Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Low-Rank Curvature Corrections for Interior-Point DC Algorithms in Large-Scale Nonconvex Quadratic Programming.

François Bertrand Akoa

francoisakoa@gmail.com

Maatai-lab https://orcid.org/0009-0007-2652-3386

Research Article

Keywords:

Posted Date: November 25th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11833500/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: The authors declare no competing interests.

Low-Rank Curvature Corrections for Interior-Point DC Algorithms in Large-Scale Nonconvex Quadratic Programming.

François Bertrand Akoa

MAATAI-LAB, France
Corresponding author: francois.akoa@maatai-lab.com

Abstract

Interior-point difference-of-convex (DC) algorithms for nonconvex quadratic programs convexify the indefinite Hessian with a uniform scalar shift, which also stiffens benign positive-curvature directions. We introduce LazyIPDCA, which replaces this shift by an event-driven metric combining sparse rank-one corrections along detected negative-curvature directions with a residual scalar safeguard. A rank-sized Karush–Kuhn–Tucker (KKT) border preserves the sparse base factorization. For a convergence-certified variant, the accepted sequence converges to a KKT point, with an $O(N^{-1/2})$ best-stationarity bound and local linear convergence. On 37 CUTEst instances, the low-rank metric reduces total time from 1409 s (fixed scalar) and 1713 s (adaptive scalar) to 381 s, while returning a best-or-tied objective on 33 instances versus 18 and 32. Against Ipopt and QPALM, LazyIPDCA is best or tied on 34 instances against each and fastest among quality-eligible solvers on 30, whereas QPALM attains the most first-order certificates. On 17 problems with 100,000–131,073 variables, it returns feasible points throughout while retaining at most seven directions; it also certifies all 72 complementarity reformulations up to five million variables.

Keywords: nonconvex quadratic programming; DC programming; interior-point methods; low-rank curvature correction; large-scale optimization

Mathematics Subject Classification: 90C20, 90C26, 90C51

1 Introduction

Consider the linearly constrained nonconvex quadratic program

$$\min_{x\in\mathbb{R}^n} \quad f(x) = \frac{1}{2}x^\top Qx + q^\top x \qquad \text{s.t.} \qquad Ax = b, \quad Cx \geq b_I, \quad \ell \leq x \leq u, \tag{1}$$

where $Q = Q^\top$ is not assumed positive semidefinite. Problems of this form occur as stand-alone models and as subproblems inside broader algorithms. Examples include indefinite SQP and direct optimal-control subproblems, where preserving the block sparsity of the discretized dynamics is essential [36], as well as quadratic reformulations of complementarity systems. The difficulty addressed here is not merely indefiniteness. At large scale, a convexification that treats all directions equally can erase useful positive curvature, worsen conditioning and force the interior-point linear algebra to work with a model substantially farther from the original objective than necessary.

Classical DC programming writes the objective as a difference of convex functions and solves a sequence of convex subproblems [18, 27]. The earlier IPDCA framework combined this idea with primal–dual interior-point steps for large nonconvex quadratic programs [28]. The present work keeps the same broad principle but changes the convexification geometry. Instead of relying on a single isotropic shift, it retains a small set of negative-curvature

directions and corrects them explicitly. A scalar term remains as a safeguard for curvature not represented by the retained low-rank model.

This design is motivated by a simple observation. If Q is indefinite only on a low-dimensional subspace, the scalar regularization $Q + \rho I \succeq 0$ modifies all n directions even though only a few need to be changed. In contrast, a matrix of the form

$$H_k = \sum_{j=1}^{r_k} \beta_{k,j} a_{k,j} a_{k,j}^\top, \qquad M_k := H_k + \rho_k I, \qquad \rho_k \geq 0, \quad \beta_{k,j} > 0, \tag{2}$$

can target detected negative curvature while keeping $r_k \ll n$ on favourable instances. The term *lazy* refers to the event-driven construction of H_k: spectral searches and new KKT borders are invoked only at prescribed curvature events and only when a candidate passes the selection and model checks, rather than at every interior-point iteration. The central questions are then: how should the directions be selected; how can the low-rank correction be inserted without destroying the sparse interior-point structure; and what convergence statement survives when the metric changes from one DCA iteration to the next?

The contributions are fourfold. First, we give an adaptive selection rule that combines spectral lift, directional activity and sparsity-aware penalties. Second, we insert the low-rank correction as a bordered KKT block, so the core sparse factorization structure is retained. Third, under a compact initial lower level set, a uniform curvature margin, bounded metrics and a relative inner-solve criterion, we prove convergence of the whole accepted sequence to a KKT point. The same argument yields a global $O(N^{-1/2})$ bound for a stationarity measure, and the piecewise linear–quadratic structure gives local linear convergence. Monotone growth of the metric is not required. Fourth, we report numerical evidence on CUTEst, including direct comparisons with Ipopt and QPALM, a 17-instance large-scale panel with dimensions from 100,000 to 131,073, a targeted sensitivity study with a binding-rank experiment and repeated timing measurements, and 72 structured complementarity reformulations up to five million variables.

1.1 Relation to recent DC and large-scale optimization developments

The DCA literature has expanded substantially beyond the classical scheme; recent surveys emphasize extrapolation, line-search boosting, variable/Bregman metrics, stochastic variants and inexact subproblem solves as distinct development axes [19]. These directions are relevant here, but they address different algorithmic questions.

Extrapolation changes how the next iterate is generated from previous iterates. The proximal DCA with extrapolation of Wen, Chen and Pong [38] augments proximal DCA with extrapolated points; Lu, Zhou and Sun [23] develop enhanced proximal DC schemes with extrapolation; and Phan and Le Thi [29] provide a further extrapolated DCA for a broader nonsmooth setting. These methods modify iterate generation, whereas the present low-rank mechanism changes the curvature matrix of the convex DCA model and therefore the KKT system solved at the current iterate.

A second line accelerates a DCA step by an additional search. The boosted DCA for smooth functions [5] and its nonsmooth extension of Aragón Artacho and Vuong [7] use a line search beyond the basic DCA point. This idea has since been extended to linearly constrained DC programs [4], to nondifferentiable DC components with a nonmonotone line search [13], and to broader double-proximal splitting schemes [6]. Recent performance-estimation analysis also gives worst-case convergence-rate comparisons between DCA and boosted DCA under bounded-curvature assumptions [2]. These developments are especially relevant to the constrained QP setting because boosting can be compatible with linear constraints, but they act on the step taken after a DC subproblem has been formed rather than on the spectral allocation of curvature inside that subproblem.

A third line is closer to the present work because it changes local geometry or the accuracy with which local models are solved. Variable-metric and Nesterov-extrapolated proximal DCA [40], Bregman proximal DCA [34], inexact successive quadratic DC approximations [22], inexact proximal DC Newton-type methods [26], and recent inexact Bregman DCA with relative stopping criteria [39] all make the metric or inner accuracy part of the algorithmic design. Stochastic DCA with variance reduction addresses another scalability bottleneck when the objective is a large finite sum [17]. More broadly, algorithms that explicitly exploit negative-curvature information provide a complementary view of why directional curvature is valuable [11].

Nonconvex QP is also approached through proximal augmented-Lagrangian, trust-region and active-set ideas. QPALM applies proximal augmented-Lagrangian iterations with semismooth Newton inner solves and proves stationary-point convergence with an R-linear rate [16]. Krylov and Lanczos methods are classical tools for large trust-region subproblems [15], and active-set schemes can exploit exact or nonglobal trust-region solutions inside constrained quadratic optimization [33]. These methods provide important algorithmic and computational reference points. LazyIPDCA differs in adapting the DC convexification metric itself and carrying the selected directions as a border of a primal–dual interior-point system.

GALAHAD QPB is a closely related primal–dual interior-point method for linearly constrained nonconvex quadratic programs, with trust-region globalization and iterative linear algebra [9, 10]. It is included here as a methodological reference. The numerical benchmark uses Ipopt and QPALM because complete matched campaigns from the common starting points, together with the original-coordinate primal–dual records needed for the KKT1 assessment, were available for all reported panels.

The low-rank mechanism studied here is complementary to these accelerations. It changes the *DC convexification metric itself* by replacing a purely isotropic correction with $H_k + \rho_k I$, where H_k is assembled from a small number of curvature directions and is tied to the sparse KKT factorization. Extrapolation can be combined with metric adaptation, while the present study isolates the curvature-allocation mechanism. Likewise, sparse convexification for optimal-control QPs uses a prescribed block structure [36], while regularized interior-point methods adapt primal/dual numerical regularization for convex QPs [30, 31]. Here the object being adapted is specifically the DC regularizer needed to convexify an indefinite QP while preserving useful curvature of Q.

Table 1: Recent DCA developments alter different parts of the algorithm. The present low-rank method modifies the convexification metric and is therefore complementary to extrapolation and boosted DCA.

Direction	Representative work	Main object changed
Extrapolation	[23, 29, 38]	Iterate generation from previous points
Boosted DCA	[4, 5, 13]	Additional step or line search after a DCA point
Variable / Bregman metrics	[34, 40]	Proximal geometry used to define the local model
Inexact subproblem solves	[22, 26, 39]	Accuracy and stopping rule for the convex/local subproblem
Present method	this paper	Directional convexification through $M_k = H_k + \rho_k I$ and a bordered KKT formulation

The remainder is organized as follows. Sections 2–3 define the variable metric and its event-driven low-rank update. Section 4 gives the bordered primal–dual realization and the

accepted-step conditions. Section 5 proves whole-sequence convergence and rates. Sections 6–7 state the numerical protocol and summarize the experiments; instance-level tables are placed in the supplementary material. The final sections discuss scope, limitations and conclusions.

2 DC model and variable-metric convexification

2.1 A metric DC decomposition

Let $\mathcal{K}$ denote the polyhedron in (1), and define the extended objective

$$F(x) := \frac{1}{2}x^\top Qx + q^\top x + \delta_{\mathcal{K}}(x). \tag{3}$$

Thus $F = f$ on $\mathcal{K}$ and $F = +\infty$ outside $\mathcal{K}$. Throughout, ∂ denotes the limiting subdifferential. For this smooth-plus-polyhedral objective,

$$\partial F(x) = Qx + q + N_{\mathcal{K}}(x), \qquad x \in \mathcal{K}.$$

The non-isotropic correction is kept separate from the scalar safeguard throughout the paper:

$$H_k = \sum_{j=1}^{r_k} \beta_{k,j} a_{k,j} a_{k,j}^\top \succeq 0, \qquad M_k := H_k + \rho_k I \succeq 0, \tag{4}$$

with $\rho_k \geq 0$. The analysis requires each accepted model to be convex on $\mathcal{K}$ and to admit the accepted subproblem solution used below. In full-space mode the sufficient condition

$$B_k := Q + M_k = Q + H_k + \rho_k I \succeq 0 \tag{5}$$

guarantees this property. A spectral test restricted to the current active face is used to select curvature directions; by itself, it is not a global positive-definiteness certificate. Semidefiniteness is sufficient for the DC decomposition itself. The convergence and rate analysis in Section 5 uses the stronger, explicitly stated condition $Q + M_k \succeq \mu I$ on accepted models for some uniform $\mu > 0$. Then

$$F(x) = \underbrace{\frac{1}{2}x^\top B_k x + q^\top x + \delta_{\mathcal{K}}(x)}_{g_k(x)} - \underbrace{\frac{1}{2}x^\top M_k x}_{h_k(x)}, \tag{6}$$

where $\delta_{\mathcal{K}}$ is the indicator of the feasible set. Linearizing h_k at x_k gives the convex DCA model

$$\min_{x \in \mathcal{K}} \quad \frac{1}{2}x^\top B_k x + q^\top x - x^\top M_k x_k, \tag{7}$$

equivalently $\min_{x\in\mathcal{K}} \Phi_k(x)$, where

$$\Phi_k(x) := F(x) + \tfrac{1}{2}(x - x_k)^\top M_k (x - x_k). \tag{8}$$

This definition is also used in the toy figures below. If this subproblem is solved exactly, its first-order condition is

$$0 \in Qx_{k+1} + q + M_k(x_{k+1} - x_k) + N_{\mathcal{K}}(x_{k+1}). \tag{9}$$

The practical algorithm solves (7) inexactly with a primal–dual interior-point method. The notation H_k henceforth always means the low-rank part, while M_k means the complete DC metric.

