AFTERIMAGE SLOW REGIONS IN FIRST-ORDER METHODS FOR LINEAR CONIC PROGRAMMING

Anonymous authors
Paper under double-blind review

ABSTRACT

First-order methods for linear conic programming often stall on long plateaus. Existing analyses characterize *when* during a run or *on which problem instances* slow convergence occurs; we instead ask *where* in the state space slow convergence is present. We introduce the *slow region family* for parameterized averaged fixed-point iterations. On a slow region, one step moves the state by only a small fraction of its distance to the fixed-point set, so an orbit starting there keeps almost its initial distance for arbitrarily many iterations. We then develop the *afterimage* principle to construct them. A *center* family and a nearby *petal* family share the same limit parameter, so their residual fields become close, while their fixed-point sets or forward drifts stay far apart. The petal geometry then certifies a slow region for the center. We verify the standing assumptions for ADMM, sGS-ADMM, and PDHG, and give an LP, SOCP, and SDP gallery showing how varied slow regions can be.

1 INTRODUCTION

Linear conic programming (conic-LP), encompassing linear programming (LP), second-order cone programming (SOCP), and semidefinite programming (SDP), provides a versatile modeling framework for problems in machine learning (Weinberger & Saul, 2004; Lanckriet et al., 2004), statistics (d'Aspremont et al., 2004), and robust and combinatorial optimization (Lobo et al., 1998). We consider the standard primal–dual pair:

$$\min_{x\in\mathcal{X}} \langle c, x\rangle_{\mathcal{X}}, \quad \text{subject to } \mathcal{A}x = b,\ x \in \mathcal{K}, \tag{P}$$

$$\max_{y\in\mathcal{Y}, s\in\mathcal{X}} \langle b, y\rangle_{\mathcal{Y}}, \quad \text{subject to } \mathcal{A}^* y + s = c,\ s \in \mathcal{K}^*, \tag{D}$$

where $b \in \mathcal{Y}$, $c \in \mathcal{X}$, and $\mathcal{A}$ is a linear map from $\mathcal{X}$ to $\mathcal{Y}$ with adjoint $\mathcal{A}^*$. Here $\mathcal{X}$ and $\mathcal{Y}$ are finite-dimensional real Euclidean spaces (we identify $\mathcal{Y}$ with $\mathbb{R}^m$), and $\mathcal{K} \subseteq \mathcal{X}$ is a closed convex cone with dual cone $\mathcal{K}^*$.

A popular route to solving conic-LP at scale is first-order methods (FOM) (Ryu & Yin, 2022), including the Alternating Direction Method of Multipliers (ADMM) (Odonoghue et al., 2016; Wen et al., 2010), the Primal–Dual Hybrid Gradient method (PDHG) (Chambolle & Pock, 2011; Applegate et al., 2021), and symmetric Gauss–Seidel based ADMM (sGS-ADMM) (Chen et al., 2017). All of them run a dynamical system on a state space $\mathcal{H}$, driven by a fixed point map $T_\theta : \mathcal{H} \to \mathcal{H}$:

$$z^{(k+1)} = T_\theta z^{(k)}, \tag{1}$$

where T_θ encodes the problem parameter $\theta := (\mathcal{A}, b, c)$, and the primal–dual variables (x, y, s) are recovered from the state $z \in \mathcal{H}$ by a method-specific map. Under mild conditions $z^{(k)}$ converges, as $k \to \infty$, to a state representing a Karush–Kuhn–Tucker (KKT) point of the conic-LP.

From slow phases to slow regions. FOM convergence curves on conic-LP are rarely uniform: the KKT residual decreases quickly during an initial transient, and then nearly stalls for a long period. *Temporal* explanations divide a run into stages. PDHG on LP first identifies the optimal basis and then converges linearly; the length of the first stage reflects how nearly degenerate the instance is, and the rate of the second is determined by a local sharpness constant (Lu & Yang, 2025a). More generally, error-bound and metric-subregularity arguments give the same local picture for splitting methods near a sufficiently regular KKT point (Han et al., 2018; Kang et al., 2025). *Instance-geometric*

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

explanations relate the whole complexity to the problem data, connecting restarted PDHG to the level-set geometry of the primal–dual pair (Xiong & Freund, 2024) and to condition measures that quantify how close an instance is to infeasibility or to multiple optima (Xiong & Freund, 2026; Xiong, 2026). Both answer *when* in a run, or *on which instances*, a method is slow. We instead ask *where*:

In which regions of the state space $\mathcal{H}$ does the iteration (1) *become very slow?*

To answer this, we need a notion of slowness that belongs to a region rather than to a stage of a run. We take the cue from the local limit dynamics of ADMM on SDP (Kang & Yang, 2026), where slowness is read from the limit residual field on the state space near a singular KKT point. There, slowness is a property of the region, not only of the instance: on one SDP, ADMM converges locally linearly near a strictly complementary KKT point (Kang et al., 2025), and nearly stalls near a singular one in the same optimal set. Such spatial diagnoses are common in numerical linear algebra: a few outlying small eigenvalues slow down conjugate gradient, and deflation removes that subspace to restore a fast rate (Saad et al., 2000). What is still missing for FOM in conic-LP is a general way to construct and certify such regions, across different cones and different methods. This paper provides one. We start with a toy example that already contains the mechanism.

Motivating example. Consider (P) with $\mathcal{X} = \mathbb{R}^2$, $\mathcal{K} = \mathbb{R}^2_+$, $\bar{\mathcal{A}} := [1,\ -1]$ and $c := (1, 0)$, and let only the right-hand side move:

$$\min_{x \in \mathbb{R}^2} \ x_1 \quad \text{subject to } x_1 - x_2 = \varepsilon,\ x \geq 0, \qquad \varepsilon \geq 0. \tag{LP$_\varepsilon$}$$

The primal optimum is the single point $(\varepsilon, 0)$ for every $\varepsilon \geq 0$. At $\varepsilon = 0$, however, the *dual* optimal set degenerates into a whole segment. The ADMM map T_θ inherits this degeneracy: at $\varepsilon = 0$ its fixed-point set is a segment $\mathcal{L}$, while for every $\varepsilon > 0$ it is a single point, which converges to one endpoint of $\mathcal{L}$ as $\varepsilon \downarrow 0$ (Figure 1, left). In contrast, the residual field $\delta_\theta(z) := T_\theta(z) - z$ shifts by the same vector at every state, of norm only $\varepsilon/\sqrt{2}$. As a result, if the iterates start from the other endpoint of $\mathcal{L}$, they move along $\mathcal{L}$ for about $2\sigma/\varepsilon$ steps, where σ is ADMM's penalty parameter, with $\|\delta_\theta(z^{(k)})\|_2$ staying at $\varepsilon/\sqrt{2}$, before they detect the collapse. In words, a vanishing perturbation leaves a long-lived *afterimage* of the unperturbed fixed-point set, and this afterimage is exactly where the iteration becomes slow.

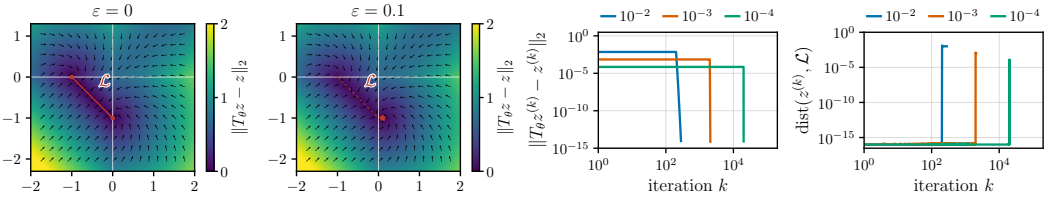

Figure 1: ADMM with $\sigma = \tau = 1$. The motivating example under ADMM. *Left two:* the residual field $T_\theta z - z$ on $\mathcal{H} = \mathbb{R}^2$, colored by its norm, at $\varepsilon = 0$ and 0.1 — nearly indistinguishable, yet $\mathrm{Fix}(T_\theta)$ collapses from the segment $\mathcal{L}$ (solid; dashed on the right) to the single star. *Right two:* from the far endpoint of $\mathcal{L}$, the same seed for every ε; each decade in ε lengthens the plateau by a decade.

Contributions. We view the dynamical system (1) as a *static* residual field δ_θ on $\mathcal{H}$, and certify slow regions by comparing two different responses to a perturbation of θ: the residual field itself changes continuously, while the asymptotic geometry it carries — its fixed points, or its forward drift — can change abruptly. Specifically, our contributions are threefold.

(i) We introduce the *slow region family* for parameterized averaged fixed-point iterations under mild regularity conditions. Slowness is asymptotic and belongs to the whole family: each region is measured against the fixed-point set of its own parameter, and the family is slow if the ratio of the one-step residual to this distance tends to zero. The regions may move with the parameter, or even diverge. We also provide calculus rules for combining and thickening such families.

(ii) We develop the *afterimage* principle to construct such families. We compare a *center* family, which carries the problems we actually solve, with a nearby *petal* family. Both are indexed by a vanishing perturbation and converge to a common *limit parameter*. Their residual fields therefore

become close, while their fixed-point sets or forward drifts stay far apart, so that the petal geometry certifies a slow region family for the center iteration. Neither the petal nor the limiting problem needs to possess a KKT point: both may be solvable, (weakly) infeasible, non-attaining, or carry a positive duality gap. The principle covers a broad range of slow-convergence phenomena, including face selection, Hölder-sensitive displacement, and escape of fixed points to infinity, each with a bound on the resulting slowness. Figure 2 summarizes the framework.

(iii) We verify the standing assumptions for ADMM, sGS-ADMM, and PDHG under mild conditions. We then present a gallery of LP, SOCP, and SDP examples, which exhibits residual fields, exact and approximate afterimages, and multiple certified slow regions, all matching the predicted vanishing moduli.

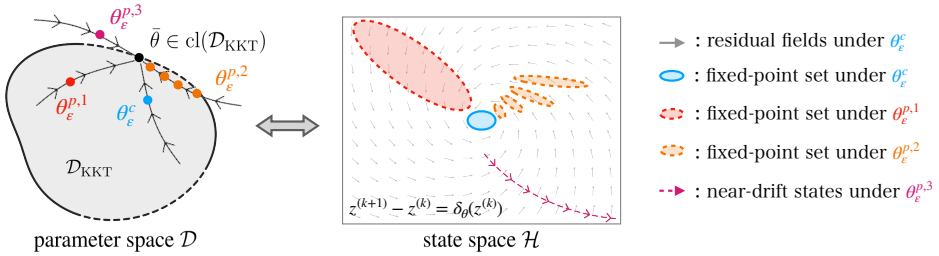

Figure 2: The afterimage framework. A *center* family (the problems we actually solve) and several *petal* families all converge to the same limit parameter $\bar\theta$, and induce nearly identical residual fields on bounded sets; their fixed-point sets and drift structures, however, differ sharply. Wherever a petal's structure is far from the center's, the center iterates make little progress toward their own fixed-point set for a long time, forming an *afterimage slow region*.

Organization. §2 sets up the assumptions and develops the afterimage principle; §3 presents the gallery; §4 concludes. Related work is reviewed in §A.1.

2 Afterimage Slow Regions in FOM for Conic-LP

2.1 Assumptions and Notation

Notation. Throughout, $(\mathcal{X}, \langle\cdot,\cdot\rangle_{\mathcal{X}})$, $(\mathcal{Y}, \langle\cdot,\cdot\rangle_{\mathcal{Y}})$, $(\mathcal{H}, \langle\cdot,\cdot\rangle_{\mathcal{H}})$ are finite-dimensional Hilbert spaces whose metrics are induced by the Euclidean inner product: $\mathcal{X}, \mathcal{Y}$ carry the conic-LP data of (P)–(D), namely $c \in \mathcal{X}$, $b \in \mathcal{Y}$, $\mathcal{A} \in \mathcal{L}(\mathcal{X}, \mathcal{Y})$ with adjoint $\mathcal{A}^*$, and a closed convex cone $\mathcal{K} \subseteq \mathcal{X}$ with dual cone $\mathcal{K}^*$, while $\mathcal{H}$ carries the state z of (1). Here $\mathcal{L}(\mathcal{X}, \mathcal{Y})$ denotes the space of linear operators from $\mathcal{X}$ to $\mathcal{Y}$ and $\mathcal{L}(\mathcal{H}) := \mathcal{L}(\mathcal{H}, \mathcal{H})$; $\|\cdot\|_{\mathrm{op}}$ and $\|\cdot\|_{\mathrm{HS}}$ are the operator and Hilbert–Schmidt norms; $\mathrm{Id}_{\mathcal{H}}$, abbreviated Id when unambiguous, is the identity on $\mathcal{H}$; and $B \succeq_{\mathcal{H}} A$, for self-adjoint $A, B \in \mathcal{L}(\mathcal{H})$, means $\langle z, (B-A)z\rangle_{\mathcal{H}} \geq 0$ for all $z \in \mathcal{H}$. If $\mathcal{A}\mathcal{A}^* \succ_{\mathcal{Y}} 0$, we write $\mathcal{A}^\dagger := \mathcal{A}^*(\mathcal{A}\mathcal{A}^*)^{-1} \in \mathcal{L}(\mathcal{Y}, \mathcal{X})$ for the Moore–Penrose pseudo-inverse, $P_{\mathcal{A}} := \mathcal{A}^\dagger\mathcal{A} \in \mathcal{L}(\mathcal{X})$ for the orthogonal projection onto $\mathrm{ran}(\mathcal{A}^*)$, and $P_{\mathcal{A}}^\perp := \mathrm{Id} - P_{\mathcal{A}}$. For a map $T : \mathcal{H} \to \mathcal{H}$, $\mathrm{ran}(T)$ and $\mathrm{Fix}(T)$ are its range and fixed point set, and we may abbreviate $T(z)$ as Tz; $\mathrm{Aff}_+ := \{L : \mathbb{R}_+ \to \mathbb{R}_+ \mid L(r) = a + br$ for some $a, b \geq 0\}$ is the set of *nonnegative affine moduli*. Given a norm $\|\cdot\|_\theta$ on $\mathcal{H}$ and a closed convex $C \subseteq \mathcal{H}$, Π_C^θ, $\mathrm{dist}_\theta(\cdot, C)$, and $\mathbb{B}_\theta(z, r)$ are the projection onto C, the distance to C, and the closed ball of radius r at z, all in $\|\cdot\|_\theta$, with $\Pi_C := \Pi_C^{\mathcal{H}}$ for the Euclidean norm; $\mathrm{cl}(C)$ is the closure of a set C, and for a set family $\{C_\varepsilon\}_{\varepsilon\downarrow 0} \subset \mathcal{H}$, $C_\varepsilon \to C$ as $\varepsilon \downarrow 0$ is meant in the Painlevé–Kuratowski sense. Finally, $[r] := \{1, \ldots, r\}$, and the conic-LP pair is fully determined by the parameter $\theta = (\mathcal{A}, b, c) \in \mathcal{D} := \mathcal{L}(\mathcal{X}, \mathcal{Y}) \times \mathcal{Y} \times \mathcal{X}$, which we equip with the product norm $\|\theta\|_{\mathcal{D}}^2 := \|\mathcal{A}\|_{\mathrm{HS}}^2 + \|b\|_{\mathcal{Y}}^2 + \|c\|_{\mathcal{X}}^2$.

