2, Lemma 4, Lemma 5 and Lemma 6, respectively, for $k = \{0, \ldots, N-1\}$ and taking the minimum over iterations differences norms:

$$F(x^0) - F(x^N) \geq p_i(\mu_1, L_1, \mu_2, L_2)N\frac{1}{2}\min_{0\leq k\leq N}\{\|x^k - x^{k+1}\|^2\},\ i = \{1,3,4,5\},$$

where p_i denotes the sum of the two coefficients multiplying $\frac{1}{2}\|x^k - x^{k+1}\|^2$ and $\frac{1}{2}\|x^{k+1} - x^{k+2}\|^2$ in each inequality. Writing (7) for the iterates x^N and x^{N+1}, we have $F(x^N) - F(x^{N+1}) \geq \frac{\mu_1+\mu_2}{2}\|x^N - x^{N+1}\|^2$. Then the rate results by summing it to the previous inequality and taking the minimum in the right-hand side. For regimes p_2 and p_6, by telescoping inequalities from Lemma 3 and 7, respectively, for $k = \{1, \ldots, N\}$ and taking the minimum over iterations:

$$F(x^1) - F(x^{N+1}) \geq p_i(\mu_1, L_1, \mu_2, L_2)N\frac{1}{2}\min_{0\leq k\leq N}\{\|x^k - x^{k+1}\|^2\},\ i = \{2,6\}.$$

Writing (7) for the iterates x^0 and x^1, summing it up and taking the minimum in the right-hand side leads to the rate. □

3 Tight Convergence Rates for Any Number of Iterations

Notation. We define $E_k : \mathbb{R}_{>0} \to \mathbb{R}_{>0}$, $E_k(z) := \sum_{j=1}^{2k} z^{-j}$, which is $\frac{-1+z^{-2k}}{1-z}$, if $z \neq 1$, and $2k$ is $z = 1$. Note that $E_0(z) = 0$. We denote $[z]_+ := \max\{0, z\}$.

3.1 Subdomains with Proved Rates

Theorem 2 *Let* $f_1 \in \mathcal{F}_{\mu_1,L_1}$ *and* $f_2 \in \mathcal{F}_{\mu_2,L_2}$, *with* $L_2 \in [0, \mu_1]$, $\mu_1 + \mu_2 > 0$ *and* $\mu_2 \in \left[\frac{-L_2\mu_1}{L_2+\mu_1}, L_2\right)$. *Then after* $N+1 \geq 2$ *iterations of DCA starting from* x^0 *it holds:* $\frac{1}{2}\|x^N - x^{N+1}\|^2 \leq \frac{F(x^0)-F(x^{N+1})}{\mu_1+\mu_2+\mu_1 E_N(\frac{L_2}{\mu_1})}$.

Proof. We take $\beta_1 = E_{k+1}(\frac{L_2}{\mu_1})$ and $\beta_2 = -1 + \frac{L_2}{\mu_1}E_{k+1}(\frac{L_2}{\mu_1})$ in (8), where $k = 0, \ldots, N-1$. After building the squares and simplifications we get:

$$\Delta F(x^k) \geq -E_k(\tfrac{L_2}{\mu_1})\tfrac{\mu_1}{2}\|\Delta x^k\|^2 + E_{k+1}(\tfrac{L_2}{\mu_1})\tfrac{\mu_1}{2}\|\Delta x^{k+1}\|^2 + \tfrac{\|G^k-\mu_1\Delta x^k\|^2}{2(L_1-\mu_1)} + \tfrac{(L_1+\mu_1)E_{k+1}(\frac{L_2}{\mu_1})\|G^{k+1}-\mu_1\Delta x^{k+1}\|^2}{2\mu_1(L_1-\mu_1)} + \tfrac{[-1+\frac{L_2+\mu_2}{\mu_1}E_{k+1}(\frac{L_2}{\mu_1})]\,\|G^{k+1}-L_2\Delta x^k\|^2}{2(L_2-\mu_2)}.$$

We have $-1+\frac{L_2+\mu_2}{\mu_1}E_{k+1}(\frac{L_2}{\mu_1}) \geq -1+\frac{L_2+\mu_2}{\mu_1}E_1[\frac{L_2}{\mu_1}] = \frac{\mu_2^2}{\mu_1(L_2-\mu_2)}(-T_1) \geq 0$, since $E_{k+1}[\frac{L_2}{\mu_1}]$ is increasing with k, thus the mixed terms can be dropped. Telescoping for $k = 0, \ldots, N-1$ and adding (B^k) with $k = N$ gives the conclusion. □