2.2 Relation with the fixed-metric IPDCA decompositions

In the proximal version of IPDCA, the complete metric is $M_k \equiv \rho I$: the convex Hessian is $Q + \rho I$, so the curvature of Q is retained but every direction pays the same shift. The projection version instead takes $Q_1 = \rho I$ and $Q_2 = \rho I - Q$; this yields a diagonal primal Hessian block but removes the anisotropic curvature of Q from the convex model. The present construction is a variable-metric extension of the proximal, not the projection, decomposition.

2.3 Why a low-rank correction

A purely scalar model chooses $H_k = 0$ and $M_k = \rho_k I$. This is simple and robust, but its largest required value is dictated by the most negative curvature direction. Every positive-curvature direction is shifted by the same amount. The proposed metric instead has the form (2). If the retained directions span most of the negative eigenspace, the residual scalar shift may become much smaller while the model remains convex.

The low-rank representation also gives a natural memory mechanism. Directions that continue to certify difficult curvature can be retained, while obsolete or redundant directions can be dropped under a rank budget. The method does not require the rank to grow monotonically and does not equate the number of stored directions with the exact matrix rank; it only guarantees $\operatorname{rank}(H_k) \le r_k$.

Rank requirement without the scalar safeguard. Let $r_- := \dim \mathcal{N}_-(Q)$ denote the dimension of the negative eigenspace of Q. If the scalar safeguard is absent and a positive-semidefinite correction H satisfies $Q + H \succeq 0$, then $\operatorname{rank}(H) \ge r_-$. Otherwise $\mathcal{N}_-(Q) \cap \ker H$ would contain a nonzero vector v, giving $v^\top (Q + H)v = v^\top Q v < 0$. This bound concerns a purely low-rank convexification. It does not impose a lower bound on the rank retained by LazyIPDCA, because the scalar term $\rho_k I$ can stabilize negative-curvature directions not represented in H_k.

2.4 Face-aware curvature detection

Simple bounds and active linear constraints can remove curvature directions that are irrelevant to feasible motion. The selection mechanism therefore uses a clearly identified spectral operator. In full-space mode it is $Q + H_k$; when the current free/affine face is enforced and Z_k spans its tangent space, it is

$$\mathcal{Q}_k^{\mathrm{sel}} = Z_k^\top (Q + H_k) Z_k. \tag{10}$$

The leftmost quantities λ_0 and λ_i^+ used by the selection score refer to this operator (or its full-space counterpart), before the residual scalar safeguard $\rho_k I$ is applied. This curvature estimate used for direction selection is not itself a global convexity certificate. The convergence analysis therefore uses a separate certified full-space condition on every accepted model; no uniform margin is inferred from face-reduced Ritz values or from an unverified scalar safety factor. A Krylov realization can apply the face projector without forming Z_k explicitly; the numerical realization used in the experiments is specified in Section 6.1.

Notation used below. The feasible set in (1) is $\mathcal{K}$; m_E and m_I denote the numbers of equality and inequality constraints. The low-rank curvature term is H_k, the complete DC metric is $M_k = H_k + \rho_k I$, and $\overline{M}$ is a uniform metric-norm bound used only in the convergence analysis. The indicator of $\mathcal{K}$ is $\delta_{\mathcal{K}}$, whereas δ_d denotes numerical dual regularization in the KKT system. The objective model is Φ_k from (8). These conventions avoid reusing solver-dimension symbols for analytical constants.

3 Adaptive low-rank DC regularization

3.1 Scalar safeguard

Let λ_0 be the current leftmost-curvature estimate before a new candidate is inserted. We write the scalar requirement as

$$\rho_{\mathrm{sc}} = \tau_\rho \,|\min(0, \lambda_0)|, \qquad \tau_\rho > 1. \tag{11}$$

The factor τ_ρ is a numerical safety parameter, not a theorem-level curvature constant. After the low-rank part has been updated, the scalar component may decrease subject to a residual floor and may increase again after unsuccessful or numerically unsafe steps. Scalar and low-rank components are therefore updated jointly rather than independently. The convergence analysis below uses only convexity of accepted models and boundedness of the resulting metrics; it does not prescribe a particular value of τ_ρ.

3.2 Candidate generation and sparsification

At a low-rank update, let $p_{\max} \in \mathbb{N}$ denote the maximum number of negative-curvature Ritz candidates considered simultaneously. A normalized candidate is truncated to the smallest support retaining at least a fraction $\tau_{\mathrm{supp}} \in (0, 1]$ of its squared Euclidean norm, subject to a support cap $s_{\max} \in \mathbb{N}$. These parameters control the amount of spectral information retained and the size of the KKT border; their numerical values belong to the experimental setting, not to the definition of the convergence model.

3.3 Selection criteria

A curvature search may return several usable negative-curvature directions. Three spectral selectors clarify the mechanism before the composite rule used by LazyIPDCA.

Rayleigh rule.
Select the candidate with the smallest post-sparsification Rayleigh quotient

$$\varrho_i = \frac{a_i^\top (Q + H_k) a_i}{a_i^\top a_i}.$$

It needs only one sparse matrix application per candidate.

Curvature-lift rule.
Temporarily add $\beta_i a_i a_i^\top$, re-estimate $\lambda_i^+ = \lambda_{\min}(Q + H_k + \beta_i a_i a_i^\top)$ and select the largest λ_i^+. This directly targets the remaining scalar correction.

Lift-per-weight rule.
Select the largest $(\lambda_i^+ - \lambda_{\min}(Q + H_k))/\beta_i$, balancing spectral lift against the retained rank-one weight.

The two lift rules require a fresh leftmost-curvature estimate for each temporarily augmented model, which is why the candidate pool is kept finite.

For a normalized sparsified candidate with $\varrho_i < 0$, its strength is parameterized as

$$\beta_i = s_\beta \,(-\varrho_i + \delta_\beta), \qquad \delta_\beta > 0, \quad s_\beta > 0. \tag{12}$$

The offset δ_β and multiplier s_β are numerical parameters.

Proposition 1 (Curvature lift and residual scalar shift)**.** *Define*

$$\rho_{\text{eff}}(H) := \max\{0, -\lambda_{\min}(Q + H)\},$$

the smallest additional *scalar shift, beyond the low-rank correction H, making $Q + H + \rho I$ positive semidefinite. For a fixed candidate set, exact curvature-lift selection minimizes $\rho_{\text{eff}}(H_k + \beta_i a_i a_i^\top)$. With approximate eigenvalue estimates, the same statement holds for the corresponding estimated residual shift.*

Proof. For candidate i, $\rho_{\text{eff}}(H_k + \beta_i a_i a_i^\top) = \max\{0, -\lambda_i^+\}$, which is decreasing in λ_i^+ while the residual shift is positive. Candidates with $\lambda_i^+ \geq 0$ all attain the minimum zero. □

When several corrections are requested, selection is greedy: after retaining the best candidate, the remaining candidates are re-scored against the updated H_k. This avoids treating correlated or nearly parallel Ritz directions as independent spectral gains.

Composite-score selection rule. Let $\varepsilon_\lambda > 0$ be the negative-curvature detection tolerance and set

$$s_\lambda = \max\{\varepsilon_\lambda, |\min(\lambda_0, -\varepsilon_\lambda)|\}, \qquad \ell_i = \frac{\max\{0, \lambda_i^+ - \lambda_0\}}{s_\lambda}, \qquad \widehat{\beta}_i = \frac{\beta_i}{s_\lambda}. \tag{13}$$

Let $\widehat{g}_k$ denote the normalized objective gradient projected on the current free face and $\widehat{s}_k$ the normalized last accepted DCA displacement when available. Define

$$\alpha_i = \max\{|a_i^\top \widehat{g}_k|, |a_i^\top \widehat{s}_k|\}, \qquad \chi_i = \min\{1, \sqrt{n_{\text{free}}}\, \alpha_i\}, \tag{14}$$

$$s_i = \frac{\text{nnz}(a_i)}{n}, \qquad e_i = \frac{\max\{0, \|a_i\|_1^2 - 1\}}{\max\{1, \text{nnz}(a_i) - 1\}}. \tag{15}$$

The scalar requirement after temporary insertion is

$$\rho_i^{\text{sp}} = \tau_\rho |\min(0, \lambda_i^+)|. \tag{16}$$

For $\rho_{\text{sc}} > 0$ define

$$\theta_i = \frac{\rho_i^{\text{sp}} + \beta_i \alpha_i^2}{\rho_{\text{sc}}}, \tag{17}$$

and set $\theta_i = 1$ when $\rho_{\text{sc}} = 0$.

With nonnegative score weights $c_\chi, c_\theta, c_s, c_e, c_\beta$, the composite score is

$$S_i = \ell_i(1 + c_\chi \chi_i) + c_\theta \min\{1, \max(0, \theta_i)\} - c_s s_i^2 - c_e e_i - c_\beta \widehat{\beta}_i. \tag{18}$$

The candidate is admissible when

$$\ell_i \geq \tau_\ell, \qquad \sqrt{n_{\text{free}}}\alpha_i \geq \tau_a, \qquad S_i \geq \tau_S, \tag{19}$$

where $\tau_\ell, \tau_a, \tau_S \geq 0$ are selection thresholds. For an admissible candidate the residual scalar target is

$$\rho_i = \max\{\rho_i^{\text{sp}}, \tau_{\text{floor}} \rho_{\text{sc}}\}, \qquad 0 \leq \tau_{\text{floor}} \leq 1. \tag{20}$$

Thus the form of the selector is algorithmic, whereas the score weights, thresholds, spectral safety factor and sparsification parameters are numerical settings. Their values in the reported experiments are collected in Section 6.2.

If the same safety factor τ_ρ is used before and after temporary insertion, then for $\lambda_0 < \lambda_i^+ < 0$ the spectral part satisfies $\rho_i^{\text{sp}}/\rho_{\text{sc}} = 1 - \ell_i$. Consequently, before the upper clip in (18) is active, the coefficient of ℓ_i is $1 + c_\chi \chi_i - c_\theta$. In particular, choosing $0 \leq c_\theta \leq 1$ preserves monotonicity of the unsaturated score with respect to curvature lift, independently of the numerical values used in a particular campaign.

Composite-score low-rank update.

1. At a prescribed curvature update, estimate the leftmost spectrum of $\mathcal{Q}_k^{\mathrm{sel}}$ and retain at most $p_{\max}$ negative Ritz candidates.
2. Lift each candidate to ambient coordinates, remove components fixed by the current face, keep the smallest support carrying a squared-norm fraction τ_{supp} subject to the cap $s_{\max}$, and renormalize it.
3. Compute $\varrho_i = a_i^\top (Q + H_k) a_i$ and set β_i by (12); discard nonnegative or redundant candidates.
4. Temporarily insert $\beta_i a_i a_i^\top$, re-estimate λ_i^+, and evaluate the descriptors (13)–(17). Reject candidates that fail (19).
5. Rank admissible candidates by (18); insert the best one and greedily re-score the remaining pool. Use (20) for the residual scalar target. If none is admissible, keep H_k unchanged and rely on the scalar safeguard ρ_k.
6. During initialization, treat a reduced scalar target as a trial value and accept it only after the model and KKT checks succeed. Outside initialization, required increases of ρ_k may be applied immediately; proposed decreases remain trial values until the same checks succeed.
7. Solve the resulting convex model inexactly and apply the acceptance tests. Rejected trials leave the accepted reference model unchanged.
8. After accepted steps, monitor retained terms and, when the down-weighting criterion of Section 3.5 is met, test a reduced weight before committing it.

3.4 Initializing the low-rank correction

Initialization repeatedly discovers negative curvature, inserts an admissible rank-one correction and recomputes the leftmost-curvature estimate. With negative-curvature tolerance $\varepsilon_\lambda > 0$, discovery seeks $\lambda_{\min}(\mathcal{Q}_k^{\mathrm{sel}}) \geq -\varepsilon_\lambda$ and otherwise stops when the initialization rank budget $r_{\mathrm{init}}^{\max}$ or the search budget N_{search} is reached; failure of the spectral estimator also terminates the attempt. These are algorithmic controls. The numerical realization, including its retained-atom cap and configuration-specific coverage and initialization controls, is stated in Sections 6.2 and 6.3.

When $\lambda_{\min}(Q) < 0$, define the initialization coverage

$$\omega_k := 1 - \frac{\lambda_{\min}(Q + H_k)}{\lambda_{\min}(Q)}. \tag{21}$$

This quantity measures how much of the initial leftmost negative curvature has been removed.