Setup. Define the set of KKT points as

$$\mathrm{KKT}(\theta) := \{(x, y, s) \mid \mathcal{A}x = b,\ \mathcal{A}^*y + s = c,\ \langle x, s\rangle_{\mathcal{X}} = 0,\ x \in \mathcal{K},\ s \in \mathcal{K}^*\}, \tag{2}$$

and we define $\mathcal{D}_{\mathrm{KKT}} := \{\theta \in \mathcal{D} \mid \mathrm{KKT}(\theta) \neq \emptyset\}$. $\mathcal{D}_{\mathrm{KKT}}$ is not closed in general. We now formalize the assumptions imposed on the fixed point map in (1). Let $\bar\theta \in \mathrm{cl}(\mathcal{D}_{\mathrm{KKT}})$ and $U \subset \mathcal{D}$ be compact

with $\bar{\theta} \in U$; U is not required to have interior. For every $\theta \in U$, we associate a single-valued map $T_\theta : \mathcal{H} \to \mathcal{H}$. Let $\delta_\theta := T_\theta - \mathrm{Id}$ be T_θ's residual map (or *residual field*, when viewed spatially) (Kang & Yang, 2026).

Assumption 2.1 (Four standing assumptions)**.** *The map family $\{T_\theta\}_{\theta \in U}$ satisfies:*

(i) $\mathrm{Fix} T_\theta \neq \emptyset$ *iff* $\theta \in \mathcal{D}_{\mathrm{KKT}}$ *for every* $\theta \in U$.

(ii) For every $\theta \in U$, T_θ is α_θ-averaged with respect to the dynamic metric $\langle \cdot, \cdot \rangle_\theta$ for some $\alpha_\theta \in (0, 1)$, where $\langle z, w \rangle_\theta := \langle z, G_\theta w \rangle_{\mathcal{H}}$ and $G_\theta \in \mathcal{L}(\mathcal{H})$ is self-adjoint positive-definite.

(iii) There exist $0 < m_U \leq M_U < \infty$ such that $m_U \cdot \mathrm{Id} \preceq_{\mathcal{H}} G_\theta \preceq_{\mathcal{H}} M_U \cdot \mathrm{Id}$ for every $\theta \in U$.

(iv) The map $\mathcal{T} : U \times \mathcal{H} \to \mathcal{H}$, $\mathcal{T}(\theta, z) := T_\theta z$, is Lipschitz continuous in the parameter, with a modulus $L_U \in \mathrm{Aff}_+$:

$$\|T_\theta z - T_{\theta'} z\|_{\mathcal{H}} \leq L_U(\|z\|_{\mathcal{H}}) \cdot \|\theta - \theta'\|_{\mathcal{D}}, \quad \forall\, \theta, \theta' \in U,\ z \in \mathcal{H}. \tag{3}$$

Assumption 2.1 (ii) is standard in monotone operator theory (Ryu & Yin, 2022), while (i) is method-specific and is verified for each map of Table 1 in Lemma A.1. Assumption 2.1 (iii) makes the dynamic norms uniformly equivalent to the ambient norm: $\sqrt{m_U}\|z\|_{\mathcal{H}} \leq \|z\|_\theta \leq \sqrt{M_U}\|z\|_{\mathcal{H}}$. Assumption 2.1 (iv) states that nearby parameters induce uniformly close residual fields on any bounded state set,

$$\sup_{z \in S} \|\delta_\theta(z) - \delta_{\theta'}(z)\|_{\mathcal{H}} \leq L_U(r_S) \cdot \|\theta - \theta'\|_{\mathcal{D}}, \qquad r_S := \sup_{z \in S} \|z\|_{\mathcal{H}}, \tag{4}$$

for all $\theta, \theta' \in U$ and all bounded $S \subset \mathcal{H}$, since $\delta_\theta - \delta_{\theta'} = T_\theta - T_{\theta'}$.

In Table 1, we show that Assumption 2.1 holds for ADMM, PDHG, and sGS-ADMM under mild conditions. Appendix A.2 proves this and derives each method's modulus L_U, omitted from the table. Several comments on Table 1. First, the two additional conditions: $\mathcal{A}\mathcal{A}^* \succeq_{\mathcal{Y}} \kappa_U \cdot \mathrm{Id}$ for ADMM/sGS-ADMM and $\gamma_U < 1$ for PDHG are rather mild. The former is what makes the y-update (12a) and the representation $\mathcal{A}^\dagger = \mathcal{A}^*(\mathcal{A}\mathcal{A}^*)^{-1}$ well-defined; the latter is the standard condition making $G_\theta \succ_{\mathcal{H}} 0$, hence T_θ averaged in that metric. Moreover, they do not need to hold simultaneously across different algorithms. Second, while the z-form of ADMM is the classical Douglas–Rachford reduction, we are not aware of an explicit statement that sGS-ADMM shares the same z-map. However, the two methods carry different primal–dual iterates (x, y, s) along the same z-orbit, so the KKT residuals evaluated at their own iterates generally differ. Third, Assumption 2.1 does not cover every fixed point method. For instance, the exact Augmented Lagrangian Method (ALM) can fail (iv). See Appendix A.2 for more details.

Partial superposition propositions. In general the linear part of $L_U(\cdot)$ cannot be 0: Appendix A.2 exhibits $\theta, \theta' \in U$ along which $\|T_\theta z - T_{\theta'} z\|_{\mathcal{H}} \to \infty$ as $\|z\|_{\mathcal{H}} \to \infty$. Whenever it can be taken 0, however, (4) becomes a state-*uniform* bound, and in favorable cases the discrepancy is *constant* in the state: for all $\theta, \theta' \in U$,

$$\sup_{z \in \mathcal{H}} \|\delta_\theta z - \delta_{\theta'} z\|_{\mathcal{H}} \leq l_U \cdot \|\theta - \theta'\|_{\mathcal{D}}, \tag{5}$$

$$\delta_\theta z - \delta_{\theta'} z = \mathcal{B}(\theta - \theta'), \quad \forall\, z \in \mathcal{H}, \tag{6}$$

where l_U depends only on U, and $\mathcal{B} : \mathcal{D} \to \mathcal{H}$ is linear in the parameter and independent of the state; since $\|\mathcal{B}\|_{\mathrm{op}} < \infty$, (6) implies (5). We call (5) and (6) the partial superposition propositions, after the Superposition Principle of electrostatic fields.

Theorem 2.2 (Partial superposition propositions)**.** *Let Assumption 2.1 hold with $U \subset \{\mathcal{A}\} \times \mathcal{Y} \times \mathcal{X}$,* i.e., *$\mathcal{A}$ is fixed while (b, c) may vary. Then (i)* (6) *holds for ADMM/sGS-ADMM; and (ii)* (5) *holds for PDHG, upgrading to* (6) *when every $\theta \in U$ also shares the same c.*

The proof is a direct computation from Table 1, deferred to Appendix A.2.

2.2 Afterimage slow region family

Given $\bar{\theta} \in \mathrm{cl}(\mathcal{D}_{\mathrm{KKT}})$, a compact $U \ni \bar{\theta}$, and Assumption 2.1, we take two families in U, both indexed by $\varepsilon \downarrow 0$: the *center* family $\{\theta^{\mathrm{c}}_\varepsilon\}_{\varepsilon \downarrow 0}$ and the *petal* family $\{\theta^{\mathrm{p}}_\varepsilon\}_{\varepsilon \downarrow 0}$, with

$$\{\theta^{\mathrm{c}}_\varepsilon\}_{\varepsilon \downarrow 0} \subset U \cap \mathcal{D}_{\mathrm{KKT}}, \quad \{\theta^{\mathrm{p}}_\varepsilon\}_{\varepsilon \downarrow 0} \subset U, \quad \theta^{\mathrm{c}}_\varepsilon \to \bar{\theta},\ \theta^{\mathrm{p}}_\varepsilon \to \bar{\theta},\ \text{as } \varepsilon \downarrow 0. \tag{7}$$

	ADMM	sGS-ADMM	PDHG
Conditions on U	$\sigma > 0$, $\tau \in (0,2)$, and $\mathcal{A}\mathcal{A}^* \succeq_{\mathcal{Y}} \kappa_U \cdot \mathrm{Id}$ for some $\kappa_U > 0$ and all $\theta \in U$		$\eta_x, \eta_y > 0$ and $\gamma_U := \sup_{\theta\in U} \sqrt{\eta_x\eta_y}\|\mathcal{A}\|_{\mathrm{op}} < 1$
$\mathcal{H}$	$\mathcal{X}$		$\mathcal{X} \times \mathcal{Y}$
G_θ	$P_{\mathcal{A}} + \tau^{-1} P_{\mathcal{A}}^{\perp}$		$\begin{bmatrix} \eta_x^{-1} \cdot \mathrm{Id}_{\mathcal{X}} & \mathcal{A}^* \\ \mathcal{A} & \eta_y^{-1} \cdot \mathrm{Id}_{\mathcal{Y}} \end{bmatrix}$
α_θ	$1/2 \cdot \max\{1, \tau\}$		$1/2$
m_U	$\min\{1, \tau^{-1}\}$		$(1-\gamma_U)\max\{\eta_x, \eta_y\}^{-1}$
M_U	$\max\{1, \tau^{-1}\}$		$(1+\gamma_U)\min\{\eta_x, \eta_y\}^{-1}$
$T_\theta z$	$z - P_{\mathcal{A}}[\Pi_{\mathcal{K}}(z) - \mathcal{A}^\dagger b]$ $- \tau P_{\mathcal{A}}^{\perp}[-\Pi_{\mathcal{K}^*}(-z) + \sigma c]$		$(x^+,\ y + \eta_y[b - \mathcal{A}(2x^+ - x)])$, $z = (x, y)$, $x^+ := \Pi_{\mathcal{K}}[x - \eta_x(c - \mathcal{A}^* y)]$

Table 1: Verification of Assumption 2.1 for three first-order methods. All entries are derived in Appendix A.2. In their z-variable form, ADMM (12) and sGS-ADMM (20) induce the *same* fixed point map, hence share every entry above. Here σ is the penalty parameter and τ is the multiplier step length of ADMM and sGS-ADMM, while η_x and η_y are the primal and the dual step sizes of PDHG; all of them are fixed over U.

The center family carries the conic-LP problems we actually study, while the petal family will certify slow regions for the center iteration. Only the center parameters are required to admit a KKT point; θ and $\theta_\varepsilon^{\mathrm{p}}$ need not. We abbreviate $T_{\theta_\varepsilon^{\mathrm{c}}}, T_{\theta_\varepsilon^{\mathrm{p}}}$ as $T_\varepsilon^{\mathrm{c}}, T_\varepsilon^{\mathrm{p}}$, and likewise for the residual map δ, and write $V_\varepsilon^{\mathrm{c}} := \mathrm{Fix}(T_\varepsilon^{\mathrm{c}})$. Throughout, all conditions on ε-indexed families are required to hold only for all sufficiently small ε.

Slow region family. We define the slow region (SR) family of $\{\theta_\varepsilon^{\mathrm{c}}\}_{\varepsilon\downarrow 0}$ by a limiting process:

Definition 2.3 (Slow region family)**.** *Suppose Assumption 2.1 holds. A set family $\{\mathcal{S}_\varepsilon\}_{\varepsilon\downarrow 0} \subset \mathcal{H}$ is called a slow region (SR) family of $\{\theta_\varepsilon^{\mathrm{c}}\}_{\varepsilon\downarrow 0}$, iff (i) $\mathcal{S}_\varepsilon \neq \emptyset$ and $\mathcal{S}_\varepsilon \cap V_\varepsilon^{\mathrm{c}} = \emptyset$; (ii) its* vanishing modulus

$$\rho_\varepsilon := \sup_{z\in\mathcal{S}_\varepsilon} \frac{\|\delta_\varepsilon^{\mathrm{c}}(z)\|_{\mathcal{H}}}{\mathrm{dist}_{\mathcal{H}}(z, V_\varepsilon^{\mathrm{c}})} \tag{8}$$

is finite and $\rho_\varepsilon \to 0$ as $\varepsilon \downarrow 0$.

Equivalently, ρ_ε is the least constant L with $\|\delta_\varepsilon^{\mathrm{c}}(z)\|_{\mathcal{H}} \le L \cdot \mathrm{dist}_{\mathcal{H}}(z, V_\varepsilon^{\mathrm{c}})$ on $\mathcal{S}_\varepsilon$; the best constant κ_ε in the error bound $\mathrm{dist}_{\mathcal{H}}(z, V_\varepsilon^{\mathrm{c}}) \le \kappa_\varepsilon \|\delta_\varepsilon^{\mathrm{c}}(z)\|_{\mathcal{H}}$ over $\mathcal{S}_\varepsilon$ therefore satisfies $\kappa_\varepsilon \ge \rho_\varepsilon^{-1}$, so an SR family is one on which this error-bound constant blows up; (i) already forces $\rho_\varepsilon > 0$. The definition is purely spatial and imposes no regularity on $\{\mathcal{S}_\varepsilon\}_{\varepsilon\downarrow 0}$ or $\{V_\varepsilon^{\mathrm{c}}\}_{\varepsilon\downarrow 0}$; in particular, $\mathcal{S}_\varepsilon$ collects starting states and need not be invariant: an orbit may leave $\mathcal{S}_\varepsilon$ after one step. For the maps of Table 1, the residual is also observable: $\|\delta_\theta(z)\|_{\mathcal{H}}$ is equivalent to the projected KKT residual of a primal–dual pair computed from the state z (Lemma A.2). Propositions 2.4–2.6 collect its basic properties. Due to page limits, all proofs in this subsection are deferred to Appendix A.3.