Theorem 3. *Let $f_1 \in \mathcal{F}_{\mu_1,L_1}$ and $f_2 \in \mathcal{F}_{\mu_2,L_2}$, with $\mu_1+\mu_2 > 0$ and $L_2+\mu_2 \leq 0$. Then after $N+1 \geq 2$ iterations of DCA starting from x^0 it holds $\frac{1}{2}\|x^N - x^{N+1}\|^2 \leq \frac{F(x^0)-F(x^{N+1})}{\mu_1+\mu_2+\mu_1 E_N(\frac{\mu_2}{\mu_1})}$.*

Proof. We take $\beta_1 = E_{k+1}(\frac{\mu_2}{\mu_1})$ and $\beta_2 = -\frac{\mu_2}{\mu_1}E_{k+1}(\frac{\mu_2}{\mu_1})$ in (8), where $k = 0, \ldots, N-1$. After building the squares and simplifications we get:

$$\Delta F(x^k) \geq -E_k(\tfrac{\mu_2}{\mu_1})\tfrac{\mu_1}{2}\|\Delta x^k\|^2 + E_{k+1}(\tfrac{\mu_2}{\mu_1})\tfrac{\mu_1}{2}\|\Delta x^{k+1}\|^2 + \tfrac{\|G^k-\mu_1\Delta x^k\|^2}{2(L_1-\mu_1)} + \tfrac{(L_1+\mu_1)E_{k+1}(\frac{\mu_2}{\mu_1})\|G^{k+1}-\mu_1\Delta x^{k+1}\|^2}{2\mu_1(L_1-\mu_1)} + \tfrac{[1-\frac{L_2+\mu_2}{\mu_1}E_{k+1}(\frac{\mu_2}{\mu_1})]\|G^{k+1}-\mu_2\Delta x^k\|^2}{2(L_2-\mu_2)}.$$

$E_{k+1}(\frac{\mu_2}{\mu_1})$ is positive and monotonically increasing with index k as $\mu_1+\mu_2 > 0$. Then due to $L_2+\mu_2 \leq 0$, we can drop all mixed squares. Telescoping for $k = 0, \ldots, N-1$ and adding (B^k) with $k = N$ gives the conclusion. □

3.2 Conjectures for Any Number of Iterations

Conjecture 1. Let $f_1 \in \mathcal{F}_{\mu_1,L_1}$ and $f_2 \in \mathcal{F}_{\mu_2,L_2}$, with $\mu_1 \geq 0$ and $\mu_1+\mu_2 > 0$. Consider $N+1 \geq 2$ iterations of DCA starting from x^0.
If $\mu_2 < 0 < \mu_1 < L_2$, thus F is nonconvex-nonconcave, then:

$$\tfrac{\mu_1}{2}\min_{0\leq k\leq N}\left\{\|x^k - x^{k+1}\|^2\right\} \leq \frac{F(x^0)-F(x^{N+1})}{1+\frac{\mu_2}{\mu_1}+\min\left\{P_N(\frac{L_2}{\mu_1},\frac{\mu_2}{\mu_1}),\, E_N(\frac{\mu_2}{\mu_1})\right\}}, \tag{9}$$

where $P_N(\eta,\rho) := \frac{(1+\eta)(1+\rho)}{\eta+\rho}\big(N + \frac{(1-\eta)(1-\rho)}{\eta-\rho}\sum_{k=0}^{N}[E_k(\eta) - E_k(\rho)]_+\big)$.
If $L_2 \leq \mu_1$ and $\mu_1 > 0$, thus F is (strongly) convex, then:

$$\tfrac{\mu_1}{2}\min_{0\leq k\leq N}\left\{\|x^k - x^{k+1}\|^2\right\} \leq \frac{F(x^0)-F(x^{N+1})}{1+\frac{\mu_2}{\mu_1}+\min\left\{E_N(\frac{L_2}{\mu_1}),\, E_N(\frac{\mu_2}{\mu_1})\right\}}. \tag{10}$$

Finally, if $0 \leq \mu_1 < L_1 \leq \mu_2 < L_2$, thus F is (strongly) concave, then:

$$\tfrac{\mu_2}{2}\min_{0\leq k\leq N}\left\{\|x^k - x^{k+1}\|^2\right\} \leq \frac{F(x^0)-F(x^{N+1})}{1+\frac{\mu_1}{\mu_2}+\min\left\{E_N(\frac{L_1}{\mu_2}),\, E_N(\frac{\mu_1}{\mu_2})\right\}}. \tag{11}$$

For $N = 1$, Conjecture 1 includes the rates corresponding to regimes p_3 and p_5 in (9) when F is nonconvex, of regimes p_4 and p_5 in (10) when F is (strongly) convex and of regime p_6 in (11) when F is concave. The rate in (10) includes the ones proved in Theorem 2 (within the domain of regime p_4) and in Theorem 3. The expressions of the rates (9) and (10) are independent of L_1, hence they also hold when f_1 is ***nonsmooth***, i.e., $L_1 = \infty$.