Proposition 2 (Monotonicity of coverage during initialization)**.** *During initialization, as long as H_k is enlarged only by positive-semidefinite rank-one terms and no term is deleted or down-weighted, H_k is nondecreasing in the Loewner order. Hence in exact arithmetic both $\lambda_{\min}(Q + H_k)$ and ω_k are nondecreasing over those initialization rounds.*

Proof. Every accepted addition has the form $\beta a a^\top \succeq 0$. Weyl monotonicity therefore gives $\lambda_{\min}(Q + H_{k+1}) \geq \lambda_{\min}(Q + H_k)$. Since the denominator in (21) is fixed and negative, the same statement is equivalent to $\omega_{k+1} \geq \omega_k$. □

A decrease in the estimated coverage from one discovery round to the next cannot therefore be caused by the PSD update itself; it indicates finite accuracy in the spectral estimate rather than newly created negative curvature.

A useful diagnostic is the spectral progress per retained correction,

$$\eta_{\mathrm{cov}}(j) := \frac{\omega_j}{j}, \qquad j > 0. \tag{22}$$

It measures the spectral improvement obtained per retained correction. It is not a convergence certificate and is not used to rank final QP solutions.

3.5 When the regularization is updated

Low-rank updates are performed at prescribed update points. Let $T_{\mathrm{curv}} \in \mathbb{N}$ denote the nominal curvature-update period and $T_\rho \in \mathbb{N}$ the period at which a trial reduction of ρ_k may be considered. A rejected Newton direction carrying sufficiently negative Rayleigh curvature may also request a search. The values of these scheduling parameters are implementation settings reported in Section 6.2.

For an accepted displacement $d_k = x_{k+1} - x_k$, define

$$\Delta f_k = \max\{f(x_k) - f(x_{k+1}), \varepsilon_f(1 + |f(x_{k+1})|)\}, \qquad p_{i,k} = \beta_i (a_i^\top d_k)^2, \tag{23}$$

with $\varepsilon_f > 0$. A retained term is flagged for possible down-weighting when

$$p_{i,k} > \tau_{\mathrm{dw}} \Delta f_k, \tag{24}$$

where $\tau_{\mathrm{dw}} > 0$. After N_{dw} repeated activations, the term is eligible for a trial

$$\beta_i \leftarrow \gamma_{\mathrm{dw}} \beta_i, \qquad 0 < \gamma_{\mathrm{dw}} < 1. \tag{25}$$

The reduced weight is committed only if the model checks succeed; otherwise the previous weight is restored. Thus the down-weighting mechanism is defined independently of the numerical threshold, persistence count and reduction factor used in a particular run.

Changes in the low-rank part H_k are made at a reference point of the majorization model. If $x_{\mathrm{ref}} = x_k$, adding or reweighting a rank-one term does not change the model value at the current iterate because the quadratic correction vanishes there. This prevents artificial jumps in the local model while allowing the metric itself to both increase and decrease. The convergence result below does not require monotone metric growth; it uses uniformly strongly convex accepted models, bounded metrics and the relative accepted-step condition stated in Section 4.

For the convergence-certified controller, a trial metric may be accepted after any mathematically certified lower bound on $Q + M_k$ proves the required margin. A symmetric inertia computation for $Q + M_k - \mu I$ or a verified spectral enclosure are two possibilities; this model-Hessian test is distinct from the primal–dual KKT inertia check in Section 4. A concrete lower bound that always supplies a safe scalar fallback is the Gershgorin bound

$$\underline{\lambda}_Q^{\mathrm{G}} := \min_i \left(q_{ii} - \sum_{j \neq i} |q_{ij}| \right) \leq \lambda_{\min}(Q). \tag{26}$$

It is obtained in one pass over the sparse matrix, hence in $O(\mathrm{nnz}(Q))$ work and $O(1)$ additional storage per processed row. The bound is usually most informative for sparse or diagonally dominant matrices and can be conservative when large off-diagonal row sums dominate. It is used only to certify a safe scalar fallback, not as the scalar shift ordinarily selected by the adaptive controller. Any tighter verified lower bound may replace (26); an unconstrained or face-reduced Ritz value is not used as a certified lower bound.

Proposition 3 (Accepted-model safeguards)**.** *Set $\underline{\lambda}_Q = \underline{\lambda}_Q^{\mathrm{G}}$ in the convergence-certified controller, or use any tighter certified lower bound. Assume that retained directions are normalized, the retained count is bounded by $r_{\max}$, the strength multiplier in (12) is bounded, the scalar controller enforces $0 \leq \rho_k \leq \rho_{\max}$, and*

$$\rho_{\max} \geq \max\{0, \mu - \underline{\lambda}_Q\} \tag{27}$$

for a prescribed $\mu > 0$. Then the candidate rule produces uniformly bounded rank-one weights and hence $0 \preceq M_k \preceq \overline{M} I$ for a finite $\overline{M}$. If every insertion, deletion, weight change or scalar decrease is committed only after the certified global test

$$Q + M_k \succeq \mu I, \tag{28}$$

then every accepted model is μ-strongly convex. A rejected change leaves the last accepted metric unchanged, so down-weighting cannot destroy (28).

Proof. For a normalized candidate a and $H_k \succeq 0$, $\varrho_i = a^\top (Q + H_k) a \geq \lambda_{\min}(Q) \geq \underline{\lambda}_Q$. The weight rule (12) and the bounded strength multiplier thus give a common upper bound $\beta_{\max}$ for every retained coefficient. Consequently, $\|H_k\| \leq \sum_{j=1}^{r_k} \beta_{k,j} \leq r_{\max} \beta_{\max}$ and $\|M_k\| \leq r_{\max} \beta_{\max} + \rho_{\max}$. The rollback rule preserves the last accepted metric whenever a trial change fails the global test. Finally, because $H_k \succeq 0$, the safe fallback $\rho_k = \rho_{\max}$ gives

$$Q + H_k + \rho_{\max} I \succeq (\underline{\lambda}_Q + \rho_{\max}) I \succeq \mu I,$$

which proves that the accepted-model margin can always be restored. The Gershgorin construction (26) makes this fallback fully specified at large scale. The face-reduced Ritz estimates used for candidate selection remain diagnostics and are not substituted for the certified global bound. □

3.6 Toy geometry

The two-dimensional example in Figure 1 isolates the geometry of scalar versus directional convexification. It uses

$$Q = R \operatorname{diag}(\tfrac{1}{2}, -3) R^\top = \begin{pmatrix} 0.3945 & 0.5985 \\ 0.5985 & -2.8945 \end{pmatrix}, \qquad q = \begin{pmatrix} 1.206 \\ 1.5 \end{pmatrix}, \qquad K = [-3, 3]^2, \tag{29}$$

where R is a rotation by 10°. The eigenvalues are 0.5 and -3; the corresponding eigenvectors v_+ and v_- are intentionally rotated relative to the coordinate axes. Starting from $x_0 = (-1, 0)$, compare

$$M_{\text{uniform}} = 3.1 I, \qquad M_{\text{LR}} = 3.1\, v_- v_-^\top. \tag{30}$$

Both choices leave a 0.1 safety margin at the left spectral edge, but only the uniform shift alters the already-positive curvature along v_+. Gray curves show the level sets of the original objective f in every panel. The low-rank model changes the negative-curvature direction while preserving the curvature of f along v_+; the uniform model shifts both directions.

Figure 2 makes the same comparison on one-dimensional cuts. Along v_+ the low-rank model follows the original curvature, whereas the uniform shift is unnecessarily stiffer. Along v_- the two convexifying models agree on the correction needed to remove negative curvature.

The iteration comparison in Figure 3 reaches the same constrained stationary point $x^\star \simeq (1.4947, -3)$. The low-rank model takes 16 outer iterations versus 164 for the uniform shift. The measured linear contraction factors are approximately 0.192 and 0.887, respectively. This example is illustrative rather than a complexity theorem, but it shows directly why preserving benign curvature can matter.

4 Primal–dual interior-point formulation

4.1 Bordered KKT system and inertia condition

Let the retained directions be the columns of U_k and $\Gamma_k = \operatorname{diag}(\beta_{k,1}, \ldots, \beta_{k,r_k})$, so that $H_k = U_k \Gamma_k U_k^\top$. The low-rank term is represented through a border rather than by explicitly

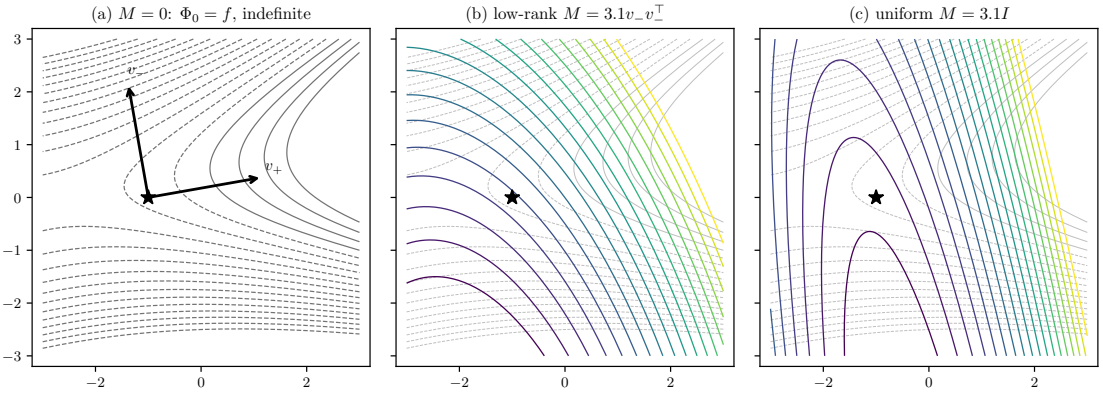

Figure 1: Original objective and two convexifying models on the rotated toy problem. Gray curves are level sets of f in every panel; arrows in panel (a) identify the positive- and negative-curvature eigenvectors v_+ and v_-.

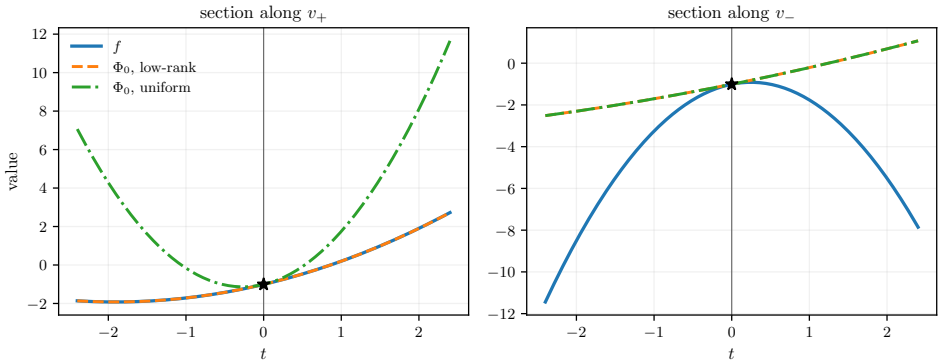

Figure 2: Directional cuts of the toy objective and its convexifying models. The low-rank model preserves the original positive curvature along v_+ while matching the required stabilization along v_-.

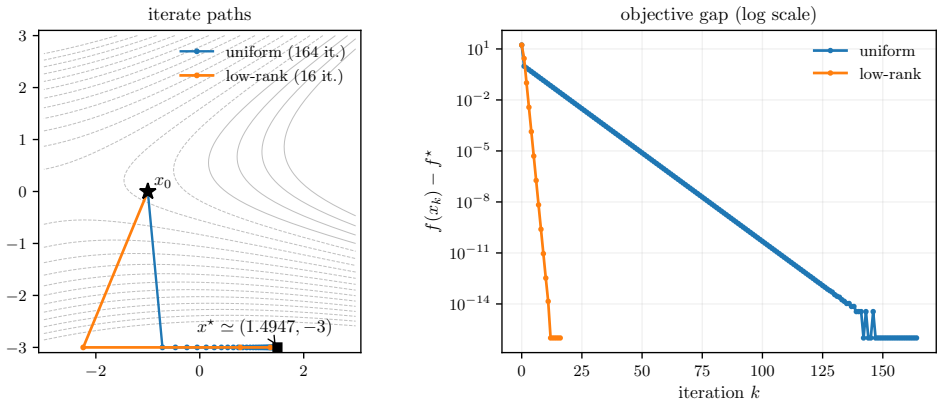

Figure 3: Toy trajectories and asymptotic contraction. Both models approach $x^\star \simeq (1.4947, -3)$ on the lower face; the low-rank model requires 16 outer iterations versus 164 for the uniform shift.

adding these outer products to the sparse base block. After the usual barrier eliminations,

write the unbordered symmetric system schematically as

$$\widehat{J}_k = \begin{pmatrix} D_k & -A^\top & -C^\top \\ -A & -\delta_d I & 0 \\ -C & 0 & -(\Theta_s^{-1} + \delta_d I) \end{pmatrix}, \qquad D_k = Q + (\rho_k + \rho_{x,k})I + \Theta_x, \tag{31}$$

where $m_E := \dim(b)$ and $m_I := \dim(b_I)$ are the numbers of equality and inequality constraints, $\Theta_x, \Theta_s \succ 0$ are diagonal barrier terms, $\delta_d > 0$ is a dual regularization, and $\rho_{x,k} \geq 0$ is a numerical primal stabilization. In particular, $\rho_{x,k}$ is not part of the DC metric $M_k = H_k + \rho_k I$. With $\widetilde{U}_k = (U_k^\top, 0, 0)^\top$, the rank-$r_k$ correction is represented by the bordered matrix

$$\mathcal{J}_k = \begin{pmatrix} \widehat{J}_k & \widetilde{U}_k \\ \widetilde{U}_k^\top & -\Gamma_k^{-1} \end{pmatrix}. \tag{32}$$

Thus adding a retained direction appends one border row and column but does not densify the base Hessian block.