Proposition 2.4 (Persistence away from the fixed point set)**.** *Under Definition 2.3's setting, fix $\beta \in (0,1)$ and define $N_\varepsilon := \lfloor (1-\beta)\sqrt{m_U/M_U} \cdot \rho_\varepsilon^{-1} \rfloor \to \infty$. Then:*

$$\mathrm{dist}_{\mathcal{H}}((T_\varepsilon^{\mathrm{c}})^k(z), V_\varepsilon^{\mathrm{c}}) \ge \sqrt{m_U/M_U}\beta \cdot \mathrm{dist}_{\mathcal{H}}(z, V_\varepsilon^{\mathrm{c}}), \quad \forall z \in \mathcal{S}_\varepsilon,\ 0 \le k \le N_\varepsilon. \tag{9}$$

Proposition 2.5 (Finite unions)**.** *Let $\{\mathcal{S}_\varepsilon^i\}_{\varepsilon\downarrow 0}$, $i \in [r]$ with $r \ge 2$, be SR families for the same $\{\theta_\varepsilon^{\mathrm{c}}\}_{\varepsilon\downarrow 0}$, with vanishing moduli $\{\rho_\varepsilon^i\}_{\varepsilon\downarrow 0}$. Then $\{\cup_{i\in[r]}\mathcal{S}_\varepsilon^i\}_{\varepsilon\downarrow 0}$ is an SR family for $\{\theta_\varepsilon^{\mathrm{c}}\}_{\varepsilon\downarrow 0}$ with vanishing modulus exactly $\{\max_{i\in[r]} \rho_\varepsilon^i\}_{\varepsilon\downarrow 0}$.*

Proposition 2.5 does not extend verbatim to countably many families (Appendix A.3).

Proposition 2.6 (Thickening)**.** *Under Definition 2.3's setting, let $\{a_\varepsilon\}_{\varepsilon\downarrow 0}$ be a real number family with $0 \le a_\varepsilon < 1$ and $a_\varepsilon \to 0$ as $\varepsilon \downarrow 0$. Define $\tilde{\mathcal{S}}_\varepsilon := \{z \in \mathcal{H} \mid \exists q \in \mathcal{S}_\varepsilon, \text{ s.t. } \|z - q\|_{\mathcal{H}} \le a_\varepsilon \cdot \mathrm{dist}_{\mathcal{H}}(q, V_\varepsilon^{\mathrm{c}})\}$. Then $\{\tilde{\mathcal{S}}_\varepsilon\}_{\varepsilon\downarrow 0}$ is also an SR family of $\{\theta_\varepsilon^{\mathrm{c}}\}_{\varepsilon\downarrow 0}$, with vanishing modulus $\tilde{\rho}_\varepsilon \le (\rho_\varepsilon + 2\sqrt{M_U/m_U} \cdot a_\varepsilon)/(1 - a_\varepsilon)$.*

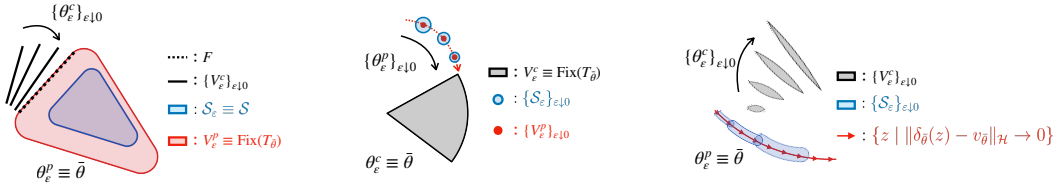

Figure 3: The three cases of Theorems 2.9–2.11, all drawn in the state space $\mathcal{H}$.

Afterimage slow region family. Finding an SR family directly is hard. However, if a petal family $\{\theta_\varepsilon^{\mathrm{p}}\}_{\varepsilon\downarrow 0}$ as in (7) has limiting behavior far from the center's, the SR family can often be certified. Before formalizing this idea, we record a fact from monotone operator theory, again proved in Appendix A.3.

Proposition 2.7 (Forward drift). *Under Assumption 2.1, take any $\theta \in U$, and let $z^{(k+1)} = T_\theta z^{(k)}$ from an arbitrary $z^{(0)} \in \mathcal{H}$. Then $\delta_\theta z^{(k)} \to v_\theta := \Pi^\theta_{\mathrm{cl}(\mathrm{ran}(\delta_\theta))}(0)$ as $k \to \infty$, and $\|\delta_\theta z^{(k)}\|_\theta$ is nonincreasing in k with limit $\|v_\theta\|_\theta$. If $\delta_\theta(z) = v_\theta$ for some $z \in \mathcal{H}$, then $\delta_\theta(z + tv_\theta) = v_\theta$ for every $t \geq 0$. Finally, if $\mathrm{Fix}(T_\theta) = \emptyset$, there is a sequence $\{z_j\} \subset \mathcal{H}$ with $\|z_j\|_{\mathcal{H}} \to \infty$ and $\delta_\theta(z_j) \to v_\theta$. We call v_θ the* forward drift.

Under Assumption 2.1, δ_θ and v_θ are well-defined whether or not θ yields a KKT point. Define $\mathrm{tag}(\theta) \in \{0, +\} \times \{\mathrm{fin}, \mathrm{inf}\}$ as follows: the first entry is 0 iff $v_\theta = 0$, and the second is fin iff $v_\theta \in \mathrm{ran}(\delta_\theta)$, *i.e.,* the projection is attained. The tag depends on the map family and its metric, not on θ alone. Then $\mathrm{tag}(\theta) = (0, \mathrm{fin})$ iff $\theta \in \mathcal{D}_{\mathrm{KKT}}$; the other three cases collect the pathological parameters — infeasible, weakly infeasible, nonattaining, or carrying a positive duality gap — and all three are realized in §3 and Appendix A.4. For comparison, homogeneous self-dual embeddings recover primal–dual solutions or certificates of strong infeasibility, while weak infeasibility and other duality pathologies may lead to a degenerate embedding (Odonoghue et al., 2016; Permenter et al., 2017). The tags $\mathrm{tag}(\bar{\theta})$, $\mathrm{tag}(\theta_\varepsilon^{\mathrm{c}})$ and $\mathrm{tag}(\theta_\varepsilon^{\mathrm{p}})$ may all differ. Abbreviating $v_{\theta_\varepsilon^{\mathrm{p}}}$ as $v_\varepsilon^{\mathrm{p}}$, we can now define the afterimage slow region (ASR) family.

Definition 2.8 (Afterimage slow region family). *Under Assumption 2.1, let $\{\mathcal{S}_\varepsilon\}_{\varepsilon\downarrow 0}$ be a slow region (SR) family for the center parameter family $\{\theta_\varepsilon^{\mathrm{c}}\}_{\varepsilon\downarrow 0}$. We call $\{\mathcal{S}_\varepsilon\}_{\varepsilon\downarrow 0}$ an afterimage slow region (ASR) family w.r.t. a petal parameter family $\{\theta_\varepsilon^{\mathrm{p}}\}_{\varepsilon\downarrow 0}$, iff there exists a nonnegative real number family $\{\mu_\varepsilon\}_{\varepsilon\downarrow 0}$, with $\mu_\varepsilon \to 0$ as $\varepsilon \downarrow 0$, s.t.*

$$\|\delta_\varepsilon^{\mathrm{p}}(z) - v_\varepsilon^{\mathrm{p}}\|_{\mathcal{H}} \leq \mu_\varepsilon, \quad \forall z \in \mathcal{S}_\varepsilon. \tag{10}$$

Similar to $V_\varepsilon^{\mathrm{c}} = \mathrm{Fix}(T_\varepsilon^{\mathrm{c}})$, we let $V_{\theta_\varepsilon^{\mathrm{p}}} := \{z \in \mathcal{H} \mid \delta_\varepsilon^{\mathrm{p}}(z) = v_\varepsilon^{\mathrm{p}}\}$, the fixed point set of the shifted map $z \mapsto T_\varepsilon^{\mathrm{p}} z - v_\varepsilon^{\mathrm{p}}$, and abbreviate it as $V_\varepsilon^{\mathrm{p}}$. If $\mathrm{tag}(\theta_\varepsilon^{\mathrm{p}})$ is $(0, \mathrm{fin})$ or $(+, \mathrm{fin})$ for all ε, then $V_\varepsilon^{\mathrm{p}} \neq \emptyset$; choosing $\mathcal{S}_\varepsilon \subseteq V_\varepsilon^{\mathrm{p}}$, as every example below with a finite petal tag does, we may then set $\mu_\varepsilon \equiv 0$. When it is $(0, \mathrm{inf})$ or $(+, \mathrm{inf})$, since $v_\varepsilon^{\mathrm{p}}$ is non-attainable and $V_\varepsilon^{\mathrm{p}} = \emptyset$, μ_ε must be positive. To see how the petal family helps to certify the ASR family, we strengthen the condition in (8) by asking for a family $\{\hat{\rho}_\varepsilon\}_{\varepsilon\downarrow 0}$ with $\hat{\rho}_\varepsilon \to 0$ and

$$\hat{\rho}_\varepsilon \cdot \mathrm{dist}_{\mathcal{H}}(z, V_\varepsilon^{\mathrm{c}}) \geq \|\delta_\varepsilon^{\mathrm{p}}(z)\|_{\mathcal{H}} + \|\delta_\varepsilon^{\mathrm{p}}(z) - \delta_\varepsilon^{\mathrm{c}}(z)\|_{\mathcal{H}}, \quad \forall z \in \mathcal{S}_\varepsilon. \tag{11}$$

By the triangle inequality, $\rho_\varepsilon \leq \hat{\rho}_\varepsilon$; hence (11) certifies (8) and exposes two residual field components we need to control. Intuitively, $\|\delta_\varepsilon^{\mathrm{p}}(z) - \delta_\varepsilon^{\mathrm{c}}(z)\|_{\mathcal{H}}$ is controlled by (4), while (10) gives $\|\delta_\varepsilon^{\mathrm{p}}(z)\|_{\mathcal{H}} \leq \|v_\varepsilon^{\mathrm{p}}\|_{\mathcal{H}} + \mu_\varepsilon$. We now illustrate Definition 2.8 through three special cases, drawn in Figure 3: $V_\varepsilon^{\mathrm{c}}$ collapses onto a proper face of $\mathrm{Fix}(T_{\bar{\theta}})$, which is closed and convex (Lemma A.3); the petal fixed points are much farther from $\mathrm{Fix}(T_{\bar{\theta}})$ than the data perturbation; and $V_\varepsilon^{\mathrm{c}}$ escapes to infinity. Each is only a sufficient condition: §3 also constructs regions that none of the three covers, working directly from Definitions 2.3 and 2.8.

Theorem 2.9 (Face selection cases). *Under Assumption 2.1, set $\mathrm{tag}(\bar{\theta}) = (0, \mathrm{fin})$ and $\theta_\varepsilon^{\mathrm{p}} \equiv \bar{\theta}$. Suppose $V_\varepsilon^{\mathrm{c}} \to F$ as $\varepsilon \downarrow 0$, where F is a proper face of $\mathrm{Fix}(T_{\bar{\theta}})$, and take $\mathcal{S}_\varepsilon \equiv \mathcal{S} \subset \mathrm{Fix}(T_{\bar{\theta}})$ nonempty, bounded, with $\inf_{z \in \mathcal{S}} \mathrm{dist}_{\mathcal{H}}(z, F) > 0$. Then $\{\mathcal{S}_\varepsilon\}_{\varepsilon\downarrow 0}$ is an ASR family with $\rho_\varepsilon = O(\|\theta_\varepsilon^{\mathrm{c}} - \bar{\theta}\|_{\mathcal{D}})$.*

Proof. Painlevé–Kuratowski convergence $V_\varepsilon^{\mathrm{c}} \to F$ with F nonempty makes $\mathrm{dist}_{\mathcal{H}}(\cdot, V_\varepsilon^{\mathrm{c}}) \to \mathrm{dist}_{\mathcal{H}}(\cdot, F)$ uniformly on bounded sets (Rockafellar & Wets, 1998, Theorem 4.35). As $\mathcal{S}$ is bounded, $d_\varepsilon := \inf_{z\in\mathcal{S}} \mathrm{dist}_{\mathcal{H}}(z, V_\varepsilon^{\mathrm{c}}) \to \inf_{z\in\mathcal{S}} \mathrm{dist}_{\mathcal{H}}(z, F) =: d_\star > 0$, so $d_\varepsilon > 0$ and $\mathcal{S} \cap V_\varepsilon^{\mathrm{c}} = \emptyset$ for small ε. Put $r_{\mathcal{S}} := \sup_{z\in\mathcal{S}} \|z\|_{\mathcal{H}} < \infty$. Since $\theta_\varepsilon^{\mathrm{p}} \equiv \bar{\theta}$ and $\mathcal{S} \subset \mathrm{Fix}(T_{\bar{\theta}})$, $\delta_\varepsilon^{\mathrm{p}} \equiv 0$ on $\mathcal{S}$, giving (10) with $\mu_\varepsilon \equiv 0$; and (4) bounds $e_\varepsilon := \sup_{z\in\mathcal{S}} \|\delta_\varepsilon^{\mathrm{c}}(z) - \delta_\varepsilon^{\mathrm{p}}(z)\|_{\mathcal{H}} \le L_U(r_{\mathcal{S}}) \cdot \|\theta_\varepsilon^{\mathrm{c}} - \bar{\theta}\|_{\mathcal{D}}$. Take $\hat{\rho}_\varepsilon := e_\varepsilon / d_\varepsilon$, so $\rho_\varepsilon \le \hat{\rho}_\varepsilon$. Every $z \in \mathcal{S}$ then obeys $\hat{\rho}_\varepsilon \cdot \mathrm{dist}_{\mathcal{H}}(z, V_\varepsilon^{\mathrm{c}}) \ge e_\varepsilon \ge \|\delta_\varepsilon^{\mathrm{p}}(z)\|_{\mathcal{H}} + \|\delta_\varepsilon^{\mathrm{p}}(z) - \delta_\varepsilon^{\mathrm{c}}(z)\|_{\mathcal{H}}$, which is (11), and $\rho_\varepsilon = O(\|\theta_\varepsilon^{\mathrm{c}} - \bar{\theta}\|_{\mathcal{D}})$ since $d_\varepsilon \to d_\star > 0$. □