Conjecture 1 is based on the exact rates developed for gradient descent (GD) derived in [18] and is confirmed by PEP numerical experiments. GD is a particular case of the proximal gradient descent (PGD), which is iteration equivalent with DCA when f_2 is smooth.

4 Deriving Rates for Proximal Gradient Descent (PGD)

Assume $F = \varphi + h$, with $\varphi \in \mathcal{F}_{\mu_\varphi, L_\varphi}$ smooth, $L_\varphi \in (0, \infty)$, $\mu_\varphi \in (-\infty, L_\varphi)$, and $h \in \mathcal{F}_{\mu_h, L_h}$ proper, closed, l.s.c. and convex, such that $\mu_h \geq 0$ and $L_h \in (0, \infty]$.

The PGD iteration with stepsize $\gamma > 0$, starting from x^k, reads $x^{k+1} = \operatorname{argmin}_{w \in \mathbb{R}^d} \{h(w) + \frac{1}{2\gamma}\|w - x^k + \gamma \nabla \varphi(x^k)\|^2\}$. This iteration is exactly the DCA one applied to $f_1 = h + \frac{\|\cdot\|^2}{2\gamma}$ and $f_2 = \frac{\|\cdot\|^2}{2\gamma} - \varphi$, as shown in [13, Section 3.3.4] and [19, Proposition 3]. Consequently, all the rates developed for DCA readily translate to rates for PGD using the mapping $\mu_1 = \gamma^{-1} + \mu_h$, $L_1 = \gamma^{-1} + L_h$, $\mu_2 = \gamma^{-1} - L_\varphi$, $L_2 = \gamma^{-1} - \mu_\varphi$. The condition $\mu_1 + \mu_2 > 0$ translates to the standard upper limit of $\gamma < \frac{2}{L_\varphi - \mu_h}$ and $\mu_2 < 0$ corresponds to stepsizes $\gamma > \frac{1}{L_\varphi}$.

When φ is convex, [24] shows various tight rates employing different performance metrics, including gradient residual for stepsize $\gamma = \frac{1}{L_\varphi}$, but not the gradient mapping. In the gradient residual, for φ nonconvex, with $\mu_\varphi = -L_\varphi$ and $\mu_h = 0$, $L_h = \infty$, the tight rate for stepsizes shorter than $\frac{1}{L_\varphi}$ is given in [1, Proposition 8.11]. Sublinear rates for the same criterion are given in [19].

The gradient descent (GD) is obtained by setting $h = 0$, hence $\mu_h = L_h = 0$. Convergence rates on the best gradient norm are proved in [18] for all curvatures of φ. These rates are a particular case of Conjecture 1 and served as a base to conjecture the tight expressions, since the projection / proximal step is not modifying the behavior of the worst-case function.

To get exact rates from DCA, one just has to plug in the expressions of the curvatures μ_i, L_i, $i = \{1, 2\}$, written with respect to curvatures of φ and h, in Theorem 1 and Conjecture 1. The convexity of h implies $\mu_h \geq 0$, thus $\mu_1 > 0$. Then regime p_1 corresponds to taking stepsizes $\gamma \leq \frac{1}{L_\varphi}$ for a splitting with φ and F nonconvex. Within the regimes p_3 and p_4, the stepsize γ lies in the interval $[\frac{1}{L_\varphi}, \bar{\gamma}^1(L_\varphi, \mu_\varphi, \mu_h)]$, where $\bar{\gamma}^1(L_\varphi, \mu_\varphi, \mu_h)$ is a threshold with an analytical expression obtained through the condition $\mu_2 = \frac{-L_2 \mu_1}{L_2 + \mu_1}$. Regime p_3 holds when φ and F are nonconvex. When $\mu_h = 0$ and $\mu_\varphi = -L_\varphi$, then $\bar{\gamma}^1 = \frac{\sqrt{3}}{L_\varphi}$, as also obtained in [2] for the GD method. Regime p_4 holds when F is strongly convex and $\mu_\varphi + \mu_h \geq 0$; in particular, this includes the case of φ being strongly convex. For stepsizes $\gamma \in (\bar{\gamma}^1(L_\varphi, \mu_\varphi, \mu_h), \frac{2}{L_\varphi + \mu_\varphi})$, regime p_5 is active if

F is nonconcave. Regimes p_2 and p_6, respectively, correspond to using stepsizes $\gamma \leq \frac{1}{L_\varphi}$ for φ concave and F nonconcave or concave, respectively.