The bordered system also gives an exact inertia test. Block congruence gives

$$\mathrm{In}(\mathcal{J}_k) = \mathrm{In}\Big(\widehat{J}_k + \widetilde{U}_k \Gamma_k \widetilde{U}_k^\top\Big) + (0, r_k, 0), \tag{33}$$

so the target sign count is

$$\mathrm{In}(\mathcal{J}_k) = (n, m_E + m_I + r_k, 0). \tag{34}$$

A second Schur-complement congruence gives the equivalent positive-primal factorization condition

$$D_k + H_k + A^\top(\delta_d I)^{-1}A + C^\top(\Theta_s^{-1} + \delta_d I)^{-1}C \succ 0. \tag{35}$$

If already $D_k \succ 0$, the matrix is symmetric quasidefinite after grouping the primal block against the remaining variables, giving the classical robust factorization setting of [35]. This condition is stronger than necessary for (34). Conversely, the constraint and barrier terms in (35) can stabilize the KKT factorization even when $Q + M_k$ is not positive semidefinite. The KKT inertia test is therefore a numerical factorization check; it does not establish convexity of the DCA subproblem. In particular, every proposed scalar decrease or rank-one down-weighting is first subjected to the certified global test (28). Only a globally certified model is passed to the inertia check, which may then trigger numerical primal or dual regularization. Passing the inertia check alone never commits a metric change. When the primal sign count is deficient, $\rho_{x,k}$ may be increased; dual rank or conditioning defects are handled through δ_d. This separation is the reason the low-rank correction can change while the sparse factorization structure remains intact; it is closely related to stabilized KKT systems in [3] but the border here represents the optimization model itself.

4.2 Mehrotra predictor–corrector step

Each convex DCA subproblem is advanced with a primal–dual predictor–corrector step of Mehrotra type [25]. The affine step provides a centering estimate, a corrector right-hand side accounts for the complementarity product, and fraction-to-the-boundary step lengths preserve strict interiority. This sparse primal–dual formulation follows standard large-scale interior-point practice; see also [37]. Every regularization strategy reported in Section 7 is realized on this same interior-point architecture.

4.3 Merit function, acceptance and numerical exits

Following [28, Section 3.2], the barrier subproblem uses the exact L^1-penalty merit function

$$\Psi^{(k)}_{\sigma,\mu}(x,s) = f^{(k)}_{\mu}(x,s) + \sigma\|Ax-b\|_1 + \sigma\|Cx-s\|_1.$$

Here $\widetilde{z}$ denotes the current vector of multiplier estimates entering the penalty update. The primal Newton direction is a descent direction whenever

$$\sigma \geq \|\widetilde{z}\|_\infty + \bar{\sigma}, \tag{36}$$

with $\bar{\sigma} > 0$; the Mayne–Polak penalty update used in the original IPDCA is retained [24]. Fraction-to-the-boundary and the primal merit-function line search follow the same logic as IPDCA, while the dual step is safeguarded when the accepted primal displacement is much shorter than the full Newton step.

For the convergence analysis, write $s_k := x_{k+1} - x_k$. At a feasible candidate $x \in \mathcal{K}$, define the model-stationarity residual

$$\mathcal{R}_k(x) := \operatorname{dist}(0, \partial\Phi_k(x)) = \min_{v \in N_{\mathcal{K}}(x)} \|Qx + q + M_k(x - x_k) + v\|. \tag{37}$$

For a polyhedron, this is a convex least-squares problem over the multipliers of the active linear constraints and bounds. The accepted inner point is required to satisfy

$$r_{k+1} \in \partial\Phi_k(x_{k+1}), \qquad \|r_{k+1}\| = \mathcal{R}_k(x_{k+1}) \leq \kappa_{\text{in}}\|s_k\|, \qquad 0 \leq \kappa_{\text{in}} < \mu/2, \tag{38}$$

where μ is the accepted-model margin in (28). This is a relative accuracy condition for the convex DCA model, not an exact subproblem solve. Exact convex-subproblem solves satisfy (38) with $\kappa_{\text{in}} = 0$. Together with the certified global curvature test, the condition implies both the descent and the relative-stationarity estimates used in Section 5; they are no longer postulated separately. The merit line search and feasibility safeguards remain the globalization mechanism used to obtain an acceptable trial.

Operational termination is kept separate from the relative accepted-step condition above. A first-order exit requires the primal feasibility, stationarity and complementarity residuals to satisfy tolerances $\varepsilon_{\text{feas}}$, $\varepsilon_{\text{stat}}$ and $\varepsilon_{\text{comp}}$. LazyIPDCA also provides a DCA-level stabilization criterion defined on accepted outer iterates rather than on individual interior-point Newton steps. Let $x^{[j]}$ denote the jth accepted DCA anchor and $F_j = f(x^{[j]})$. Given $\varepsilon^{\text{abs}}_{\text{DCA}}, \varepsilon^{\text{rel}}_{\text{DCA}} \geq 0$ and an integer $N_{\text{DCA}} \geq 1$, objective stabilization is declared when

$$|F_j - F_{j-1}| \leq \varepsilon^{\text{abs}}_{\text{DCA}} + \varepsilon^{\text{rel}}_{\text{DCA}} \max\{1, |F_j|\} \tag{39}$$

holds for N_{DCA} consecutive accepted anchors. Rejected trial steps do not advance this counter. Criterion (39) is independent of the inner interior-point residuals and is not used as a substitute for the relative accepted-model condition (38). The numerical tolerances used in the experiments are stated in Section 6.2.

4.4 LazyIPDCA algorithm

The complete method is summarized below. The low-rank candidate construction and selection are those of Section 3.3, while weight reductions are governed by Section 3.5. The scalar correction ρ_k remains a safeguard throughout; low-rank corrections supplement it rather than replace it.

Algorithm 1 LazyIPDCA accepted-anchor logic

Require: a feasible anchor x_0, an accepted metric $M_0 = H_0 + \rho_0 I$, a certified $\underline{\lambda}_Q$, $\mu > 0$, $\rho_{\max} \geq \max\{0, \mu - \underline{\lambda}_Q\}$, $0 \leq \kappa_{\text{in}} < \mu/2$, stopping tolerances and computational limits

for $k = 0, 1, 2, \ldots$ **do**
  Generate event-driven trial changes to H_k and ρ_k from the last accepted metric
  Certify $Q + H_k + \rho_k I \succeq \mu I$; use $\rho_{\max}$ as the safe fallback if needed
  Form Φ_k and the bordered system (32); use inertia only to regularize the factorization
  Advance the convex model until a feasible candidate satisfies (38)
  Apply the merit and model-decrease acceptance tests
  if the candidate is accepted **then**
    Commit the new anchor and controller state
    For every proposed decrease of ρ_k or β_i, rerun the global test before the separate inertia check
  else
    Restore the last accepted anchor and metric, revise the model, and recompute
  end if
  if a first-order test, (39), or a computational limit is satisfied **then**
    stop
  end if
end for

Algorithm 1 separates the three checks that matter for the analysis: global model convexity, relative inner accuracy, and accepted-step globalization. The KKT inertia check concerns the linear system and is never used in place of the global model test.

5 Convergence and rates

The analysis is stated for the accepted outer sequence, with $x_0 \in \mathcal{K}$. Define F by (3), $s_k = x_{k+1} - x_k$, and

$$\mathcal{L}_0 := \{x \in \mathcal{K} : F(x) \leq F(x_0)\}. \tag{40}$$

The compactness assumption below is automatic when the variable bounds make $\mathcal{K}$ compact, and also holds whenever F is level-bounded on $\mathcal{K}$.

Assumption 1. *(C1) The lower level set $\mathcal{L}_0$ is nonempty and compact.*

(C2) There exist $\mu > 0$ and $\overline{M} < \infty$ such that every accepted metric satisfies

$$0 \preceq M_k \preceq \overline{M} I, \qquad Q + M_k \succeq \mu I.$$

(C3) The accepted model point satisfies the relative residual condition (38) for a fixed $0 \leq \kappa_{\text{in}} < \mu/2$.

Condition (C2) is enforced in the convergence version of the controller by the global accepted-model test of Proposition 3; since $\mathcal{K}$ is nonempty and closed, it also makes every model subproblem coercive and uniquely solvable. The KKT signed-spectrum test remains a factorization check and is not used as a substitute for model convexity. Condition (C3) is the computable inner stopping rule (37)– (38). The proof therefore rests on controller properties checked at each accepted outer step, rather than on independent summability and vanishing-residual assumptions. The finite-tolerance experiments below are not used to infer the asymptotic rate.

Theorem 1 (Descent and global stationarity complexity)**.** *Under Assumption 1, set*

$$a := \mu/2 - \kappa_{\text{in}} > 0, \qquad b := \overline{M} + \kappa_{\text{in}}, \qquad F_{\text{inf}} := \min_{x \in \mathcal{L}_0} F(x).$$

Then, for every k,

$$F(x_k) - F(x_{k+1}) \geq a\|s_k\|^2, \tag{41}$$

$$\operatorname{dist}(0, \partial F(x_{k+1})) \leq b\|s_k\|. \tag{42}$$

Consequently, $\{x_k\} \subset \mathcal{L}_0$, $\sum_k \|s_k\|^2 < \infty$, every accumulation point is a KKT point of (1), and for every $N \geq 1$,

$$\min_{0\leq k<N} \operatorname{dist}(0, \partial F(x_{k+1})) \leq \frac{b}{\sqrt{aN}} \sqrt{F(x_0) - F_{\inf}}. \tag{43}$$

Thus the best stationarity residual among the first N accepted outer steps is $O(N^{-1/2})$.

Proof. Since $Q + M_k \succeq \mu I$, the model Φ_k is μ-strongly convex on $\mathcal{K}$. For r_{k+1} in (38), the strong subgradient inequality at x_{k+1} with comparison point x_k gives

$$\Phi_k(x_k) - \Phi_k(x_{k+1}) \geq -\langle r_{k+1}, s_k\rangle + \frac{\mu}{2}\|s_k\|^2 \geq a\|s_k\|^2.$$

Because $\Phi_k(x_k) = F(x_k)$ and $\Phi_k(x_{k+1}) = F(x_{k+1}) + \frac{1}{2}s_k^\top M_k s_k \geq F(x_{k+1})$, (41) follows. It keeps the sequence in $\mathcal{L}_0$ and, after summation, yields $a\sum_{k=0}^{N-1}\|s_k\|^2 \leq F(x_0) - F(x_N) \leq F(x_0) - F_{\inf}$.

Next write $r_{k+1} = w_{k+1} + M_k s_k$ with $w_{k+1} \in \partial F(x_{k+1})$. Then $\|w_{k+1}\| \leq \|r_{k+1}\| + \|M_k\|\|s_k\| \leq b\|s_k\|$, proving (42). Compactness gives accumulation points, while square summability gives $s_k \to 0$; closedness of the graph of $\partial F = Qx + q + N_{\mathcal{K}}(x)$ then gives KKT stationarity at every accumulation point. Finally, sum $\operatorname{dist}(0, \partial F(x_{k+1}))^2 \leq b^2\|s_k\|^2$ and take the minimum over $k = 0, \ldots, N-1$ to obtain (43). □

The $N^{-1/2}$ order in (43) matches the worst-case stationarity order established for smooth DCA under its corresponding termination measure. It matches the worst-case gradient-norm order proved for smooth DCA in [1], where that order is shown to be exact for the corresponding termination measure. The $O(N^{-1})$ result in the same reference concerns a different nonsmooth DCA criterion. Here the controlled quantity is the limiting-subdifferential residual of a constrained, inexact, variable-metric QP model, so neither the constants nor the stopping measures are interchangeable. The local linear result below is instead KL-based, in the same broad tradition as whole-sequence and local-rate analyses for pDCA$_e$ [21].