Theorem 2.10 (Hölder sensitivity cases)**.** *Under Assumption 2.1, set* $\mathrm{tag}(\bar{\theta}) = (0, \mathrm{fin})$ *and* $\theta_\varepsilon^{\mathrm{c}} \equiv \bar{\theta}$. *Suppose that for some* $\gamma \in (0,1)$ *and all sufficiently small* ε: $\theta_\varepsilon^{\mathrm{p}} \ne \bar{\theta}$, $\mathrm{tag}(\theta_\varepsilon^{\mathrm{p}}) = (0, \mathrm{fin})$, *and there are* witnesses $z_\varepsilon \in V_\varepsilon^{\mathrm{p}}$ *with* $\sup_\varepsilon \|z_\varepsilon\|_{\mathcal{H}} < \infty$ *and* $\mathrm{dist}_{\mathcal{H}}(z_\varepsilon, \mathrm{Fix}(T_{\bar{\theta}})) = \Theta(\|\theta_\varepsilon^{\mathrm{p}} - \bar{\theta}\|_{\mathcal{D}}^{\gamma})$. *Then there is an ASR family* $\{\mathcal{S}_\varepsilon\}_{\varepsilon\downarrow 0}$ *with* $\rho_\varepsilon = O(\|\theta_\varepsilon^{\mathrm{p}} - \bar{\theta}\|_{\mathcal{D}}^{1-\gamma})$*; under* (5) *the witnesses need not be bounded.*

Proof. Write $h_\varepsilon := \|\theta_\varepsilon^{\mathrm{p}} - \bar{\theta}\|_{\mathcal{D}} > 0$ and $\tilde{d}_\varepsilon := \mathrm{dist}_{\mathcal{H}}(z_\varepsilon, \mathrm{Fix}(T_{\bar{\theta}}))$, so that $C_0 h_\varepsilon^{\gamma} \ge \tilde{d}_\varepsilon \ge c_0 h_\varepsilon^{\gamma} > 0$ for some $C_0 \ge c_0 > 0$. Here $V_\varepsilon^{\mathrm{c}} = \mathrm{Fix}(T_{\bar{\theta}})$ because $\theta_\varepsilon^{\mathrm{c}} \equiv \bar{\theta}$, and $\mathrm{tag}(\theta_\varepsilon^{\mathrm{p}}) = (0, \mathrm{fin})$ gives $v_\varepsilon^{\mathrm{p}} = 0$ with $V_\varepsilon^{\mathrm{p}} \ne \emptyset$. Fix $\alpha \in (0,1)$ and set $\mathcal{S}_\varepsilon := V_\varepsilon^{\mathrm{p}} \cap \mathbb{B}_{\mathcal{H}}(z_\varepsilon, \alpha\tilde{d}_\varepsilon)$, which contains z_ε. As $\mathcal{S}_\varepsilon \subseteq V_\varepsilon^{\mathrm{p}}$, $\delta_\varepsilon^{\mathrm{p}} \equiv 0$ there and (10) holds with $\mu_\varepsilon \equiv 0$. By 1-Lipschitzness of $\mathrm{dist}_{\mathcal{H}}(\cdot, V_\varepsilon^{\mathrm{c}})$, $\inf_{z\in\mathcal{S}_\varepsilon} \mathrm{dist}_{\mathcal{H}}(z, V_\varepsilon^{\mathrm{c}}) \ge (1-\alpha)\tilde{d}_\varepsilon \ge c \cdot h_\varepsilon^{\gamma} > 0$ with $c := (1-\alpha)c_0$; in particular $\mathcal{S}_\varepsilon \cap V_\varepsilon^{\mathrm{c}} = \emptyset$. Moreover $\sup_{z\in\mathcal{S}_\varepsilon} \|z\|_{\mathcal{H}} \le \sup_\varepsilon (\|z_\varepsilon\|_{\mathcal{H}} + \alpha C_0 h_\varepsilon^{\gamma}) =: R < \infty$, where only the witnesses enter, so (4) gives $\sup_{z\in\mathcal{S}_\varepsilon} \|\delta_\varepsilon^{\mathrm{p}}(z) - \delta_\varepsilon^{\mathrm{c}}(z)\|_{\mathcal{H}} \le L_U(R) h_\varepsilon$. Take $\hat{\rho}_\varepsilon := L_U(R)/c \cdot h_\varepsilon^{1-\gamma}$, so $\rho_\varepsilon \le \hat{\rho}_\varepsilon$, which tends to 0 as $\gamma < 1$. Every $z \in \mathcal{S}_\varepsilon$ obeys $\hat{\rho}_\varepsilon \cdot \mathrm{dist}_{\mathcal{H}}(z, V_\varepsilon^{\mathrm{c}}) \ge L_U(R) h_\varepsilon \ge \|\delta_\varepsilon^{\mathrm{p}}(z)\|_{\mathcal{H}} + \|\delta_\varepsilon^{\mathrm{p}}(z) - \delta_\varepsilon^{\mathrm{c}}(z)\|_{\mathcal{H}}$, which is (11). Under (5), l_U replaces $L_U(R)$ throughout and no bound on R is needed. □

Theorem 2.11 (Escape-to-horizon cases)**.** *Under Assumption 2.1, set* $\theta_\varepsilon^{\mathrm{p}} \equiv \bar{\theta}$ *and suppose* $D_\varepsilon := \inf_{z\in V_\varepsilon^{\mathrm{c}}} \|z\|_{\mathcal{H}} \to \infty$ *as* $\varepsilon \downarrow 0$. *Fix* $\gamma \in (0,1)$ *and a nonnegative family* $\{\nu_\varepsilon\}_{\varepsilon\downarrow 0}$ *with* $\nu_\varepsilon \to 0$. *With* $g_{\bar{\theta}}(D) := \min_{\|z\|_{\mathcal{H}} \le D} \|\delta_{\bar{\theta}}(z) - v_{\bar{\theta}}\|_{\mathcal{H}}$, *let* $\mathcal{S}_\varepsilon := \{z \in \mathbb{B}_{\mathcal{H}}(0, D_\varepsilon^{\gamma}) \mid \|\delta_{\bar{\theta}}(z) - v_{\bar{\theta}}\|_{\mathcal{H}} \le g_{\bar{\theta}}(D_\varepsilon^{\gamma}) + \nu_\varepsilon\}$. *Then* $\{\mathcal{S}_\varepsilon\}_{\varepsilon\downarrow 0}$ *is an ASR family with* $\rho_\varepsilon = O(D_\varepsilon^{-1} + D_\varepsilon^{\gamma-1} \cdot \|\theta_\varepsilon^{\mathrm{c}} - \bar{\theta}\|_{\mathcal{D}})$.

Proof. Here $g_{\bar{\theta}}$ is nonincreasing with $g_{\bar{\theta}}(D) \to 0$ as $D \to \infty$: when $\mathrm{Fix}(T_{\bar{\theta}}) = \emptyset$, Proposition 2.7 supplies $\{z_j\}$ with $\|z_j\|_{\mathcal{H}} \to \infty$ and $\delta_{\bar{\theta}}(z_j) \to v_{\bar{\theta}}$; otherwise $v_{\bar{\theta}} = 0$ and $g_{\bar{\theta}}$ vanishes on every ball meeting $\mathrm{Fix}(T_{\bar{\theta}})$. The minimum defining $g_{\bar{\theta}}(D_\varepsilon^{\gamma})$ is attained, $\mathbb{B}_{\mathcal{H}}(0, D_\varepsilon^{\gamma})$ being compact and $\delta_{\bar{\theta}}$ continuous, and any minimizer lies in $\mathcal{S}_\varepsilon$, so $\mathcal{S}_\varepsilon \ne \emptyset$. Since $\theta_\varepsilon^{\mathrm{p}} \equiv \bar{\theta}$ we have $\delta_\varepsilon^{\mathrm{p}} = \delta_{\bar{\theta}}$ and $v_\varepsilon^{\mathrm{p}} = v_{\bar{\theta}}$, so (10) holds with $\mu_\varepsilon := g_{\bar{\theta}}(D_\varepsilon^{\gamma}) + \nu_\varepsilon \to 0$. Every $z \in \mathcal{S}_\varepsilon$ has $\|z\|_{\mathcal{H}} \le D_\varepsilon^{\gamma}$ while every $w \in V_\varepsilon^{\mathrm{c}}$ has $\|w\|_{\mathcal{H}} \ge D_\varepsilon$, so $\mathrm{dist}_{\mathcal{H}}(z, V_\varepsilon^{\mathrm{c}}) \ge D_\varepsilon - D_\varepsilon^{\gamma} > 0$ for small ε; in particular $\mathcal{S}_\varepsilon \cap V_\varepsilon^{\mathrm{c}} = \emptyset$. On $\mathcal{S}_\varepsilon$, $\|\delta_\varepsilon^{\mathrm{p}}(z)\|_{\mathcal{H}} \le \|v_{\bar{\theta}}\|_{\mathcal{H}} + \mu_\varepsilon$, while (4) gives $\|\delta_\varepsilon^{\mathrm{p}}(z) - \delta_\varepsilon^{\mathrm{c}}(z)\|_{\mathcal{H}} \le L_U(D_\varepsilon^{\gamma}) \cdot \|\theta_\varepsilon^{\mathrm{c}} - \bar{\theta}\|_{\mathcal{D}}$. Hence (11) holds with

$$\hat{\rho}_\varepsilon := \frac{L_U(D_\varepsilon^{\gamma}) \cdot \|\theta_\varepsilon^{\mathrm{c}} - \bar{\theta}\|_{\mathcal{D}} + \|v_{\bar{\theta}}\|_{\mathcal{H}} + \mu_\varepsilon}{D_\varepsilon - D_\varepsilon^{\gamma}} = O(D_\varepsilon^{\gamma-1} \cdot \|\theta_\varepsilon^{\mathrm{c}} - \bar{\theta}\|_{\mathcal{D}} + D_\varepsilon^{-1}).$$

The order holds because $L_U \in \mathrm{Aff}_+$ makes $L_U(D_\varepsilon^{\gamma})/(D_\varepsilon - D_\varepsilon^{\gamma}) = O(D_\varepsilon^{\gamma-1})$ while the rest of the numerator stays bounded. As $\gamma < 1$, $D_\varepsilon \to \infty$ and $\|\theta_\varepsilon^{\mathrm{c}} - \bar{\theta}\|_{\mathcal{D}} \to 0$, we get $\rho_\varepsilon \le \hat{\rho}_\varepsilon \to 0$. □

3 Afterimage Example Gallery

We exhibit ASR families across all three problem classes — LP, with $\mathcal{X} = \mathbb{R}^n$ and $\mathcal{K} = \mathbb{R}^n_+$; SOCP, with $\mathcal{X} = \mathbb{R}^n$ and $\mathcal{K} = \mathcal{Q}^n := \{(t, v) : t \ge \|v\|_2\}$; and SDP, with $\mathcal{X} = \mathbb{S}^n$ and $\mathcal{K} = \mathbb{S}^n_+$, where capital letters denote matrices — under both ADMM/sGS-ADMM and PDHG. Example 3.1 treats both maps and verifies Assumption 2.1.

Example 3.1 (Motivating example continued). Recall (LP_ε), in which only b moves: $\bar{\theta} := (\bar{A}, 0, c)$, $\theta_\varepsilon^{\mathrm{c}} := (\bar{A}, \varepsilon, c)$, $\theta_\varepsilon^{\mathrm{p}} \equiv \bar{\theta}$ and $\|\theta_\varepsilon^{\mathrm{c}} - \bar{\theta}\|_{\mathcal{D}} = \varepsilon$. Both tags equal $(0, \mathrm{fin})$, and Assumption 2.1 holds on a small compact $U \ni \bar{\theta}$, since Table 1's side conditions are strict at $\bar{\theta}$ (Appendix A.4.1). The dual degeneracy turns into a collapse of the fixed point sets: $\mathcal{L} := \mathrm{Fix}T_{\bar{\theta}}$ is $\{(-(1-t)\sigma, -t\sigma)\}_{t\in[0,1]}$ for ADMM and $\{(0, 0, y)\}_{y\in[0,1]}$ for PDHG, while $V_\varepsilon^{\mathrm{c}}$ is the single point $(\varepsilon, -\sigma)$, resp. $(\varepsilon, 0, 1)$, collapsing onto an endpoint F. Only b moves, so (6) holds and Theorem 2.9 gives, on $\mathcal{S} := \{z \in \mathcal{L} : \mathrm{dist}_{\mathcal{H}}(z, F) \ge c_0\}$ for any $c_0 > 0$ below the length of $\mathcal{L}$, $\mu_\varepsilon \equiv 0$ and $\rho_\varepsilon \le \varepsilon/(\sqrt{2}c_0)$, resp. $\eta_y \varepsilon / c_0$ — the scalings displayed in Figures 1 and 4. Trimming the margin to $c_\star \varepsilon^{\alpha}$, any $c_\star > 0$ below the length of $\mathcal{L}$ when $\alpha = 0$, gives $\rho_\varepsilon = \Theta(\varepsilon^{1-\alpha})$ for $\alpha < 1$ and $\Theta(1)$ beyond, an ASR family

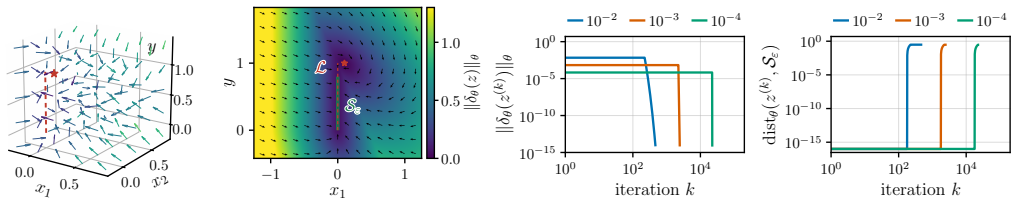

Figure 4: Example 3.1 under PDHG, $\eta_x = \eta_y = 0.45$. *Left two:* δ_θ on $\mathcal{H} = \mathbb{R}^3$ at $\varepsilon = 0.1$, and its slice $\{x_2 = 0\}$ carrying $\mathcal{L} = \mathrm{Fix}T_{\bar\theta}$ (dashed), $V_\varepsilon^{\mathrm{c}}$ (star) and $\mathcal{S}$ (solid); arrows are in-slice components. *Right two:* $\|\delta_\theta(z^{(k)})\|_\theta$ and $\mathrm{dist}_\theta(z^{(k)}, \mathcal{S})$ against k, from one ε-independent seed, at $\varepsilon = 10^{-2}, 10^{-3}, 10^{-4}$.

precisely for $\alpha \in [0, 1)$, certified directly in Appendix A.4.1 rather than by Theorem 2.9. Among ε-independent singletons, $\mathcal{L} \setminus \{F\}$ is exactly the slow set, and fixed-parameter restart and Halpern schemes change only the constants of the $\Theta(\varepsilon^{-1})$ plateau (Appendix A.4.1).