Lemma 8 provides an example of convergence rate translation between PGD and DCA, corresponding to regime p_3 from Lemma 4, assuming the standard setup of PGD with $\mu_h = 0$ and $L_h = \infty$.

Lemma 8. *Let $\varphi \in \mathcal{F}_{\mu_\varphi, L_\varphi}$, $h \in \mathcal{F}_{0,\infty}$ and $\gamma \in [\frac{1}{L_\varphi}, \frac{2}{L_\varphi})$. Then after two iterations of PGD with stepsize γ connecting x^k, x^{k+1} and x^{k+2} it holds*

$$\Delta F(x^k) \geq \frac{[(2-L_\varphi)(2-\mu_\varphi)-1]}{\gamma(2-L_\varphi-\mu_\varphi)} \frac{1}{2}\|\Delta x^k\|^2 + \frac{1}{\gamma(2-L_\varphi-\mu_\varphi)} \frac{1}{2}\|\Delta x^{k+1}\|^2.$$

Lemma 8 is the extension to PGD of the same result obtained for GD in [19, Lemma 4.3]. Therein, it is involved in deriving a stepsize schedule gradually increasing from $\bar{\gamma}^1(L_\varphi, \mu_\varphi, 0)$ towards $\frac{2}{L_\varphi+\mu_\varphi}$ (if φ is convex), which gives a better worst-case guarantee than the best constant stepsize. Based on Lemma 8, the same schedule should give a similar performance for PGD. Further, we show how it may improve the DCA convergence.

Curvature Shifting and Schedules for DCA

Proposition 2. *One PGD iteration with stepsize γ on the splitting $F = \varphi + h$ produces the same iterate as one iteration with stepsize $\tilde{\gamma}$ on $F = \tilde{\varphi} + \tilde{h}$, with $\tilde{\varphi} = \varphi + \lambda\frac{\|\cdot\|^2}{2}$, $\tilde{h} = h - \lambda\frac{\|\cdot\|^2}{2}$, $\tilde{\gamma}^{-1} = \gamma^{-1} - \lambda$, where $\lambda < \gamma^{-1}$.*

This relates to curvature shifting in DCA: $f_i^\lambda := f_i - \frac{\lambda\|\cdot\|^2}{2}$, $i = \{1, 2\}$, where λ is a scalar. Based on a single-iteration analysis, [19] studies the best curvature shifting, observing that it usually reduces the convexity of both . The procedure can be improved, as (optimized) stepsize schedules $\{\gamma^k(L_\varphi, \mu_\varphi)\}$ for PGD can be translated to DCA. Consider the standard setting of PGD with $L_\varphi > 0$, $\mu_h = 0$, $L_h = \infty$, assume $\mu_1 > 0$ and choose $\gamma = \mu_1^{-1}$. Using Proposition 2, a curvature splitting schedule for DCA is defined as $\{\lambda^k\} = \mu_1 - \{\gamma^k(\mu_1 - L_2, \mu_1 - \mu_2)\}$.

5 Conclusion

We studied rates in the gradient mapping for DCA and PGD. Examining the behavior of DCA after two iterations, we identified six regimes partitioning the parameter space, and proved the corresponding six tight sublinear convergence rates. Three of those sublinear rates hold for any number of iterations and, for the other regimes, we conjectured their exact convergence rates. We also show how PGD rates follow directly from DCA ones. This connection suggests the potential for an optimized splitting of the objective function used in DCA, possibly varying at each iteration, inspired by stepsize schedules used for (proximal) gradient descent. We leave the practical testing of those as future work.

Acknowledgements. This research was funded within the framework of the Global PhD Partnership KU Leuven - UCLouvain.