Because $\mathcal{K}$ is polyhedral, $0 \in \partial F(x^\star)$ is equivalent to the standard KKT system for (1); no additional constraint qualification is needed for this normal-cone representation.

The preceding result does not yet use the special algebraic form of the QP. That form strengthens the cluster-point conclusion to convergence of the complete sequence and supplies a local rate.

Lemma 1 (KL exponent of the QP objective)**.** *The extended objective F in (3) is a proper closed piecewise linear–quadratic function, continuous on $\operatorname{dom}\partial F = \mathcal{K}$, and has the Kurdyka–Łojasiewicz property with exponent $1/2$ at every critical point.*

Proof. Corollary 5.2 of [20] covers a finite minimum of functions $\frac{1}{2}x^\top M_i x + u_i^\top x + \beta_i + P_i(x)$ with M_i symmetric, not necessarily positive semidefinite, and P_i proper, closed and polyhedral, provided the resulting function is continuous on its limiting-subdifferential domain. Apply it with one piece: $M_1 = Q \in \mathbb{S}^n$, $u_1 = q$, $\beta_1 = 0$, and $P_1 = \delta_{\mathcal{K}}$. Because $\mathcal{K}$ is a nonempty closed polyhedron, $\delta_{\mathcal{K}}$ is proper, closed and polyhedral, and $N_{\mathcal{K}}(x)$ contains zero for every $x \in \mathcal{K}$. The smooth sum rule therefore gives

$$\partial F(x) = Qx + q + N_{\mathcal{K}}(x), \qquad \operatorname{dom}\partial F = \mathcal{K}.$$

The quadratic part is continuous and $\delta_{\mathcal{K}}$ vanishes on $\mathcal{K}$, so F is continuous relative to $\mathcal{K}$. All hypotheses of the cited corollary are therefore satisfied, including when Q is indefinite. □

Theorem 2 (Whole-sequence and local linear convergence)**.** *Under Assumption 1, the accepted sequence has finite length,*

$$\sum_{k=0}^{\infty} \|x_{k+1} - x_k\| < \infty, \tag{44}$$

and converges to a single KKT point $x^\star$. Moreover, either the sequence terminates at a stationary point in finitely many accepted steps or there are constants $q \in (0,1)$ and $C > 0$ such that, for all sufficiently large k,

$$0 \le F(x_{k+1}) - F(x^\star) \le q\,[F(x_k) - F(x^\star)], \qquad \|x_k - x^\star\| \le C q^{k/2}. \tag{45}$$

Proof. Equations (41) and (42) are precisely the sufficient-decrease and relative-error hypotheses of the KL descent theorem [8]. The compact lower level set provides a cluster point, and $F(x_k) \downarrow F(x^\star)$ along every convergent subsequence because F is continuous relative to $\mathcal{K}$. The uniformized KL inequality on the compact cluster set, together with Lemma 1, therefore gives finite length and convergence of the whole sequence to one critical point.

For the rate, let $\Delta_k := F(x_k) - F(x^\star)$. The exponent-1/2 KL inequality gives a constant $c > 0$ such that, for all large k with $\Delta_{k+1} > 0$, $\operatorname{dist}(0, \partial F(x_{k+1})) \ge c\sqrt{\Delta_{k+1}}$. Using (42) and (41),

$$\Delta_k - \Delta_{k+1} \ge a\|s_k\|^2 \ge \frac{ac^2}{b^2}\Delta_{k+1}.$$

Hence (45) holds with $q = (1 + ac^2/b^2)^{-1}$. The finite-length KL estimate bounds the remaining path length by a constant multiple of $\sqrt{\Delta_k}$, which yields the stated R-linear rate for the iterates. □

5.1 Relation to finite-tolerance implementations

Theorems 1–2 concern a convergence-certified variant: accepted anchors are feasible, every metric change passes the global test (28), and the inner residual is purely relative. A fixed absolute residual floor produces a quantifiable stationarity floor rather than exact asymptotic stationarity.

Proposition 4 (Finite-accuracy stationarity bound)**.** *Assume (C2), let $F(x_k) \ge F_{\mathrm{low}} > -\infty$, and suppose*

$$\|r_{k+1}\| \le \kappa\|s_k\| + \nu, \qquad 0 \le \kappa < \mu/2,$$

for constants κ and $\nu \ge 0$. Set $a = \mu/2 - \kappa$ and $b = \overline{M} + \kappa$. Then, for every $N \ge 1$,

$$\min_{0 \le k < N} \operatorname{dist}(0, \partial F(x_{k+1})) \le \left[\frac{4b^2(F(x_0) - F_{\mathrm{low}})}{aN} + 2\left(1 + \frac{b^2}{a^2}\right)\nu^2\right]^{1/2}. \tag{46}$$

Proof. The strong-subgradient argument used in Theorem 1 gives

$$F(x_k) - F(x_{k+1}) \ge a\|s_k\|^2 - \nu\|s_k\| \ge \frac{a}{2}\|s_k\|^2 - \frac{\nu^2}{2a}.$$

Summation yields $\sum_{k<N} \|s_k\|^2 \le 2(F(x_0) - F_{\mathrm{low}})/a + N\nu^2/a^2$. Moreover, $\operatorname{dist}(0, \partial F(x_{k+1})) \le b\|s_k\| + \nu$. Squaring, averaging and using $(u+v)^2 \le 2u^2 + 2v^2$ proves (46). □

Thus a constant floor preserves the $O(N^{-1/2})$ transient only down to an $O(\nu)$ neighborhood. Exact KKT convergence requires the purely relative rule or a suitable vanishing error sequence; a fixed feasibility tolerance likewise gives an approximate rather than exact KKT statement.

The reported production campaign uses absolute feasibility, stationarity and complementarity tolerances together with the accepted-objective stabilization rule (39). It does not expose numerical values of μ and κ_{in} and is not claimed to instantiate the convergence-certified variant. The terminal residuals in the supplement range from values well below 10^{-6} to substantially larger values on some stabilization or recovery exits. This is qualitatively consistent with a finite-accuracy floor, but the model residual r_{k+1} required to estimate ν was not archived, so no numerical value of ν is inferred. Likewise, the accepted-anchor objective histories were not stored at the resolution needed to estimate the local contraction factor in (45); the local linear rate is therefore a theoretical statement rather than an empirical fit to these runs.

Second-order interpretation. A KKT point of a nonconvex QP need not be a local minimizer. For a KKT point $x^\star$, let

$$\mathcal{C}(x^\star) := \{d \in T_{\mathcal{K}}(x^\star) : \langle Qx^\star + q, d\rangle = 0\}$$

be the critical cone. The strict condition $d^\top Qd > 0$ for every nonzero $d \in \mathcal{C}(x^\star)$ is sufficient for a strict local minimum. LazyIPDCA is designed to reach stationarity; it does not make this second-order property automatic.

6 Experimental protocol

6.1 Implementation and comparison protocol

All CUTEst experiments were run on the same AMD Ryzen Threadripper 7960X workstation with 256 GiB of memory and GCC 13.3, using one computational thread per run. The reported LazyIPDCA realization uses sparse Lanczos curvature estimation, the bordered KKT representation of Section 4, the sequential sparse symmetric-indefinite LDL backend, and the Mehrotra predictor–corrector scheme. The external references are Ipopt 3.14.19, called through cyipopt 1.7.0 with MUMPS 5.8.2, and QPALM 1.3.0 in nonconvex mode.

All direct comparisons start from the original CUTEst point, retain fixed variables and use a wall-time limit of 36,000 s. LazyIPDCA and Ipopt use a 3000-iteration safeguard, while QPALM uses 50,000 iterations. These counters are not commensurate: a QPALM proximal augmented-Lagrangian update is neither an accepted DCA anchor nor an interior-point Newton/KKT factorization. The common wall-time guard is therefore the cross-solver resource limit; the solver-specific iteration caps prevent unbounded internal progress without being used as an efficiency metric. The same LazyIPDCA–Ipopt protocol is used on the 17-instance large-scale panel. For STNQP1, an external wall-time guard terminates the Ipopt run before the first iteration is completed, so no final point is returned. Reported dimensions are the actual CUTEst decoded dimensions.

A returned point enters the objective comparison only if its largest constraint violation in original coordinates is at most 10^{-6}. If only one method returns a feasible point within the budget, that method is credited in the comparison; when both do, their objectives are compared using the tie statistic

$$g_f(f; f^\star) := \frac{|f - f^\star|}{\max\{1, |f^\star|\}}, \qquad g_f \le 10^{-4} \quad \Longrightarrow \quad \text{tie}. \tag{47}$$

The lower bound of one in the denominator is important for near-zero objectives such as the BLOWEY family. The CUTEst Ipopt reference requests tolerance 10^{-8}, uses gradient-based scaling and the monotone barrier update, and disables acceptable-level early termination. QPALM requests an absolute tolerance of 10^{-6}. Independently of solver-reported status,

the matched 37-instance post-processing evaluates the same four original-coordinate KKT1 components and applies the same threshold 10^{-6} to all three returns: primal feasibility, relative stationarity, complementarity and multiplier sign. The multiplier source is solver specific. LazyIPDCA and Ipopt use their exported terminal multiplier vectors, whereas QPALM multipliers on the near-active set are reconstructed by constrained sparse least squares. Thus KKT1 is harmonized at the level of residual definitions, scaling and tolerance, but it does not claim that all three multiplier vectors were produced by an identical recovery routine. The reported median wall time per solver iteration is used only to quantify the different scales of the three iteration counters; it is not treated as a common work unit.

The complementarity comparisons use the certificate in (50), rather than a generic QP solver status. For QPALM, absolute tolerances are tried successively in $\{10^{-6}, 10^{-8}, 10^{-10}, 10^{-12}, 10^{-13}\}$ until that certificate is satisfied or the sequence is exhausted. The tables report the wall time of the selected run, not the cumulative time of preceding tolerance trials; the certificate-matched Ipopt results use the same reporting convention.

6.2 Numerical settings of LazyIPDCA

Sections 3–4.4 define the method in terms of symbolic controls. Table 2 gives the single set of values used in the reported CUTEst experiments. These are campaign-level settings and are not assumptions of the convergence theorem; no value is tuned per problem.

Table 2: Numerical parameter values used in the reported CUTEst runs.

Symbol	Role	Reported value
τ_ρ	relative scalar curvature safety factor	1.10
τ_{floor}	residual scalar floor fraction	0.10
$p_{\max}$	negative Ritz candidates per update	8
τ_{supp}	retained squared-norm fraction	0.999999
$s_{\max}$	support cap per retained direction	1000
δ_β	additive rank-one weight offset	10^{-6}
s_β	rank-one strength multiplier	1
ε_λ	negative-curvature tolerance	10^{-8}
$L_{\max}$	Lanczos iteration cap per spectral call	60
s_{res}	Lanczos residual-safety multiplier	1
$r_{\max}$	global retained-atom cap	500
T_{curv}	nominal curvature-update period	50 iterations
T_ρ	scalar-decrease trial period	2 iterations
ε_f	objective-decrease floor in (23)	10^{-12}
τ_{dw}	down-weighting activity threshold	0.50
N_{dw}	repeated activations before a weight trial	5
γ_{dw}	multiplicative weight reduction	0.50
$\varepsilon_{\text{feas}}, \varepsilon_{\text{stat}}, \varepsilon_{\text{comp}}$	first-order termination tolerances	10^{-6}
$\varepsilon_{\text{DCA}}^{\text{abs}}$	absolute objective-stabilization tolerance	10^{-8}
$\varepsilon_{\text{DCA}}^{\text{rel}}$	relative objective-stabilization tolerance	10^{-10}
N_{DCA}	consecutive accepted anchors in (39)	5

The composite score (18) uses

$$(c_\chi, c_\theta, c_s, c_e, c_\beta) = (0.50, 0.50, 0.10, 0.10, 0.05) \tag{48}$$

and admissibility thresholds

$$(\tau_\ell, \tau_a, \tau_S) = (10^{-4}, 10^{-2}, 10^{-4}). \tag{49}$$

The scalar controller further uses a trial-decay factor 0.5, a relative floor margin 10^{-6} and a rejected-step growth factor 2. These quantities belong to the numerical realization rather than to the abstract convergence model. The convergence-only parameters μ and κ_{in} are not listed in Table 2, because the production campaign did not enforce the pure relative rule (38); no numerical values are inferred retrospectively.