Each example below uses one of the two maps, takes Assumption 2.1 as given, and draws its residual field on a representative 2-D slice, with $\sigma = \tau = 1$ for ADMM and $\eta_x = \eta_y = 0.45$ for PDHG. Each contributes a feature the others lack. (i) Example 3.2 realizes Theorem 2.10 at $\gamma = 0.5$, where the residual and ρ_ε scale by different powers of the perturbation. (ii) Example 3.3 attaches two petals to one center, and in the plotted run a single orbit passes slowly through both carriers in succession. (iii) Example 3.4 has $\mathrm{Fix}T_{\bar\theta} = \emptyset$ because the limit is strongly infeasible, and the residual on its first region is independent of ε. Examples 3.3 and 3.4 each carry a region outside the three cases of §2.2. Appendix A.4 adds three more: (iv) Example A.4, where the petal fixed set, the surviving face and the slow region are all positive-dimensional, with $V_\varepsilon^{\mathrm{c}} \subsetneq V_\varepsilon^{\mathrm{p}}$ exact at every $\varepsilon > 0$; (v) Example A.5, whose drift is zero and unattained; and (vi) Example A.6, which loses attainment and zero drift together, since $\mathcal{A}(\mathcal{K})$ fails to be closed. The last two have $V_\varepsilon^{\mathrm{p}} = \emptyset$, so $\mu_\varepsilon > 0$ is forced. Two kinds of slowness appear below: on some regions the residual itself tends to zero, while in the escape examples it stays of order one and the slowness is relative to a fixed-point set that runs away.

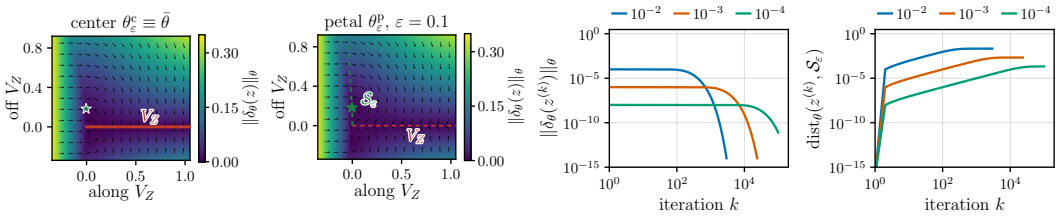

Figure 5: Example 3.2 under ADMM; layout as in Figure 4, on the plane through $V_\varepsilon^{\mathrm{c}}$ spanned by the ray and the dislocation; dashed green traces $V_\varepsilon^{\mathrm{p}} = \{Z_\varepsilon\}$ (star) as $\varepsilon \downarrow 0$. Here the seed Z_ε *moves with* ε: the residual sits near ε^2 while $\mathrm{dist}_\theta(z^{(k)}, \mathcal{S}_\varepsilon)$ grows immediately, $\mathcal{S}_\varepsilon$ being one point.

Example 3.2 (SDP square-root Hölder). We now *freeze* the center and move the petal. Let $\mathcal{K} = \mathbb{S}^2_+$, $A_1 := \left[\begin{smallmatrix} 0 & 1 \\ 1 & -1 \end{smallmatrix}\right]$, $\bar{\mathcal{A}}(X) := \langle A_1, X\rangle_{\mathcal{X}}$ and E_{ij} the matrix units: $\bar\theta = \theta_\varepsilon^{\mathrm{c}} :\equiv (\bar{\mathcal{A}},\, 0,\, E_{22})$, $\theta_\varepsilon^{\mathrm{p}} := (\bar{\mathcal{A}},\, \varepsilon^3,\, E_{22} + \varepsilon^2 E_{11})$ and $h_\varepsilon := \|\theta_\varepsilon^{\mathrm{p}} - \bar\theta\|_{\mathcal{D}} = \sqrt{\varepsilon^6 + \varepsilon^4}$, with $\mathrm{tag}(\bar\theta) = (0, \mathrm{fin})$. The center carries the ray $V_\varepsilon^{\mathrm{c}} = \mathrm{Fix}T_{\bar\theta} = \{\mathrm{diag}\,(t, -1) : t \geq 0\}$ and the petal the single point $V_\varepsilon^{\mathrm{p}} = \{Z_\varepsilon\}$, dislocated by $\mathrm{dist}_{\mathcal{H}}(Z_\varepsilon, V_\varepsilon^{\mathrm{c}}) = \sqrt{3}\,\varepsilon + O(\varepsilon^2) = \Theta(h_\varepsilon^{1/2})$ (Appendix A.4.2). This is Theorem 2.10 at $\gamma = 0.5$: taking $\mathcal{S}_\varepsilon = V_\varepsilon^{\mathrm{p}}$, whose state-free defect has size $\varepsilon^2\sqrt{1 + \varepsilon^2/3}$, gives $\rho_\varepsilon = \Theta(\varepsilon)$; see Figure 5. Here $\mathcal{S}_\varepsilon$ is a single *moving* point, so the seed depends on ε, and there is no invariant carrier to move along.

Example 3.3 (Two afterimages, two exponents). We keep the SDP data of Example 3.2 but read it *backwards* — what was the center there is a petal here — and add a second petal. Writing $C_\varepsilon := E_{22} + \varepsilon^2 E_{11}$: $\bar\theta := (\bar{\mathcal{A}},\, 0,\, E_{22})$, $\theta_\varepsilon^{\mathrm{c}} := (\bar{\mathcal{A}},\, \varepsilon^3,\, C_\varepsilon)$, $\theta_\varepsilon^{\mathrm{p},1} :\equiv \bar\theta$ and $\theta_\varepsilon^{\mathrm{p},2} := (\bar{\mathcal{A}},\, 0,\, C_\varepsilon)$, with gaps $\Theta(\varepsilon^2)$ and ε^3 and $\mathrm{tag}(\bar\theta) = (0, \mathrm{fin})$ as before. Three fixed sets are involved (Appendix A.4.3): the ray $V_\varepsilon^{\mathrm{p},1} = V_Z := \{Q_t := \mathrm{diag}\,(t, -1) : t \geq 0\}$; the point $V_\varepsilon^{\mathrm{c}} = \{Z_\varepsilon\}$ of Example 3.2, which collapses onto Q_0 yet stays $\sqrt{3}\varepsilon + O(\varepsilon^2)$ *off* the ray; and the segment $V_\varepsilon^{\mathrm{p},2} = \{q(y) := -C_\varepsilon + yA_1\}_{y \in [y_-, y_+]}$, with $y_\pm$ the roots of the dual PSD condition $y^2 = \varepsilon^2(1 + y)$. Both petals differ from the center

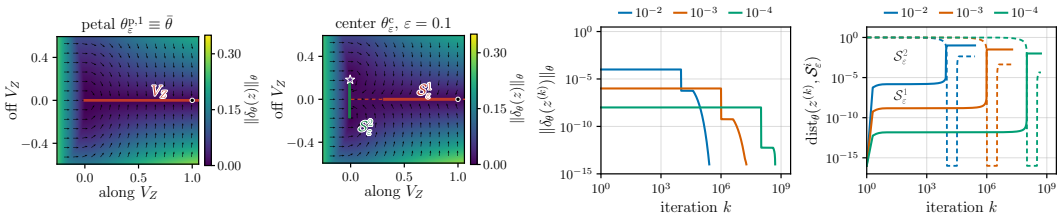

Figure 6: Example 3.3 under ADMM; layout as in Figure 4. *Left two:* the ray V_Z, then its collapse to Z_ε (star), both petal carriers surviving with trimmed parts solid. *Right two:* one orbit from Q_{t_1}, $\alpha = 0.5$, $\eta = 0.25$; solid and dashed are the distances to $\mathcal{S}_\varepsilon^1$, $\mathcal{S}_\varepsilon^2$.

in b and/or c only, so Theorem 2.2 (i) makes both defects state-free; each $\delta_\varepsilon^{\mathrm{p}}$ vanishes on its own carrier, so $\mu_\varepsilon \equiv 0$. Fix $t_1 \geq 1$, $\alpha \in [0, 1)$ and $\eta \in (0, 1)$: the regions $\mathcal{S}_\varepsilon^1 := \{Q_t : \varepsilon^\alpha \leq t \leq t_1\}$ and $\mathcal{S}_\varepsilon^2 := \{q(y) : y_- \leq y \leq (1-\eta)y_+\}$ have $\rho_\varepsilon^1 = \Theta(\varepsilon^{2-\alpha})$ and $\rho_\varepsilon^2 = \Theta(\varepsilon^2)$. Region 1 is face selection with a tunable trim, certified directly in Appendix A.4.3. Region 2 falls outside §2.2: $\theta_\varepsilon^{\mathrm{c}}$ and $\theta_\varepsilon^{\mathrm{p},2}$ both move and neither equals $\bar{\theta}$, so Definitions 2.3 and 2.8 certify it directly. One center thus carries two afterimages whose *residuals* differ by a full order of ε; see Figure 6.

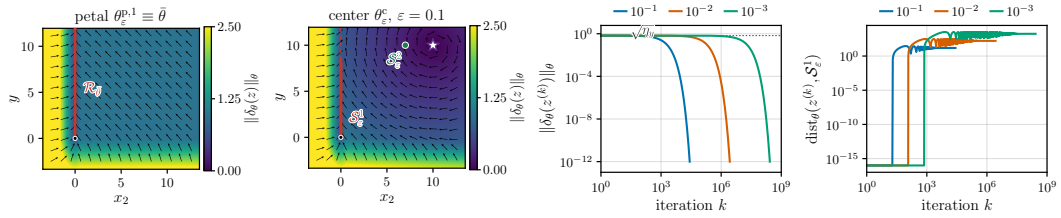

Figure 7: Example 3.4 under PDHG; layout as in Figure 4, on the slice $\{x_1 = 0\}$. *Left two:* with no fixed point at the limit the color never reaches 0, so $\mathcal{R}_{\bar{\theta}}$ is drawn instead; then one appears (star) and escapes, $\mathcal{S}_\varepsilon^1$ solid and $\mathcal{S}_\varepsilon^2$ the green dot. *Right two:* one orbit from $z^{(0)} = 0$, $\gamma = 0.8$; the three plateaus share the same height $\sqrt{\eta_y}$ and only lengthen.

Example 3.4 (Escape to the horizon). Let $\mathcal{K} = \mathbb{R}_+^2$, $c := (0, 1)$, $a := 1/\varepsilon$ and $0 < \xi_\varepsilon \to 0$: $\bar{\theta} := ([-1 \;\; 0], 1, c)$, $\theta_\varepsilon^{\mathrm{c}} := ([-1 \;\; \varepsilon], 1, c)$, $\theta_\varepsilon^{\mathrm{p},1} :\equiv \bar{\theta}$ and $\theta_\varepsilon^{\mathrm{p},2} := ([-1 \;\; \varepsilon], 1 - \xi_\varepsilon, c)$, with gaps ε and ξ_ε. The limit is strongly infeasible, so $\mathrm{Fix} T_{\bar{\theta}} = \emptyset$, while its drift $v_{\bar{\theta}} = (0, 0, \eta_y)$ is attained on the ray $\mathcal{R}_{\bar{\theta}} := \{(0, 0, y) : y \geq 0\} = V_\varepsilon^{\mathrm{p},1}$: $\mathrm{tag}(\bar{\theta}) = (+, \mathrm{fin})$. The center is solvable, but its unique fixed point $V_\varepsilon^{\mathrm{c}} = \{(0, a, a)\}$ *escapes*, $D_\varepsilon := \inf_{z \in V_\varepsilon^{\mathrm{c}}} \|z\|_{\mathcal{H}} = \sqrt{2}/\varepsilon \to \infty$. For any $\gamma \in (0, 1)$, the regions $\mathcal{S}_\varepsilon^1 := \mathcal{R}_{\bar{\theta}} \cap \mathbb{B}_{\mathcal{H}}(0, D_\varepsilon^\gamma)$ and $\mathcal{S}_\varepsilon^2 := V_\varepsilon^{\mathrm{p},2} = \{(0, (1 - \xi_\varepsilon)a, a)\}$ have $\mu_\varepsilon \equiv 0$, $\rho_\varepsilon^1 = \Theta(\varepsilon)$ and $\rho_\varepsilon^2 = \eta_y \varepsilon$ exactly (Appendix A.4.4). On $\mathcal{S}_\varepsilon^1$ the residual equals $\|v_{\bar{\theta}}\|_\theta = \sqrt{\eta_y}$, independent of ε; Region 2 again falls outside §2.2, since both $\theta_\varepsilon^{\mathrm{c}}$ and $\theta_\varepsilon^{\mathrm{p},2}$ move and neither equals $\bar{\theta}$.