References

1. Abbaszadehpeivasti, H.: Performance analysis of optimization methods for machine learning. CentER dissertation. CentER, Tilburg University (2024)
2. Abbaszadehpeivasti, H., Klerk, E.D., Zamani, M.: The exact worst-case convergence rate of the gradient method with fixed step lengths for l-smooth functions. Optimization Lett. (2021)
3. Abbaszadehpeivasti, H., Klerk, E.D., Zamani, M.: On the rate of convergence of the difference-of-convex algorithm (DCA). J. Optimiz. Theory and Appl. (2023)
4. Bauschke, H.H., Moursi, W.M., Wang, X.: Generalized monotone operators and their averaged resolvents. Math. Program. **189**(1), 55–74 (2021)
5. Davis, D., Drusvyatskiy, D.: Stochastic model-based minimization of weakly convex functions. SIAM J. Optim. **29**(1), 207–239 (2019)
6. Dinh, T.P., Le Thi, H. A.: Convex analysis approach to d.c. programming: theory, algorithm and applications. In: Acta Matehmatica Vietnamica, vol. 22 (1997)
7. Drori, Y., Teboulle, M.: Performance of first-order methods for smooth convex minimization: a novel approach. Math. Program. **145**(1–2), 451–482 (2014)
8. Faust, O., Fawzi, H., Saunderson, J.: A Bregman divergence view on the Difference-of-Convex Algorithm. In: Proceedings of The 26th International Conference on Artificial Intelligence and Statistics, volume 206 of Proceedings of Machine Learning Research, pp. 3427–3439. PMLR, 25–27 (2023)
9. Goujaud, B., Moucer, C., Glineur, F., Hendrickx, J.M., Taylor, A.B., Dieuleveut, A.: PEPit: computer-assisted worst-case analyses of first-order optimization methods in python. Math. Program. Comput. **16**(3), 337–367 (2024)
10. Horst, R., Thoai, N.V.: DC Programming: overview. J. Optim. Theory Appl. **103**(1), 1–43 (1999)
11. Lanckriet, G., Sriperumbudur, B. K.: On the convergence of the concave-convex procedure. Advances in Neural Information Processing Systems, 22 (2009)
12. Le Thi, H. A., Huynh, V. N., Dinh, T. P.: Convergence analysis of difference-of-convex algorithm with subanalytic data. J. Optimiz. Theory Appl. 179(1):103–126 (2018)
13. Hoai An Le Thi and Tao Pham Dinh: DC programming and DCA: thirty years of developments. Math. Program. **169**(1), 5–68 (2018)
14. Lipp, T., Boyd, S.: Variations and extension of the convex-concave procedure. Optim. Eng. **17**(2), 263–287 (2016)
15. Nesterov, Y.: Gradient methods for minimizing composite functions. Math. Program. **140**(1), 125–161 (2013)
16. Lectures on Convex Optimization. SOIA, vol. 137. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-91578-4_7
17. Rockafellar, R.T., Wets, R J.-B.: Variational Analysis. Springer Verlag, Heidelberg, Berlin, New York (1998)
18. Rotaru, T., Glineur, F., Patrinos, P.: Exact worst-case convergence rates of gradient descent: a complete analysis for all constant stepsizes over nonconvex and convex functions. arXiv preprint arXiv:2406.17506 (2024)
19. Rotaru, T., Patrinos, P., Glineur, F.: Tight analysis of difference-of-convex algorithm (DCA) improves convergence rates for proximal gradient descent. In: AIStats 2025 Proceedings (2025)
20. Tao, P.D., An, L.T. H.: A D.C. optimization algorithm for solving the trust-region subproblem. SIAM J. Optimiz. 8(2):476–505 (1998)

21. Taylor, A., Hendrickx, J.T., Glineur, F.: Smooth strongly convex interpolation and exact worst-case performance of first-order methods. Mathematical Programming, 161(1-2):307–345 (2017)
22. Taylor, A.B., Hendrickx, J.M., Glineur, F.: Exact worst-case performance of first-order methods for composite convex optimization. SIAM J. Optim. **27**(3), 1283–1313 (2017)
23. Taylor, A. B., Hendrickx, J. M., Glineur, F.: Performance estimation toolbox (PESTO): automated worst-case analysis of first-order optimization methods. In: 2017 IEEE 56th Annual Conference on Decision and Control (CDC), pp. 1278–1283 (2017)
24. Taylor, A.B., Hendrickx, J.M., Glineur, F.: Exact worst-case convergence rates of the proximal gradient method for composite convex minimization. J. Optim. Theory Appl. **178**(2), 455–476 (2018)
25. Toland, J.F.: A duality principle for non-convex optimisation and the calculus of variations. Arch. Ration. Mech. Anal. **71**(1), 41–61 (1979)
26. Yao, C., Jiang, X.: A globally convergent difference-of-convex algorithmic framework and application to log-determinant optimization problems (2023)
27. Yuille, A. L., Rangarajan, A.: The concave-convex procedure (CCCP). In: Dietterich, T., Becker, S., Ghahramani, Z., (eds) Advances in Neural Information Processing Systems, volume 14. MIT Press (2001)
28. Yurtsever, A., Sra, S.: CCCP is Frank-Wolfe in disguise. Adv. Neural Inf. Proc. Syst. **35**, 35352–35364 (2022)