6.3 Low-rank configurations

Four low-rank configurations are considered. A and B use *coverage-target selection* with target $\omega_{\text{targ}} = 0.25$ in (21); C and D use *composite-score selection* (18). B and D additionally impose the initialization cap $r_{\text{init}}^{\max} = 20$, whereas A and C use the global rank/search budgets without this additional cap. Otherwise all four share the implementation, controller settings, globalization strategy and comparison rules stated above. The first numerical experiment compares these four choices and selects the default used in all subsequent LazyIPDCA results.

6.4 Sensitivity and timing design

The sensitivity study evaluates the scalar-decrease period T_ρ on ten instances. Five of them form the common subset used for every other one-factor perturbation: NCVXQP1 at $n = 1000$, BLOWEYA at decoded dimension $n = 4002$, STNQP1 at $n = 8193$, and BLOCKQP3 and NCVXQP3 at dimensions 10010 and 10000. On this subset, the reported default LazyIPDCA configuration is compared with 12 one-factor perturbations, giving 65 runs, and five additional runs impose the active cap $r_{\max} = 4$. Because T_ρ is the only moderate control that changes trajectories on the common subset, its values 1, 2 and 5 are also evaluated on QPNBOEI1, QPNSTAIR, NCVXBQP3 at $n = 10000$, NCVXQP6 at $n = 1000$ and QPNBAND at $n = 50000$. The T_ρ study therefore contains 30 runs on ten instances, 15 of which are already included in the 65-run design. The parameter evidence contains 85 distinct runs. Timing reproducibility is assessed separately through ten independent repetitions on each instance of the common five-instance subset, for 50 further runs.

The Lanczos iteration cap takes values 30, 60 and 120; the scalar-decrease trial period T_ρ takes values 1, 2 and 5; the nonbinding global atom caps take values 100, 500 and 1000; and the Lanczos residual-safety multiplier takes values 0.5, 1 and 2. The two reward coefficients of the composite score are scaled jointly by 0.75 and 1.25, and the three penalty coefficients are perturbed in the same way. The active-cap runs compare $r_{\max} = 4$ with the reference value 500. Conclusions use objective, feasibility, the common KKT1 components, reported stationarity, retained rank and Lanczos-call count. One-factor wall times are not used to rank settings; the repeat campaign is summarized by the median, standard deviation and coefficient of variation of wall time.

6.5 Benchmark scope

The main comparison among the four LazyIPDCA configurations contains 37 distinct CUTEst problem–dimension instances after duplicate manifest aliases are merged using the decoded problem name and actual dimension. They come from 24 distinct CUTEst problems; the NCVXQP/NCVXBQP families predominate. A separate large-scale panel contains 17 QPs with dimensions from 100,000 to 131,073. Eight are NCVXBQP/NCVXQP cases at $n =$ 100,000, and QPNBAND at $n = 100{,}000$ is an official predefined CUTEst size. To broaden structural coverage, STNQP1–2, BLOWEYA–C and BLOCKQP1–3 are evaluated at larger parametric sizes obtained from local copies of their official SIF definitions. In each of these eight extensions, exactly the active numerical size parameter is changed and the mathematical problem definition is otherwise left unchanged; their source and modified SHA-256 hashes are

archived with the numerical evidence. These eight sizes are therefore reported as parametric extensions, not as additional predefined CUTEst catalog sizes. The 37-instance set is compared with both Ipopt and QPALM. The large-scale direct comparison also contains complete 17-instance campaigns for both references. The eight NCVXBQP/NCVXQP Ipopt cases use the $n = 100{,}000$ reference runs, while the remaining nine use the same Ipopt configuration on the broadened panel. QPALM is evaluated on the same 17 instances with the first-order post-check stated above, and on all 72 complementarity reformulations under the original-problem certificate. CUTEst decoding follows [14].

7 Numerical results

The main text reports aggregate results and the comparisons needed to interpret them. Complete instance-level objectives, feasibility measures, termination labels, iteration counts and timing data are provided in the supplementary material.

7.1 Selection rule and initialization cap

Table 3: Low-rank configuration comparison on the 37 CUTEst instances. Feasibility uses 10^{-6} and objective ties use (47).

Configuration	feasible	best within tol.	total t (s)
Coverage-target selection (A)	37/37	32/37	772.089
Coverage-target selection, rank cap 20 (B)	37/37	33/37	412.918
Composite-score selection (C)	37/37	35/37	381.251
Composite-score selection, rank cap 20 (D)	37/37	35/37	370.380

Table 3 shows that composite-score selection is feasible throughout and attains the best feasible objective within tolerance on 35 instances. Configurations C and D have identical objectives, feasibility measures, iteration counts and termination labels on every instance; the 20-direction initialization cap in D is inactive. C is therefore retained as the default LazyIPDCA configuration. The coverage-target cap is active: B changes two returned objectives relative to A and reduces the aggregate time.

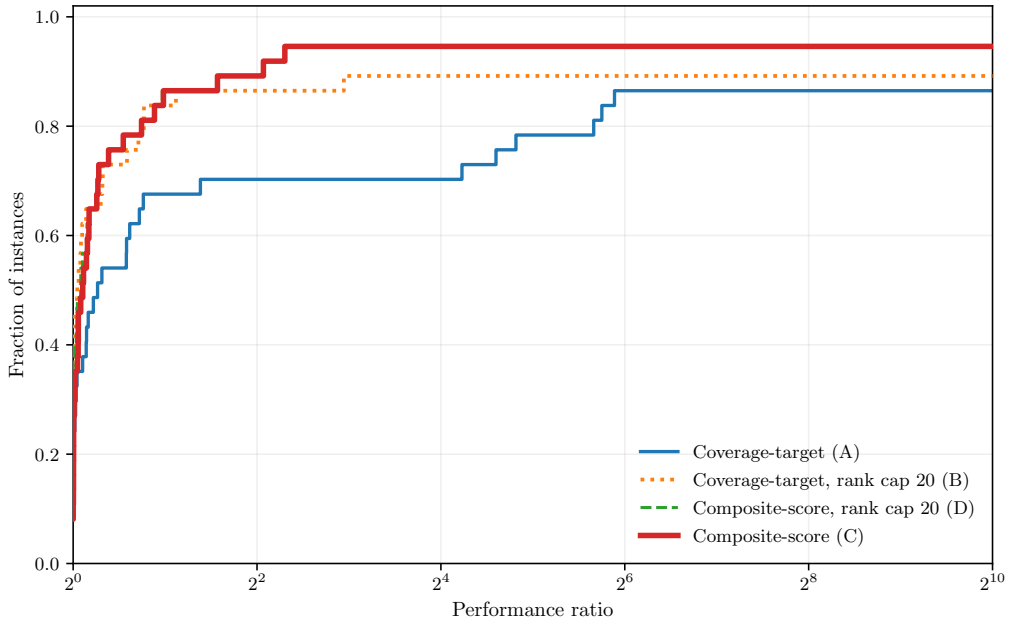

Figure 4: Quality-filtered wall-time performance profile for configurations A–D. On each instance, an eligible run is feasible and lies within the objective tolerance in (47); its performance ratio is its wall time divided by the smallest wall time among eligible configurations on that instance. Ineligible runs have infinite ratio. The plateaus are 32/37, 33/37, 35/37 and 35/37.

7.2 Parameter sensitivity and timing reproducibility

Table 4: Local one-factor sensitivity of the default configuration. 'Final diagnostics' denotes objective, feasibility, reported stationarity, retained rank and Lanczos-call count. The T_ρ axis is assessed on ten instances; every other one-factor perturbation and the binding-rank-cap experiment use the common five-instance subset. All 85 parameter-study returns satisfy the 10^{-6} feasibility threshold.

Control	Values around the default	Observed result
Lanczos iteration cap $L_{\max}$	30/60/120	final diagnostics unchanged on 5/5 instances
Scalar-decrease period T_ρ	1/2/5	on 10 instances: 10/10 feasible and 6/10 KKT1 for each value; several trajectories change
Nonbinding atom caps $r_{\max}$	100/500/1000	final diagnostics unchanged; no cap binds
Active atom cap $r_{\max}$	4 versus 500	binds on 3/5 instances and changes those trajectories
Lanczos residual safety s_{res}	0.5/1/2	final diagnostics unchanged on 5/5 instances
Reward coefficients	0.75/1/1.25 times default	final diagnostics unchanged on 5/5 instances
Penalty coefficients	0.75/1/1.25 times default	final diagnostics unchanged on 5/5 instances

Table 4 shows that moderate changes in the Lanczos cap and residual-safety factor, and joint ±25% changes in the reward or penalty coefficients of the composite score, leave the recorded outcome unchanged on all five instances. The caps 100, 500 and 1000 are also inactive because the reference ranks range only from zero to seven. These tests therefore rule out fine tuning within the sampled neighborhood without implying global parameter insensitivity.

The scalar-decrease trial period is the only moderate perturbation that changes trajectories. On the expanded ten-instance panel, $T_\rho = 1, 2, 5$ each returns 10/10 feasible points and 6/10 KKT1 passes. The aggregate feasibility and first-order pass counts are therefore unchanged, although four of the five added instances exhibit visible changes in objective, stationarity or spectral work; QPNBAND at $n = 50000$ is essentially invariant. On NCVXQP3 at $n = 10000$, changing T_ρ from the default 2 to 1 raises the reported stationarity measure from 1.346×10^{-1} to 9.204×10^{-1} and the Lanczos-call count from 1837 to 2512; $T_\rho = 5$ gives 1.193×10^{-1} and 2089 calls. On NCVXQP1 at $n = 1000$, $T_\rho = 5$ raises the stationarity measure from 5.082×10^{-13} to 8.660×10^{-7}. These results support $T_\rho = 2$ as an intermediate default rather than an instance-wise optimum. The median wall times of the three settings are reported in the supplement only; they are not ranked because the extension combines retained runs with new single runs.

The active cap $r_{\max} = 4$ binds on NCVXQP1, BLOCKQP3 and NCVXQP3. On NCVXQP1 it preserves the objective and stationarity to the reported precision while reducing the retained rank from five to four and the Lanczos-call count from 271 to 37. The capped BLOCKQP3 and NCVXQP3 trajectories also remain inside the 10^{-4} objective-tie band of their references, but change their stationarity and spectral work. In particular, the NCVXQP3 stationarity measure decreases from 1.346×10^{-1} to 1.349×10^{-6} and its Lanczos-call count from 1837 to 37. Both of these latter capped runs end with a nonzero process return code, so the experiment demonstrates the effect of a binding cap but does not motivate replacing the default.

Table 5: Timing reproducibility over ten independent runs per instance. Objective, feasibility, reported stationarity, retained rank and Lanczos-call count are identical across the ten repetitions of every instance.

Instance	median t (s)	CV (%)	min. t (s)	max. t (s)
NCVXQP1, $n = 1000$	0.321	1.24	0.317	0.329
BLOWEYA, $n = 4002$	1.260	0.41	1.257	1.273
STNQP1, $n = 8193$	2.756	0.29	2.742	2.768
BLOCKQP3, $n = 10010$	13.000	0.34	12.961	13.109
NCVXQP3, $n = 10000$	171.909	0.39	171.012	172.906

Table 5 gives coefficients of variation from 0.29% to 1.24%. The identical algorithmic diagnostics across repetitions show that the observed dispersion is timing noise rather than trajectory variation. Full one-factor, active-cap and repetition records are provided in the supplementary material and accompanying numerical archive.

7.3 Scalar regularization against LazyIPDCA

Table 6: Aggregate regularization comparison on the 37 CUTEst instances.

Regularization	feasible	best within tol.	total time (s)
Fixed scalar	37/37	18/37	1408.796
Adaptive scalar	37/37	32/37	1713.414
LazyIPDCA	37/37	33/37	381.251

Table 6 compares the two scalar metrics with the default low-rank metric. All returns are feasible. Adaptive scalar is 21.6% more expensive in aggregate than fixed scalar and produces the strictly lower objective on 18 instances, with 19 ties and no fixed-scalar win. LazyIPDCA has the smallest aggregate time and the largest three-way best-or-tied count. These results show that scalar adaptation improves objective quality at a substantial spectral and computational cost, whereas the low-rank metric achieves the stronger aggregate time–quality balance.

7.4 Direct comparison with Ipopt and QPALM

Table 7: Matched aggregate comparison on the 37 CUTEst instances. KKT1 uses the same residual definitions and tolerance 10^{-6} for all three solvers; multiplier sources are stated in Section 6.1. The total-time ratios use LazyIPDCA as reference. Median time per reported iteration is descriptive because the iteration counters represent different algorithmic work.