4 Conclusion

We introduced a framework that certifies slow regions in first-order methods for conic-LP: a region family is slow once the residual vanishes relative to the distance to the fixed-point set of its own parameter. The afterimage principle explains where such families come from: a vanishing perturbation moves the residual field continuously, while the geometry it carries can change abruptly, so that the geometry of a neighboring problem survives as a slow region of the problem we actually solve. The standing assumptions hold for ADMM, sGS-ADMM, and PDHG. Moreover, the mechanism requires neither a polyhedral cone nor a well-posed limit, so the gallery covers infeasible and non-attaining problems. Our framework also has two limitations. First, it is purely spatial: it measures the strength of the residual field on a region, but says nothing about the dynamics — whether typical orbits enter a slow region, or for how long they stay. Second, every certificate starts from a petal family designed by hand; we do not yet know how to find one for a given, non-designed instance. Two directions are worth exploring in future work: locating afterimage slow regions in specific conic-LP families, and designing acceleration schemes around them.

REPRODUCIBILITY STATEMENT

All assumptions, proofs, and derivations are provided in the main text and appendices. The figures involving numerical iterations are generated from the algorithms described in the paper; the corresponding code will be open-sourced upon acceptance.

AI USE STATEMENT

The framework of this paper — its assumptions, definitions, theorems, and the example gallery with its derivations — was developed and written by the authors. Large language models (LLMs) were used to copy-edit and compress the exposition throughout, and to review the draft adversarially. The authors take full responsibility for this paper.

REFERENCES

Jacob M Aguirre, Diego Cifuentes, Vincent Guigues, Renato DC Monteiro, Victor Hugo Nascimento, and Arnesh Sujanani. cuhallar: A gpu accelerated low-rank augmented lagrangian method for large-scale semidefinite programming. *arXiv preprint arXiv:2505.13719*, 2025.

David Applegate, Mateo Díaz, Oliver Hinder, Haihao Lu, Miles Lubin, Brendan O'Donoghue, and Warren Schudy. Practical large-scale linear programming using primal-dual hybrid gradient. In *Conference on Neural Information Processing Systems (NeurIPS)*, volume 34, pp. 20243–20257, 2021.

David Applegate, Oliver Hinder, Haihao Lu, and Miles Lubin. Faster first-order primal-dual methods for linear programming using restarts and sharpness. *Mathematical Programming*, 201(1):133–184, 2023.

David Applegate, Mateo Díaz, Haihao Lu, and Miles Lubin. Infeasibility detection with primal-dual hybrid gradient for large-scale linear programming. *SIAM Journal on Optimization*, 34(1):459–484, 2024.

Jean-Bernard Baillon, Ronald E. Bruck, and Simeon Reich. On the asymptotic behavior of nonexpansive mappings and semigroups in banach spaces. *Houston Journal of Mathematics*, 4(1):1–9, 1978.

Goran Banjac. On the minimal displacement vector of the douglas–rachford operator. *Operations Research Letters*, 49(2):197–200, 2021.

Heinz H. Bauschke and Patrick L. Combettes. *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*. Springer, Heidelberg, Germany, 2 edition, 2017.

Antonio Bellon, Didier Henrion, Vyacheslav Kungurtsev, and Jakub Mareček. Parametric semidefinite programming: geometry of the trajectory of solutions. *Mathematics of Operations Research*, 50 (1):410–430, 2025.

Daniel Boley. Local linear convergence of the alternating direction method of multipliers on quadratic or linear programs. *SIAM Journal on Optimization*, 23(4):2183–2207, 2013.

Ronald E Bruck and Simeon Reich. Nonexpansive projections and resolvents of accretive operators in banach spaces. *Houston J. Math*, 3(4):459–470, 1977.

Antonin Chambolle and Thomas Pock. A first-order primal-dual algorithm for convex problems with applications to imaging. *Journal of mathematical imaging and vision*, 40(1):120–145, 2011.

Raymond H Chan, Shiqian Ma, and Junfeng Yang. Inertial primal-dual algorithms for structured convex optimization. *arXiv preprint arXiv:1409.2992*, 2014.

Liang Chen, Defeng Sun, and Kim-Chuan Toh. An efficient inexact symmetric gauss–seidel based majorized admm for high-dimensional convex composite conic programming. *Mathematical Programming*, 161(1):237–270, 2017.

Liang Chen, Xudong Li, Defeng Sun, and Kim-Chuan Toh. On the equivalence of inexact proximal alm and admm for a class of convex composite programming: L. chen et al. *Mathematical Programming*, 185(1):111–161, 2021.

Ying Cui, Defeng Sun, and Kim-Chuan Toh. On the asymptotic superlinear convergence of the augmented lagrangian method for semidefinite programming with multiple solutions. *arXiv preprint arXiv:1610.00875*, 2016.

Alexandre d'Aspremont, Laurent Ghaoui, Michael Jordan, and Gert Lanckriet. A direct formulation for sparse pca using semidefinite programming. *Advances in neural information processing systems*, 17, 2004.

Zhanwang Deng, Jiang Hu, Kangkang Deng, and Zaiwen Wen. An efficient primal dual semismooth newton method for semidefinite programming. *arXiv preprint arXiv:2504.14333*, 2025.

Lijun Ding and Madeleine Udell. A strict complementarity approach to error bound and sensitivity of solution of conic programs: L. ding, m. udell. *Optimization Letters*, 17(7):1551–1574, 2023.

Lijun Ding, Haihao Lu, and Jinwen Yang. New understandings and computation on augmented lagrangian methods for low-rank semidefinite programming. *arXiv preprint arXiv:2505.15775*, 2025.

Lijun Ding, Haihao Lu, and Jinwen Yang. On the local linear convergence of operator splitting methods for conic programming. *arXiv preprint arXiv:2608.16054*, 2026.

Robert M Freund and Jorge R Vera. Some characterizations and properties of the "distance to ill-posedness" and the condition measure of a conic linear system. *Mathematical Programming*, 86(2):225–260, 1999.

Daniel Gabay and Bertrand Mercier. A dual algorithm for the solution of nonlinear variational problems via finite element approximation. *Computers & mathematics with applications*, 2(1): 17–40, 1976.

Michael Garstka, Mark Cannon, and Paul Goulart. Cosmo: A conic operator splitting method for convex conic problems. *Journal of Optimization Theory and Applications*, 190(3):779–810, 2021.

Antoine Groudiev, Shucheng Kang, and Heng Yang. cuADMM: GPU-Accelerated First-Order Optimization for Large-Scale Multi-Block Semidefinite Programs. In *RSS Workshop on Fast Motion Planning and Control in the Era of Parallelism*, 2025.

John Guckenheimer and Philip Holmes. *Nonlinear oscillations, dynamical systems, and bifurcations of vector fields*. Springer Science & Business Media, 2013.

Deren Han, Defeng Sun, and Liwei Zhang. Linear rate convergence of the alternating direction method of multipliers for convex composite programming. *Mathematics of Operations Research*, 43(2):622–637, 2018.

Qiushi Han, Chenxi Li, Zhenwei Lin, Caihua Chen, Qi Deng, Dongdong Ge, Huikang Liu, and Yinyu Ye. A low-rank admm splitting approach for semidefinite programming. *INFORMS Journal on Computing*, 2025.

Alan J Hoffman. On approximate solutions of systems of linear inequalities. In *Selected Papers Of Alan J Hoffman: With Commentary*, pp. 174–176. World Scientific, 2003.

Di Hou, Tianyun Tang, and Kim-Chuan Toh. A low-rank augmented lagrangian method for doubly nonnegative relaxations of mixed-binary quadratic programs. *Operations Research*, 74(3):1626–1647, 2026.

Tao Jiang, Walaa M Moursi, and Stephen A Vavasis. Range of the displacement operator of pdhg with applications to quadratic and conic programming. *arXiv preprint arXiv:2309.15009*, 2023.

Xin Jiang. Local linear convergence of the primal-dual hybrid gradient method for semidefinite programming. *arXiv preprint arXiv:2607.08035*, 2026.

Shucheng Kang and Heng Yang. Local second-order limit dynamics of the alternating direction method of multipliers for semidefinite programming. *arXiv preprint arXiv:2602.20103*, 2026.

Shucheng Kang, Xin Jiang, and Heng Yang. Local linear convergence of the alternating direction method of multipliers for semidefinite programming under strict complementarity. *arXiv preprint arXiv:2503.20142*, 2025.

Yuri A Kuznetsov. *Elements of applied bifurcation theory*. Springer, 1998.

Gert RG Lanckriet, Nello Cristianini, Peter Bartlett, Laurent El Ghaoui, and Michael I Jordan. Learning the kernel matrix with semidefinite programming. *Journal of Machine learning research*, 5(Jan):27–72, 2004.

Hongpei Li, Yicheng Huang, Huikang Liu, Dongdong Ge, and Yinyu Ye. Gpu-accelerated conic quadratic programming with local linear convergence under strict complementarity. *arXiv preprint arXiv:2608.09159*, 2026.

Yongfeng Li, Zaiwen Wen, Chao Yang, and Ya-xiang Yuan. A semismooth newton method for semidefinite programs and its applications in electronic structure calculations. *SIAM Journal on Scientific Computing*, 40(6):A4131–A4157, 2018.

Zhenwei Lin, Zikai Xiong, Dongdong Ge, and Yinyu Ye. Pdcs: A primal-dual large-scale conic programming solver with gpu enhancements. *arXiv preprint arXiv:2505.00311*, 2025.

Tianhao Liu and Haihao Lu. A new crossover algorithm for lp inspired by the spiral dynamic of pdhg. *arXiv preprint arXiv:2409.14715*, 2024.

Yanli Liu, Ernest K Ryu, and Wotao Yin. A new use of douglas-rachford splitting and admm for identifying infeasible, unbounded, and pathological conic programs. *arXiv preprint arXiv:1706.02374*, 2017.

Miguel Sousa Lobo, Lieven Vandenberghe, Stephen Boyd, and Hervé Lebret. Applications of second-order cone programming. *Linear algebra and its applications*, 284(1-3):193–228, 1998.

Haihao Lu and Jinwen Yang. Restarted halpern pdhg for linear programming. *arXiv preprint arXiv:2407.16144*, 2024.

Haihao Lu and Jinwen Yang. On the geometry and refined rate of primal–dual hybrid gradient for linear programming. *Mathematical Programming*, 212(1):349–387, 2025a.

Haihao Lu and Jinwen Yang. cupdlp. jl: A gpu implementation of restarted primal-dual hybrid gradient for linear programming in julia. *Operations Research*, 73(6):3440–3452, 2025b.

Olvi L Mangasarian and RR Meyer. Nonlinear perturbation of linear programs. *SIAM Journal on Control and Optimization*, 17(6):745–752, 1979.

Renato DC Monteiro, Arnesh Sujanani, and Diego Cifuentes. A low-rank augmented lagrangian method for large-scale semidefinite programming based on a hybrid convex-nonconvex approach: Rdc monteiro et al. *Mathematical Programming*, pp. 1–57, 2026.

Brendan Odonoghue, Eric Chu, Neal Parikh, and Stephen Boyd. Conic optimization via operator splitting and homogeneous self-dual embedding. *Journal of Optimization Theory and Applications*, 169(3):1042–1068, 2016.

Amnon Pazy. Asymptotic behavior of contractions in hilbert space. *Israel Journal of Mathematics*, 9 (2):235–240, 1971.

Frank Permenter, Henrik A Friberg, and Erling D Andersen. Solving conic optimization problems via self-dual embedding and facial reduction: a unified approach. *SIAM Journal on Optimization*, 27(3):1257–1282, 2017.

Clarice Poon and Jingwei Liang. Trajectory of alternating direction method of multipliers and adaptive acceleration. *Advances in neural information processing systems*, 32, 2019.

Clarice Poon and Jingwei Liang. Geometry of first-order methods and adaptive acceleration. *arXiv preprint arXiv:2003.03910*, 2020.

James Renegar. Some perturbation theory for linear programming. *Mathematical programming*, 65 (1):73–91, 1994.

R Tyrrell Rockafellar. Monotone operators and the proximal point algorithm. *SIAM journal on control and optimization*, 14(5):877–898, 1976.

R Tyrrell Rockafellar and Roger JB Wets. *Variational analysis*. Springer, Heidelberg, Germany, 1998.

Ernest K Ryu and Wotao Yin. *Large-scale convex optimization: algorithms & analyses via monotone operators*. Cambridge University Press, Cambridge, 2022.

Yousef Saad, Man Yeung, Jocelyne Erhel, and Frédéric Guyomarc'h. A deflated version of the conjugate gradient algorithm. *SIAM Journal on Scientific Computing*, 21(5):1909–1926, 2000.

Jos F Sturm. Error bounds for linear matrix inequalities. *SIAM Journal on Optimization*, 10(4): 1228–1248, 2000.

Tianyun Tang and Kim-Chuan Toh. A feasible method for general convex low-rank sdp problems. *SIAM Journal on Optimization*, 34(3):2169–2200, 2024.

Lloyd N. Trefethen and Mark Embree. *Spectra and Pseudospectra: The Behavior of Nonnormal Matrices and Operators*. Princeton University Press, Princeton, NJ, USA, 2005.

Jie Wang, Liangbing Hu, and Bican Xia. A dual riemannian admm algorithm for low-rank sdps with unit diagonal. *arXiv preprint arXiv:2512.04406*, 2025.

Kilian Q. Weinberger and Lawrence K. Saul. Unsupervised learning of image manifolds by semidefinite programming. In *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 2, pp. 988–995, 2004. doi: 10.1109/CVPR.2004.1315272.

Zaiwen Wen, Donald Goldfarb, and Wotao Yin. Alternating direction augmented Lagrangian methods for semidefinite programming. *Mathematical Programming Computation*, 2(3):203–230, 2010.

Zikai Xiong. Accessible complexity bounds for restarted pdhg on linear programs with a unique optimizer. *Mathematics of Operations Research*, 2026.

Zikai Xiong and Robert M Freund. The role of level-set geometry on the performance of pdhg for conic linear optimization. *arXiv preprint arXiv:2406.01942*, 2024.

Zikai Xiong and Robert M Freund. Computational guarantees for restarted pdhg for lp based on "limiting error ratios" and lp sharpness. *Mathematical Programming*, pp. 1–62, 2026.

Liuqin Yang, Defeng Sun, and Kim-Chuan Toh. SDPNAL+: A majorized semismooth Newton-CG augmented Lagrangian method for semidefinite programming with nonnegative constraints. *Mathematical Programming Computation*, 7(3):331–366, 2015.

Xin-Yuan Zhao, Defeng Sun, and Kim-Chuan Toh. A newton-cg augmented lagrangian method for semidefinite programming. *SIAM Journal on Optimization*, 20(4):1737–1765, 2010.

Yang Zheng, Giovanni Fantuzzi, Antonis Papachristodoulou, Paul Goulart, and Andrew Wynn. Fast ADMM for homogeneous self-dual embedding of sparse SDPs. *IFAC-PapersOnLine*, 50(1): 8411–8416, 2017.

A Appendix

A.1 Related Work

Our afterimage slow region framework is related to three lines of work: recent solver developments for FOM on conic-LP, convergence theory of FOM on conic-LP, and dynamical system theory.

Recent solver developments in FOM for conic-LP. ADMM, sGS-ADMM and PDHG have become popular choices for large-scale conic-LP, and a number of solvers are built upon them (Odonoghue et al., 2016; Zheng et al., 2017; Garstka et al., 2021). A related family of methods maintains a second-order outer loop without ever forming the full Newton system. Semismooth Newton (SSN) is applied either to the Augmented Lagrangian Method (ALM) subproblem (Zhao et al., 2010; Yang et al., 2015) or directly to the fixed-point residual of (1) (Li et al., 2018; Deng et al., 2025), with the Newton system solved inexactly by conjugate gradient. Alternatively, a low-rank factorization reduces the size of the subproblem, which is then handled by Riemannian methods on the resulting manifold (Tang & Toh, 2024; Hou et al., 2026) or by a hybrid convex–nonconvex scheme (Monteiro et al., 2026). Like FOM, these methods remain sensitive to conditioning. More recently, GPU implementations of FOM for conic-LP have appeared (Lu & Yang, 2025b; Lin et al., 2025; Han et al., 2025; Groudiev et al., 2025; Aguirre et al., 2025; Ding et al., 2025).

Recent theoretical developments in FOM for conic-LP. As general splitting schemes, FOM have well-established worst-case sublinear rates in monotone operator theory (Ryu & Yin, 2022), provided the fixed point set is nonempty. On conic-LP, however, local linear convergence is observed much more frequently than the worst-case rate suggests, even on degenerate instances. For LP this is classical: global sharpness follows from polyhedral error bounds (Hoffman, 2003; Mangasarian & Meyer, 1979), and later work has further refined it for PDHG (Applegate et al., 2021; Lu & Yang, 2025a). For ADMM on LP/QP, Boley (2013) identifies constant-step transients before the local linear phase. General conic-LP is much harder. Metric subregularity of the KKT operator, which would imply a local linear rate, can fail (Cui et al., 2016). Han et al. (2018) recover it under uniqueness of the KKT point together with a two-sided second-order regularity condition. For a non-unique KKT point set, Kang et al. (2025) prove local linear convergence of ADMM on SDP whenever the limit point satisfies strict complementarity, and further generalizations have followed (Jiang, 2026; Li et al., 2026; Ding et al., 2026; Wang et al., 2025). These analyses are anchored at a KKT point of the instance being solved; the afterimage certificate is instead anchored at a neighboring problem. Another line works on the pathological instance itself, using the minimal displacement vector of DRS/ADMM and PDHG as an infeasibility certificate (Banjac, 2021; Liu et al., 2017; Jiang et al., 2023; Applegate et al., 2024); the afterimage instead uses a *neighboring* problem's drift to certify slow states of a solvable center. A separate line of work measures an instance's distance from ill-posedness. For a feasible conic system, the size of the smallest data perturbation that destroys feasibility governs both the geometry of the feasible region and the cost of solving it (Renegar, 1994; Freund & Vera, 1999); more recent works on restarted PDHG map an LP or conic instance to a condition measure, and then to an iteration count (Xiong & Freund, 2024; 2026; Xiong, 2026). We instead locate the slow states inside $\mathcal{H}$, allowing non-polyhedral cones and ill-posed limits. Closest to us, Bellon et al. (2025) classify how the optimal set of a parametric SDP degenerates; we study how such degeneration affects the iterates.

Dynamical system theory. Locating slow transients inside a convergent dynamical system is a classical difficulty. Bifurcation theory, especially center manifold theory, isolates slow motion by linearizing at an equilibrium and splitting the spectrum (Guckenheimer & Holmes, 2013; Kuznetsov, 1998). For averaged nonsmooth maps, however, a smooth linearization to split is generally unavailable, and there is no distinguished critical direction. We therefore certify a transient region instead of an invariant manifold, and this region need not contain a fixed point. Many algorithms in numerical analysis can likewise be viewed as dynamical systems. Pseudospectral theory, for instance, explains the transient behavior of non-normal operators and of the eigenvalue algorithms built on them (Trefethen & Embree, 2005). Within optimization, Poon & Liang (2019; 2020) study the trajectory and geometry of FOM; for LP, Liu & Lu (2024) identify the spiral trajectories of PDHG within a fixed basis; and for ADMM on degenerate SDP, Kang & Yang (2026) give a second-order spatial analysis of the residual field and of the slow regions surrounding a singular KKT point. These works motivate us to look for slow regions in FOM for conic-LP in general.

A.2 Additional material for §2.1

In this appendix, we prove Table 1 by verifying Assumption 2.1 for ADMM, sGS-ADMM, and PDHG.

Why exact ALM is not covered. Exact ALM is the proximal point method applied to the dual (Rockafellar, 1976), with state $y \in \mathcal{Y}$ and $T_\theta(y) = \Pi_{D_\theta}(y + \sigma b)$, where $D_\theta := \{y \in \mathcal{Y} \mid c - \mathcal{A}^* y \in \mathcal{K}^*\}$ itself carries θ and need not vary continuously with it. Take $\mathcal{X} = \mathbb{R}^2$, $\mathcal{K} = \mathbb{R}^2_+$, $\mathcal{Y} = \mathbb{R}$, and $\theta_\varepsilon := (\mathcal{A}_\varepsilon, 0, (1, 0))$ with $\mathcal{A}_\varepsilon := [1 \ \varepsilon]$, $\varepsilon \geq 0$. Every θ_ε is a solvable LP of optimal value 0 with attained dual; the rank condition of Table 1 holds uniformly. Since $c - \mathcal{A}_\varepsilon^* y = (1 - y, \ -\varepsilon y)$, we get $D_{\theta_\varepsilon} = (-\infty, 0]$ for every $\varepsilon > 0$ while $D_{\theta_0} = (-\infty, 1]$: the dual feasible set collapses discontinuously in the limit. At the single state $y = 0.5$, $T_{\theta_\varepsilon}(0.5) = 0$ but $T_{\theta_0}(0.5) = 0.5$, yet $\|\theta_\varepsilon - \theta_0\|_{\mathcal{D}} = \varepsilon$; Assumption 2.1 (iv) would demand $0.5 \leq L_U(0.5)\,\varepsilon$ for all small ε, which no $L_U \in \mathrm{Aff}_+$ satisfies.

ADMM case. Fix $\sigma > 0$ and $\tau \in (0, 2)$. Suppose $\mathcal{A}\mathcal{A}^* \succeq_{\mathcal{Y}} \kappa_U \cdot \mathrm{Id}$ with constant $\kappa_U > 0$. Consider the classical three-step ADMM with multiplier step length τ:

$$y^{(k+1)} = (\mathcal{A}\mathcal{A}^*)^{-1}[\sigma^{-1} b - \mathcal{A}(\sigma^{-1} x^{(k)} + s^{(k)} - c)], \tag{12a}$$

$$s^{(k+1)} = \Pi_{\mathcal{K}^*}(c - \mathcal{A}^* y^{(k+1)} - \sigma^{-1} x^{(k)}), \tag{12b}$$

$$x^{(k+1)} = x^{(k)} + \tau\sigma(\mathcal{A}^* y^{(k+1)} + s^{(k+1)} - c). \tag{12c}$$

For $k \geq 0$, define the auxiliary state variable $z^{(k)} := P_{\mathcal{A}}^{\perp} x^{(k)} + \mathcal{A}^\dagger b - \sigma P_{\mathcal{A}} s^{(k)} - \sigma P_{\mathcal{A}}^{\perp} c$. Then, from (12b), $s^{(k+1)} = \Pi_{\mathcal{K}^*}[c - \sigma^{-1}\mathcal{A}^\dagger b + P_{\mathcal{A}}(\sigma^{-1} x^{(k)} + s^{(k)} - c) - \sigma^{-1} x^{(k)}] = \sigma^{-1}\Pi_{\mathcal{K}^*}(-z^{(k)})$. Furthermore, from (12c),

$$\begin{aligned}
x^{(k+1)} &= x^{(k)} + \tau[\mathcal{A}^\dagger b - P_{\mathcal{A}}(x^{(k)} + \sigma s^{(k)} - \sigma c) + \Pi_{\mathcal{K}^*}(-z^{(k)}) - \sigma c] \\
&= (1 - \tau) x^{(k)} + \tau\Pi_{\mathcal{K}^*}(-z^{(k)}) + \tau(P_{\mathcal{A}}^{\perp} x^{(k)} - \sigma P_{\mathcal{A}} s^{(k)} + \mathcal{A}^\dagger b - \sigma P_{\mathcal{A}}^{\perp} c) \\
&= (1 - \tau) x^{(k)} + \tau\Pi_{\mathcal{K}^*}(-z^{(k)}) + \tau z^{(k)} = (1 - \tau) x^{(k)} + \tau\Pi_{\mathcal{K}}(z^{(k)}),
\end{aligned}$$

where we use Moreau identity $z = \Pi_{\mathcal{K}}(z) - \Pi_{\mathcal{K}^*}(-z)$ in the last equality. Put $x^{(k+1)}$ and $s^{(k+1)}$'s new formula into $z^{(k+1)} = P_{\mathcal{A}}^{\perp} x^{(k+1)} + \mathcal{A}^\dagger b - \sigma P_{\mathcal{A}} s^{(k+1)} - \sigma P_{\mathcal{A}}^{\perp} c$, and with the observation that $P_{\mathcal{A}}^{\perp} z^{(k)} = P_{\mathcal{A}}^{\perp} x^{(k)} - \sigma P_{\mathcal{A}}^{\perp} c$ for all $k \geq 0$, we get

$$\begin{aligned}
z^{(k+1)} &= (1 - \tau) P_{\mathcal{A}}^{\perp} x^{(k)} + \tau P_{\mathcal{A}}^{\perp}\Pi_{\mathcal{K}}(z^{(k)}) + \mathcal{A}^\dagger b - P_{\mathcal{A}}\Pi_{\mathcal{K}^*}(-z^{(k)}) - \sigma P_{\mathcal{A}}^{\perp} c \\
&= (1 - \tau) P_{\mathcal{A}}^{\perp}(z^{(k)} + \sigma c) + \tau P_{\mathcal{A}}^{\perp}[z^{(k)} + \Pi_{\mathcal{K}^*}(-z^{(k)})] + \mathcal{A}^\dagger b + P_{\mathcal{A}}[z^{(k)} - \Pi_{\mathcal{K}}(z^{(k)})] - \sigma P_{\mathcal{A}}^{\perp} c \\
&= z^{(k)} - P_{\mathcal{A}}[\Pi_{\mathcal{K}}(z^{(k)}) - \mathcal{A}^\dagger b] - \tau P_{\mathcal{A}}^{\perp}[-\Pi_{\mathcal{K}^*}(-z^{(k)}) + \sigma c],
\end{aligned} \tag{13}$$

which is exactly $T_\theta(z^{(k)})$'s form. To recover the primal–dual variable iterates from $z^{(k)}$'s:

$$x^{(k+1)} = (1 - \tau) x^{(k)} + \tau\Pi_{\mathcal{K}}(z^{(k)}), \quad s^{(k+1)} = \sigma^{-1}\Pi_{\mathcal{K}^*}(-z^{(k)}), \tag{14}$$

and $y^{(k+1)}$ from (12a) for $k \geq 0$.

The admissible multiplier step-length range $\tau \in (0, 2)$ is known for two-block ADMM when one objective block is linear; see (Gabay & Mercier, 1976) and, for full-sequence convergence to the set of KKT points, (Chen et al., 2021). Here we also provide a short proof of its averagedness. Let $G_\theta := P_{\mathcal{A}} + \tau^{-1} P_{\mathcal{A}}^{\perp}$. Clearly,

$$\min\{1, \tau^{-1}\} \cdot \mathrm{Id} \preceq_{\mathcal{X}} G_\theta \preceq_{\mathcal{X}} \max\{1, \tau^{-1}\} \cdot \mathrm{Id}. \tag{15}$$

Abbreviate $P_{\mathcal{A}}[\Pi_{\mathcal{K}}(z) - \mathcal{A}^\dagger b]$ as $r_\theta^{\parallel}(z)$ and $P_{\mathcal{A}}^{\perp}[-\Pi_{\mathcal{K}^*}(-z) + \sigma c]$ as $r_\theta^{\perp}(z)$, respectively. Then for ADMM, $T_\theta(z) = z - r_\theta^{\parallel}(z) - \tau r_\theta^{\perp}(z)$ from (13). For $\tau = 1$, it is known from Douglas–Rachford splitting that T_θ is firmly nonexpansive (*i.e.,* 1/2-averaged) under $\langle \cdot, \cdot \rangle_{\mathcal{X}}$ (Ryu & Yin, 2022), and thus $-\delta_\theta$ is 1-cocoercive: $\forall z, w \in \mathcal{X}$,

$$\langle r_\theta^{\parallel}(z) + r_\theta^{\perp}(z) - r_\theta^{\parallel}(w) - r_\theta^{\perp}(w), z - w \rangle_{\mathcal{X}} \geq \|r_\theta^{\parallel}(z) + r_\theta^{\perp}(z) - r_\theta^{\parallel}(w) - r_\theta^{\perp}(w)\|_{\mathcal{X}}^2. \tag{16}$$