Solver	feasible	KKT1	total t (s)	t/t_L	median t/iter. (s)
LazyIPDCA	37/37	27/37	381.251	1.00	8.719×10^{-3}
Ipopt	37/37	28/37	3428.100	8.99	2.301×10^{-2}
QPALM	37/37	35/37	6412.156	16.82	9.270×10^{-4}

Table 8: Pairwise objective partitions on the same 37 feasible returns. Ties use (47); each row partitions all 37 instances.

Pairwise objective comparison	LazyIPDCA lower	reference lower	tied
Ipopt	21	3	13
QPALM	20	3	14

Table 7 reports feasible points from all three solvers on all 37 instances. The harmonized KKT1 test is satisfied on 27 LazyIPDCA returns, 28 Ipopt returns and 35 QPALM returns. The corresponding total wall times are 381.251 s, 3428.100 s and 6412.156 s. These measures describe different attributes of the returned points: the highest KKT1 count does not coincide with the smallest aggregate time or with the largest number of lower objectives.

All ten LazyIPDCA KKT1 misses remain primal feasible. Eight exceed the stationarity threshold and seven exceed the complementarity threshold, with overlap; no separate multiplier-sign failure is recorded. Six terminate on the accepted-objective DCA stabilization rule before all KKT1 components reach 10^{-6}. Two follow an unsuccessful bounded terminal recovery, for which the pre-recovery point is restored after residual degradation. The remaining two are accompanied by ill-conditioned multiplier recovery, a diagnostic of possible active-constraint degeneracy rather than a verified LICQ failure, and exhaust the

scalar safeguard. Thus DCA stabilization is the dominant terminal mode, but not the sole explanation for the 27/37 count. The ten cases and their residual components are listed in the supplement.

The median seconds per reported iteration in Table 7 quantify the scale of the three iteration counters. The LazyIPDCA median is 9.41 times the QPALM median, whereas the Ipopt median is 2.64 times the LazyIPDCA median. Since the iterations perform different algorithmic work, these ratios are descriptive rather than work-equivalence factors. They also show why the safeguards of 3000 iterations for LazyIPDCA and Ipopt and 50,000 for QPALM are not equal computational budgets; the common wall-time limit is the cross-solver resource control.

Table 8 shows that LazyIPDCA has the strictly lower feasible objective on 21 instances against Ipopt and on 20 against QPALM. Each reference has three strict wins. The remaining 13 and 14 cases are tied, respectively. Against Ipopt, tightening the objective threshold from 10^{-4} to 10^{-6} changes the partition from 21/3/13 to 21/7/9.

The two QPALM returns that fail KKT1 both reach the 50,000-iteration safeguard and are limited by stationarity rather than feasibility or complementarity. On the bounds-only NCVXBQP3 instance at $n = 10{,}000$, primal feasibility is 5.33×10^{-15} and relative complementarity is 7.16×10^{-12}, while relative stationarity is 1.81×10^{-6}. On BLOCKQP3 the corresponding values are 1.90×10^{-9}, 1.61×10^{-12} and 1.46×10^{-4}. The data identify incomplete first-order convergence at the iteration limit, without indicating a feasibility or multiplier-sign failure.

The three strict QPALM objective wins occur on QPNBOEI1, BLOWEYB and BLOCKQP1. On BLOCKQP1, values near -4994 for QPALM and 5 for LazyIPDCA show that the methods enter different basins. On the 14 LazyIPDCA–QPALM ties, LazyIPDCA records the smaller wall time on 11 instances; QPALM is faster on QPNSTAIR, QPNBAND at $n = 10{,}000$ and BLOCKQP2. Complete instance-level objectives and residuals are reported in the supplementary material.

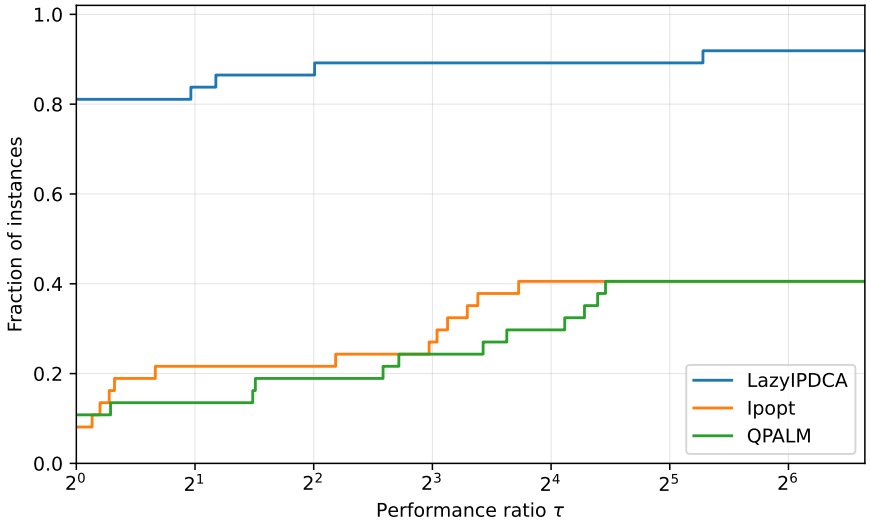

Figure 5: Quality-filtered three-solver wall-time performance profile on the 37 matched instances, truncated at $\tau = 100$. Eligibility is joint rather than pairwise: a return must be feasible and lie within the band in (47) of the best feasible objective among all three solvers. In general, ineligibility can result from infeasibility or objective quality; here all 37 returns of every solver are feasible, so the 3, 21 and 22 ineligible LazyIPDCA, Ipopt and QPALM returns arise only from objective values outside the band. The ratio is wall time divided by the smallest eligible wall time on the instance. At $\tau = 1$, the counts are 30/37, 3/37 and 4/37; the eligibility plateaus are 34/37, 16/37 and 15/37.

Figure 5 complements the aggregate totals with their instance-level distribution. At $\tau = 2$ the counts are 31/8/5 for LazyIPDCA/Ipopt/QPALM; at $\tau = 5$ they are 33/9/7, and at $\tau = 100$ they are 34/15/15. LazyIPDCA is the fastest eligible solver on 30 instances. The remaining plateau differences reflect objective quality alone on this matched set, not feasibility failures.

7.5 Large-scale panel

Table 9: Summary of the 17-instance large-scale comparison. LazyIPDCA returns a feasible point on all 17 instances. The objective partition credits LazyIPDCA when the reference does not return a feasible point; otherwise it uses the tie rule in (47). QPALM's KKT1 column is the independent first-order post-check.

Reference	ref. feasible	ref. KKT1	Lazy/ref./tie	t_L (s)	t_R (s)	retained r
Ipopt	16/17	–	9/4/4	114873.208	232940.574	0–7
QPALM	11/17	8/17	9/1/7	114873.208	201987.952	0–7

Table 9 extends the same feasible-return comparison used on the 37-instance set to dimensions from 100,000 to 131,073. Against Ipopt, LazyIPDCA is credited on nine instances, Ipopt on four, and four are tied. Against QPALM the corresponding partition is 9/1/7: QPALM returns a feasible point on 11 instances and satisfies the independent KKT1 post-check on eight. Restricting the objective comparison to the 11 mutually feasible LazyIPDCA–QPALM pairs gives three lower objectives for LazyIPDCA, one for QPALM, and seven ties. The six remaining QPALM runs do not satisfy the common primal-feasibility threshold and are therefore credited to LazyIPDCA. Five are the linearly constrained NCVXQP1, 3, 5, 6 and

9 instances at $n = 100{,}000$; each reaches the wall-time limit, with final violations from 1.75×10^{-3} to 4.77 and relative stationarity residuals from 0.555 to 1.005. In contrast, all three bounds-only NCVXBQP instances at the same dimension are feasible. The sixth case, BLOWEYA, reaches the iteration cap with a smaller violation of 3.23×10^{-6}. The observed pattern is therefore concentrated in the large, linearly constrained NCVXQP rows rather than in all problems above a given dimension; it is an empirical diagnosis, not a proof of a solver limitation.

The family split remains informative. On the eight NCVXBQP/NCVXQP instances at $n = 100{,}000$, both reference comparisons give the same objective partition 7/0/1 in favor of LazyIPDCA, although QPALM returns a feasible point on only three of those eight. On the nine added QPNBAND, STNQP, BLOWEY and BLOCKQP structures, QPALM is feasible on eight and the partition is 2/1/6, whereas Ipopt gives 2/4/3. This reinforces the pattern already visible on the 37-instance set: the relative solver behaviour is strongly structure dependent. The retained correction is endogenous, with rank two on the first subset and values from zero to seven on the additional structures.

7.6 Large-scale complementarity reformulations

The generated LCP families use the formulation

$$x \geq 0, \qquad w = \mathsf{L}x - g \geq 0, \qquad x^\top w = 0,$$

embedded as a linearly constrained QP whose objective is $x^\top w$. A return is certified when

$$\max\{0, -\min_i x_i, -\min_i w_i\} \leq 10^{-6}, \qquad \|x \odot w\|_\infty \leq 10^{-6}, \qquad |x^\top w| \leq 10^{-6}. \tag{50}$$

When required, a structure-specific semismooth Newton postsolve is accepted only after these original-problem conditions are re-evaluated.

Table 10: Certificate-matched complementarity comparisons. LazyIPDCA satisfies (50) on all 72 instances. The Ipopt study contains 67 instances because its matched sweep was not continued for five LCP10 sizes above 5×10^5; the QPALM study covers all 72. Time ratios use the selected certified run and exclude an uncertified reference return.

Reference	evaluated	certified	joint cert.	Lazy faster	reference faster	median t_R/t_L
Ipopt	67	66	66	56	10	1.31
QPALM	72	71	71	8	63	0.08

Table 10 shows two distinct reference patterns. QPALM satisfies the certificate at its baseline tolerance on 31 instances and on 40 more after tolerance tightening. Its only uncertified return is LCP8 at $n = 5 \times 10^6$: the aggregate gap is 7.160×10^{-6} although the componentwise complementarity and primal violations are 2.294×10^{-10} and 4.590×10^{-10}. On the 71 jointly certified instances, QPALM has the smaller selected-run wall time on 63 and the median ratio is $t_Q/t_L = 0.08$. The distribution is not uniform: on LCP7 at $n = 4 \times 10^6$, QPALM requires 8549.030 s, compared with 57.63 s for LazyIPDCA.

The Ipopt and QPALM studies certify 66 of the same 67-instance subset, but on different exceptions: Ipopt misses LCP10 at $n = 500{,}000$, whereas QPALM misses LCP8 at $n = 5 \times 10^6$. The family definitions, tolerance selections, residuals and instance-level times are reported in the supplementary material.

LCP10 is structurally different from the nearly uniformly scaled LCP7–LCP8 families: its matrix is $\mathsf{L} = \mathrm{diag}(1/n, 2/n, \ldots, 1)$, so its diagonal condition number is n and the complementary solution has components $x_i = n/i$. This increasing scale separation is

consistent with the hundreds of KKT factorizations seen on most large LCP10 runs, versus seven for LCP7 and LCP8. The $n = 2 \times 10^6$ run follows a distinct path with 78 factorizations, compared with 763–770 at one, three, four and five million variables. This reduction in recorded linear-algebra work accounts for most of the lower wall time, but the stored aggregate records do not identify the controller event that caused the trajectory change. No asymptotic inference is therefore made from this isolated run.

8 Discussion

The experiments support the central computational premise: a small, structure-dependent correction can be preferable to shifting every direction uniformly. The retained rank is zero on several BLOWEY and BLOCKQP instances, two on the original $n = 100{,}000$ NCVX subset, and at most seven on the broader large-scale panel. The scalar ablation shows that this advantage belongs to the complete regularization strategy rather than to scalar adaptation alone.

The theoretical analysis uses enforceable accepted-step properties. A global curvature test and the rollback rule protect strong convexity after insertions or weight reductions; finite rank and scalar caps bound the metric; and a relative model residual supplies descent, stationarity complexity, whole-sequence convergence and a local linear rate. The compact-level-set assumption is automatic for bounded feasible sets but remains a condition for general unbounded formulations. The limit is a KKT point and need not satisfy the second-order condition for a local minimum.

The empirical scope has two limitations. First, the 37-instance set is dominated by the NCVXQP/NCVXBQP families, although the 17-instance panel broadens the structures represented. Second, T_ρ, the only moderate control that changes trajectories, is assessed on ten instances, whereas every remaining one-factor perturbation and the binding-rank-cap experiment use the common five-instance subset. The design isolates local one-factor effects; parameter interactions remain outside its scope, and timing reproducibility is examined separately on that five-instance subset. The two external references are complementary: Ipopt represents a general nonlinear interior-point method, whereas QPALM is designed specifically for nonconvex quadratic programs. On the matched 37-instance set, the common KKT1 counts are 27, 28 and 35 for LazyIPDCA, Ipopt and QPALM, while the three-solver profile identifies LazyIPDCA as the fastest eligible solver on 30 instances. These diagnostics show why feasibility, first-order quality, objective value and time must be reported separately.

The method is most natural when indefiniteness is concentrated in a modest set of directions and sparse factorization dominates the cost. It is less attractive when negative curvature is genuinely high-rank, when repeated spectral searches dominate, or when a specialized structure provides a more direct solver. The external comparisons illustrate this dependence. On the large QP panel, QPALM's nonfeasible returns are concentrated in five heavily constrained NCVXQP instances that exhaust the wall-time budget, while the bounds-only NCVXBQP instances at the same dimension remain feasible. On the LCP set, QPALM is faster on most jointly certified instances but misses one certificate at five million variables. In the 67-instance certificate-matched Ipopt study, LazyIPDCA certifies all 67 instances and Ipopt 66; among the 66 jointly certified cases, LazyIPDCA is faster on 56 and the median ratio t_I/t_L is 1.31. The LCP postsolve remains structure-specific and is kept separate from the generic QP method.

9 Conclusion

LazyIPDCA replaces uniform DC convexification by an event-driven metric that combines sparse rank-one curvature corrections with a residual scalar safeguard. The bordered KKT realization retains the sparse base factorization, while the convergence-certified variant admits whole-sequence convergence, a global $O(N^{-1/2})$ best-stationarity bound and a local linear rate for the quadratic–polyhedral objective. The finite-accuracy analysis states explicitly how fixed inner tolerances replace exact stationarity by a residual floor.

The experiments support the intended mechanism rather than a uniform ranking of solvers. On the matched CUTEst set, LazyIPDCA combines complete feasibility, competitive first-order quality and the strongest objective–time profile; Ipopt and QPALM remain preferable on particular structures. The large-scale and complementarity studies show the same dependence on problem structure, with retained rank adapting from zero to seven and original-problem certification separating solver status from useful accuracy. Moderate parameter perturbations preserve the aggregate feasibility and KKT1 counts, whereas a binding rank cap and the scalar-decrease period visibly alter some trajectories. These results support adaptive low-rank convexification when negative curvature is concentrated in a modest subspace and sparse factorization dominates the cost, while leaving high-rank indefiniteness and strongly specialized problem structure as natural limits of the approach.

Acknowledgements

The author gratefully acknowledges the foundational joint work with T. Pham Dinh and H. A. Le Thi that led to IPDCA and motivated the present extension.

Statements and Declarations

Funding. The author received no specific funding for this work.

Competing interests. The author declares no competing interests.

Author contributions. François Bertrand Akoa: conceptualization, methodology, software, formal analysis, investigation, validation, data curation, visualization, writing–original draft, and writing–review and editing.

Data and code availability. The LazyIPDCA source code and the reproducibility materials supporting this study are available at https://github.com/francoisakoa/lazyipdca-qp. The repository contains the experiment scripts, CUTEst manifests, generated LCP definitions and processed tables. CUTEst problem definitions remain subject to their original distribution terms. The instance-level numerical results are supplied as supplementary information.

Use of AI-assisted tools. OpenAI's ChatGPT assisted with language editing, LaTeX formatting and consistency checks. The author verified the mathematical arguments, numerical data, references and final text and accepts full responsibility for the content.

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] H. Abbaszadehpeivasti, E. de Klerk and A. Taylor, *On the convergence rate of the boosted difference-of-convex algorithm (DCA)*, Optimization Letters, 2026, https://doi.org/10.1007/s11590-026-02305-w.

[3] A. Altman and J. Gondzio, *Regularized symmetric indefinite systems in interior point methods for linear and quadratic optimization*, Optimization Methods and Software, 11(1):275–302, 1999.

[4] F.J. Aragón-Artacho, R. Campoy and P.T. Vuong, *The Boosted DC Algorithm for Linearly Constrained DC Programming*, Set-Valued and Variational Analysis, 30:1265–1289, 2022, https://doi.org/10.1007/s11228-022-00656-x.

[5] F.J. Aragón Artacho, R.M.T. Fleming and P.T. Vuong, *Accelerating the DC algorithm for smooth functions*, Mathematical Programming, 169:95–118, 2018, https://doi.org/10.1007/s10107-017-1180-1.

[6] F.J. Aragón-Artacho, P. Pérez-Aros and D. Torregrosa-Belén, *The Boosted Double-proximal Subgradient Algorithm for nonconvex optimization*, Mathematical Programming, 214:491–537, 2025, https://doi.org/10.1007/s10107-024-02190-0.

[7] F.J. Aragón-Artacho and P.T. Vuong, *The boosted difference of convex functions algorithm for nonsmooth functions*, SIAM Journal on Optimization, 30(1):980–1006, 2020.

[8] H. Attouch, J. Bolte and B.F. Svaiter, *Convergence of descent methods for semi-algebraic and tame problems: proximal algorithms, forward–backward splitting, and regularized Gauss–Seidel methods*, Mathematical Programming, 137:91–129, 2013, https://doi.org/10.1007/s10107-011-0484-9.

[9] A.R. Conn, N.I.M. Gould, D. Orban and Ph.L. Toint, A primal-dual trust-region algorithm for non-convex nonlinear programming, Math. Program. 87, 215–249 (2000). https://doi.org/10.1007/s101070050112.

[10] N.I.M. Gould, D. Orban and Ph.L. Toint, GALAHAD, a library of thread-safe Fortran 90 packages for large-scale nonlinear optimization, ACM Trans. Math. Softw. 29(4), 353–372 (2003). https://doi.org/10.1145/962437.962438.

[11] F.E. Curtis and D.P. Robinson, *Exploiting negative curvature in deterministic and stochastic optimization*, Mathematical Programming, 176:69–94, 2019.

[12] E. D. Dolan and J. J. Moré, 'Benchmarking optimization software with performance profiles,' *Mathematical Programming*, vol. 91, pp. 201–213, 2002.

[13] O.P. Ferreira, E.M. Santos and J.C.O. Souza, *A boosted DC algorithm for non-differentiable DC components with non-monotone line search*, Computational Optimization and Applications, 88:783–818, 2024, https://doi.org/10.1007/s10589-024-00578-4.

[14] N.I.M. Gould, D. Orban and Ph.L. Toint, *CUTEst: a constrained and unconstrained testing environment with safe threads for mathematical optimization*, Computational Optimization and Applications, 60(3):545–557, 2015.

[15] N.I.M. Gould, S. Lucidi, M. Roma and Ph.L. Toint, *Solving the trust-region subproblem using the Lanczos method*, SIAM Journal on Optimization, 9(2):504–525, 1999, https://doi.org/10.1137/S1052623497322735.

[16] B. Hermans, A. Themelis and P. Patrinos, *QPALM: a proximal augmented Lagrangian method for nonconvex quadratic programs*, Mathematical Programming Computation, 14(3):497–541, 2022, https://doi.org/10.1007/s12532-022-00218-0.

[17] H.A. Le Thi, H.P.H. Luu, H.M. Le and T. Pham Dinh, *Stochastic DCA with variance reduction and applications in machine learning*, Journal of Machine Learning Research, 23(206):1–44, 2022.

[18] H.A. Le Thi and T. Pham Dinh, *DC programming and DCA: thirty years of developments*, Mathematical Programming, 169(1):5–68, 2018.

[19] H.A. Le Thi and T. Pham Dinh, *Open issues and recent advances in DC programming and DCA*, Journal of Global Optimization, 88(3):533–590, 2024, https://doi.org/10.1007/s10898-023-01272-1.

[20] G. Li and T.K. Pong, *Calculus of the exponent of the Kurdyka–Łojasiewicz inequality and its applications to linear convergence of first-order methods*, Foundations of Computational Mathematics, 18:1199–1232, 2018, https://doi.org/10.1007/s10208-017-9366-8.

[21] T. Liu, T.K. Pong and A. Takeda, *A refined convergence analysis of* pDCA$_e$ *with applications to simultaneous sparse recovery and outlier detection*, Computational Optimization and Applications, 73(1):69–100, 2019, https://doi.org/10.1007/s10589-019-00067-z.

[22] T. Liu and A. Takeda, *An inexact successive quadratic approximation method for a class of difference-of-convex optimization problems*, Computational Optimization and Applications, 82:141–173, 2022, https://doi.org/10.1007/s10589-022-00357-z.

[23] Z. Lu, Z. Zhou and Z. Sun, *Enhanced proximal DC algorithms with extrapolation for a class of structured nonsmooth DC minimization*, Mathematical Programming, 176:369–401, 2019, https://doi.org/10.1007/s10107-018-1318-9.

[24] D.Q. Mayne and E. Polak, *A superlinearly convergent algorithm for constrained optimization problems*, Mathematical Programming Study, 16:45–61, 1982.

[25] S. Mehrotra, *On the implementation of a primal-dual interior point method*, SIAM Journal on Optimization, 2(4):575–601, 1992.

[26] S. Nakayama, Y. Narushima and H. Yabe, *Inexact proximal DC Newton-type method for nonconvex composite functions*, Computational Optimization and Applications, 87:611–640, 2024, https://doi.org/10.1007/s10589-023-00525-9.

[27] T. Pham Dinh and H.A. Le Thi, *Convex analysis approach to DC programming: theory, algorithms and applications*, Acta Mathematica Vietnamica, 22(1):289–355, 1997.

[28] T. Pham Dinh, H.A. Le Thi and F. Akoa, *Combining DCA (DC Algorithms) and interior point techniques for large-scale nonconvex quadratic programming*, Optimization Methods and Software, 23(4):609–629, 2008.

[29] D.N. Phan and H.A. Le Thi, *Difference-of-convex algorithm with extrapolation for nonconvex, nonsmooth optimization problems*, Mathematics of Operations Research, 49(3):1973–1985, 2024, https://doi.org/10.1287/moor.2020.0393.

[30] S. Pougkakiotis and J. Gondzio, *Dynamic non-diagonal regularization in interior point methods for linear and convex quadratic programming*, Journal of Optimization Theory and Applications, 181:905–945, 2019.

[31] S. Pougkakiotis and J. Gondzio, *An interior point-proximal method of multipliers for convex quadratic programming*, Computational Optimization and Applications, 78:307–351, 2021.

[32] L. Qi and J. Sun, *A nonsmooth version of Newton's method*, Mathematical Programming, 58:353–367, 1993, https://doi.org/10.1007/BF01581275.

[33] N. Rontsis, P.J. Goulart and Y. Nakatsukasa, *An active-set algorithm for norm constrained quadratic problems*, Mathematical Programming, 193:447–483, 2022, https://doi.org/10.1007/s10107-021-01617-2.

[34] S. Takahashi, M. Fukuda and M. Tanaka, *New Bregman proximal type algorithms for solving DC optimization problems*, Computational Optimization and Applications,

83:893–931, 2022, https://doi.org/10.1007/s10589-022-00411-w.

[35] R.J. Vanderbei, *Symmetric quasidefinite matrices*, SIAM Journal on Optimization, 5(1):100–113, 1995.

[36] R. Verschueren, M. Zanon, R. Quirynen and M. Diehl, *A sparsity preserving convexification procedure for indefinite quadratic programs arising in direct optimal control*, SIAM Journal on Optimization, 27(3):2085–2109, 2017.

[37] A. Wächter and L.T. Biegler, *On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming*, Mathematical Programming, 106(1):25–57, 2006.

[38] B. Wen, X. Chen and T.K. Pong, *A proximal difference-of-convex algorithm with extrapolation*, Computational Optimization and Applications, 69(2):297–324, 2018, https://doi.org/10.1007/s10589-017-9954-1.

[39] L. Yang, J. Hu and K.-C. Toh, *An inexact Bregman proximal difference-of-convex algorithm with two types of relative stopping criteria*, Journal of Scientific Computing, 103:91, 2025, https://doi.org/10.1007/s10915-025-02904-2.

[40] Y. You and Y.-S. Niu, *A variable metric and Nesterov extrapolated proximal DCA with backtracking for a composite DC program*, Journal of Industrial and Management Optimization, 19(10):7716–7734, 2023, https://doi.org/10.3934/jimo.2023016.