For $\tau \neq 1$, fix $z, w \in \mathcal{X}$ and abbreviate $u := r_\theta^{\parallel}(z) - r_\theta^{\parallel}(w) \in \mathrm{ran}(\mathcal{A}^*)$, $v := r_\theta^{\perp}(z) - r_\theta^{\perp}(w) \in \ker(\mathcal{A})$, so that $\delta_\theta(z) - \delta_\theta(w) = -(u + \tau v)$. Since G_θ acts as the identity on $\mathrm{ran}(\mathcal{A}^*)$ and as $\tau^{-1}\mathrm{Id}$ on $\ker(\mathcal{A})$, we have the key identity $G_\theta(u + \tau v) = u + v$. Therefore,

$$\begin{aligned}
&-\langle \delta_\theta(z) - \delta_\theta(w), z - w \rangle_\theta = \langle G_\theta(u + \tau v), z - w \rangle_{\mathcal{X}} = \langle u + v, z - w \rangle_{\mathcal{X}} \\
&\geq \|u + v\|_{\mathcal{X}}^2 = \|G_\theta(\delta_\theta(z) - \delta_\theta(w))\|_{\mathcal{X}}^2 \\
&\geq \min\{1, \tau^{-1}\} \langle G_\theta(\delta_\theta(z) - \delta_\theta(w)), \delta_\theta(z) - \delta_\theta(w) \rangle_{\mathcal{X}} = \min\{1, \tau^{-1}\} \|\delta_\theta(z) - \delta_\theta(w)\|_\theta^2,
\end{aligned}$$

where the first inequality is (16) and the second holds because $G_\theta^2 \succeq_{\mathcal{X}} \min\{1,\tau^{-1}\}\cdot G_\theta$, which follows by comparing $\|G_\theta w\|_{\mathcal{X}}^2 = \|P_{\mathcal{A}} w\|_{\mathcal{X}}^2 + \tau^{-2}\|P_{\mathcal{A}}^{\perp} w\|_{\mathcal{X}}^2$ with $\langle G_\theta w, w\rangle_{\mathcal{X}} = \|P_{\mathcal{A}} w\|_{\mathcal{X}}^2 + \tau^{-1}\|P_{\mathcal{A}}^{\perp} w\|_{\mathcal{X}}^2$. Thus, $-\delta_\theta$ is $\min\{1,\tau^{-1}\}$-cocoercive in $\langle\cdot,\cdot\rangle_\theta$. Equivalently, T_θ is α_θ-averaged with $\alpha_\theta = 0.5\max\{1,\tau\}$; note that $\alpha_\theta \in (0,1)$ precisely because $\tau \in (0,2)$, which is exactly the admissible multiplier range.

We now verify Assumption 2.1 (iv). Since U is compact, $M_U^{\mathcal{A}} := \sup_{\theta\in U}\|\mathcal{A}\|_{\mathrm{op}}$, $M_U^b := \sup_{\theta\in U}\|b\|_{\mathcal{Y}}$ and $M_U^c := \sup_{\theta\in U}\|c\|_{\mathcal{X}}$ are finite. Write $\theta = (\mathcal{A}, b, c)$, $\theta' = (\mathcal{A}', b', c')$ in U, and note $\|\mathcal{A} - \mathcal{A}'\|_{\mathrm{op}} \le \|\mathcal{A} - \mathcal{A}'\|_{\mathrm{HS}} \le \|\theta - \theta'\|_{\mathcal{D}}$.

The condition $\mathcal{A}\mathcal{A}^* \succeq_{\mathcal{Y}} \kappa_U \cdot \mathrm{Id}$ makes $\mathcal{A}\mathcal{A}^*$ uniformly invertible on U, so the resolvent identity $X^{-1} - Y^{-1} = X^{-1}(Y - X)Y^{-1}$ gives $\|(\mathcal{A}\mathcal{A}^*)^{-1} - (\mathcal{A}'\mathcal{A}'^*)^{-1}\|_{\mathrm{op}} \le \kappa_U^{-2}\|\mathcal{A}\mathcal{A}^* - \mathcal{A}'\mathcal{A}'^*\|_{\mathrm{op}} \le 2M_U^{\mathcal{A}}\kappa_U^{-2}\|\mathcal{A} - \mathcal{A}'\|_{\mathrm{op}}$. Since $\mathcal{A}^\dagger = \mathcal{A}^*(\mathcal{A}\mathcal{A}^*)^{-1}$ and $P_{\mathcal{A}} = \mathcal{A}^\dagger\mathcal{A}$ are products of maps that are bounded and Lipschitz in $\mathcal{A}$ over U, there is a finite c_U, depending only on κ_U and $M_U^{\mathcal{A}}$, with

$$\|\mathcal{A}^\dagger - \mathcal{A}'^\dagger\|_{\mathrm{op}} \le c_U\|\mathcal{A} - \mathcal{A}'\|_{\mathrm{op}}, \qquad \|P_{\mathcal{A}} - P_{\mathcal{A}'}\|_{\mathrm{op}} = \|P_{\mathcal{A}}^{\perp} - P_{\mathcal{A}'}^{\perp}\|_{\mathrm{op}} \le c_U\|\mathcal{A} - \mathcal{A}'\|_{\mathrm{op}}; \quad (17)$$

moreover $\|\mathcal{A}^\dagger\|_{\mathrm{op}} \le \kappa_U^{-1/2}$ and $\|P_{\mathcal{A}}\|_{\mathrm{op}}, \|P_{\mathcal{A}}^{\perp}\|_{\mathrm{op}} \le 1$ on U.

Expanding (13) as $T_\theta(z) = z - P_{\mathcal{A}}\Pi_{\mathcal{K}}(z) + P_{\mathcal{A}}\mathcal{A}^\dagger b + \tau P_{\mathcal{A}}^{\perp}\Pi_{\mathcal{K}^*}(-z) - \tau\sigma P_{\mathcal{A}}^{\perp} c$, the identity terms cancel in the difference:

$$\begin{aligned} T_\theta(z) - T_{\theta'}(z) = &-(P_{\mathcal{A}} - P_{\mathcal{A}'})\Pi_{\mathcal{K}}(z) + \tau(P_{\mathcal{A}}^{\perp} - P_{\mathcal{A}'}^{\perp})\Pi_{\mathcal{K}^*}(-z) \\ &+ (P_{\mathcal{A}}\mathcal{A}^\dagger b - P_{\mathcal{A}'}\mathcal{A}'^\dagger b') - \tau\sigma(P_{\mathcal{A}}^{\perp} c - P_{\mathcal{A}'}^{\perp} c'). \end{aligned}$$

Because $\mathcal{K}, \mathcal{K}^*$ are closed convex cones, $\Pi_{\mathcal{K}}(0) = \Pi_{\mathcal{K}^*}(0) = 0$, so 1-Lipschitzness gives $\|\Pi_{\mathcal{K}}(z)\|_{\mathcal{X}} \le \|z\|_{\mathcal{X}}$ and $\|\Pi_{\mathcal{K}^*}(-z)\|_{\mathcal{X}} \le \|z\|_{\mathcal{X}}$. With (17), the first line is thus at most $(1+\tau)c_U\|z\|_{\mathcal{X}}\|\theta - \theta'\|_{\mathcal{D}}$. The second line does not involve z; splitting each product as $\|P_{\mathcal{A}} - P_{\mathcal{A}'}\|_{\mathrm{op}}\|\mathcal{A}^\dagger b\|_{\mathcal{X}} + \|P_{\mathcal{A}'}\|_{\mathrm{op}}\|\mathcal{A}^\dagger b - \mathcal{A}'^\dagger b'\|_{\mathcal{X}}$ and using $\|\mathcal{A}^\dagger b\|_{\mathcal{X}} \le \kappa_U^{-1/2} M_U^b$, it is at most $\kappa_U'\|\theta - \theta'\|_{\mathcal{D}}$ for a finite κ_U' depending only on $\kappa_U, M_U^{\mathcal{A}}, M_U^b, M_U^c, \sigma, \tau$. Hence (3) holds with a modulus of the form

$$L_U(r) = \kappa_U' + (1+\tau)c_U \cdot r, \quad (18)$$

which lies in Aff_+.

The linear growth in $\|z\|_{\mathcal{X}}$ cannot be removed. Suppose $\mathcal{K}$ spans $\mathcal{X}$, as it does for $\mathbb{R}_+^n$, $\mathcal{Q}^n$ and $\mathbb{S}_+^n$, and take $b = b'$, $c = c'$ and $z = td$ with $d \in \mathcal{K}$, $t > 0$. Then $\Pi_{\mathcal{K}}(z) = td$, $\Pi_{\mathcal{K}^*}(-z) = 0$, so

$$T_\theta(z) - T_{\theta'}(z) = -t(P_{\mathcal{A}} - P_{\mathcal{A}'})d + w, \qquad w := P_{\mathcal{A}}\mathcal{A}^\dagger b - P_{\mathcal{A}'}\mathcal{A}'^\dagger b + \tau\sigma(P_{\mathcal{A}} - P_{\mathcal{A}'})c, \quad (19)$$

where w does *not* vanish merely because $b = b'$ and $c = c'$, since $\mathcal{A}^\dagger \ne \mathcal{A}'^\dagger$ in general; it is, however, independent of t. If $P_{\mathcal{A}} \ne P_{\mathcal{A}'}$ then, $\mathcal{K}$ spanning $\mathcal{X}$, some $d \in \mathcal{K}$ has $(P_{\mathcal{A}} - P_{\mathcal{A}'})d \ne 0$, and $\|T_\theta(z) - T_{\theta'}(z)\|_{\mathcal{X}} \ge t\|(P_{\mathcal{A}} - P_{\mathcal{A}'})d\|_{\mathcal{X}} - \|w\|_{\mathcal{X}} \to \infty$. A uniform Lipschitz constant on $U \times \mathcal{X}$ is therefore impossible.

sGS-ADMM case. Fix $\sigma > 0$ and $\tau \in (0,2)$. Suppose $\mathcal{A}\mathcal{A}^* \succeq_{\mathcal{Y}} \kappa_U \cdot \mathrm{Id}$ with constant $\kappa_U > 0$. Consider sGS-ADMM for conic-LP (Chen et al., 2017):

$$y^{(k+1/2)} = (\mathcal{A}\mathcal{A}^*)^{-1}[\sigma^{-1}b - \mathcal{A}(\sigma^{-1}x^{(k)} + s^{(k)} - c)], \quad (20a)$$

$$s^{(k+1)} = \Pi_{\mathcal{K}^*}(c - \mathcal{A}^* y^{(k+1/2)} - \sigma^{-1}x^{(k)}), \quad (20b)$$

$$y^{(k+1)} = (\mathcal{A}\mathcal{A}^*)^{-1}[\sigma^{-1}b - \mathcal{A}(\sigma^{-1}x^{(k)} + s^{(k+1)} - c)], \quad (20c)$$

$$x^{(k+1)} = x^{(k)} + \tau\sigma(\mathcal{A}^* y^{(k+1)} + s^{(k+1)} - c). \quad (20d)$$

Compared to ADMM, sGS-ADMM adds one additional y-update in (20c). Similar to the ADMM case, define the auxiliary state variable $z^{(k)} := P_{\mathcal{A}}^{\perp} x^{(k)} + \mathcal{A}^\dagger b - \sigma P_{\mathcal{A}} s^{(k)} - \sigma P_{\mathcal{A}}^{\perp} c$. From (20a) and (20b), $s^{(k+1)} = \sigma^{-1}\Pi_{\mathcal{K}^*}(-z^{(k)})$. From (20d), we get $x^{(k+1)} = (\mathrm{Id} - \tau P_{\mathcal{A}})x^{(k)} + \tau\sigma P_{\mathcal{A}}^{\perp} s^{(k+1)} + \tau\mathcal{A}^\dagger b -$

$\tau\sigma P_{\mathcal{A}}^{\perp}c$. With the observation that $P_{\mathcal{A}}^{\perp}x^{(k)} = P_{\mathcal{A}}^{\perp}z^{(k)} + \sigma P_{\mathcal{A}}^{\perp}c$, we calculate $z^{(k+1)}$:

$$\begin{aligned}
&z^{(k+1)} = P_{\mathcal{A}}^{\perp}x^{(k+1)} + \mathcal{A}^{\dagger}b - \sigma P_{\mathcal{A}}s^{(k+1)} - \sigma P_{\mathcal{A}}^{\perp}c\\
=&P_{\mathcal{A}}^{\perp}x^{(k)} + \sigma(\tau P_{\mathcal{A}}^{\perp} - P_{\mathcal{A}})s^{(k+1)} + \mathcal{A}^{\dagger}b - (1+\tau)\sigma P_{\mathcal{A}}^{\perp}c\\
=&P_{\mathcal{A}}^{\perp}z^{(k)} + (\tau P_{\mathcal{A}}^{\perp} - P_{\mathcal{A}})\Pi_{\mathcal{K}^*}(-z^{(k)}) + \mathcal{A}^{\dagger}b - \sigma\tau P_{\mathcal{A}}^{\perp}c\\
=&z^{(k)} - P_{\mathcal{A}}[\Pi_{\mathcal{K}}(z^{(k)}) - \Pi_{\mathcal{K}^*}(-z^{(k)})] - P_{\mathcal{A}}\Pi_{\mathcal{K}^*}(-z^{(k)}) + \mathcal{A}^{\dagger}b + \tau P_{\mathcal{A}}^{\perp}\Pi_{\mathcal{K}^*}(-z^{(k)}) - \sigma\tau P_{\mathcal{A}}^{\perp}c\\
=&z^{(k)} - P_{\mathcal{A}}[\Pi_{\mathcal{K}}(z^{(k)}) - \mathcal{A}^{\dagger}b] - \tau P_{\mathcal{A}}^{\perp}[-\Pi_{\mathcal{K}^*}(-z^{(k)}) + \sigma c]. \qquad (21)
\end{aligned}$$

The form of (21) is exactly the same as that of (13). Therefore, in its z-variable formula, sGS-ADMM inherits all dynamical properties from ADMM, including $T_\theta, m_U, M_U, \alpha_\theta, G_\theta$, and in particular Assumption 2.1 (iv) with the same modulus (18).

What differentiates sGS-ADMM from ADMM is its variable recovery strategy from z to (x, y, s). For sGS-ADMM:

$$x^{(k+1)} = (\mathrm{Id} - \tau P_{\mathcal{A}})x^{(k)} + \tau P_{\mathcal{A}}^{\perp}(\Pi_{\mathcal{K}^*}(-z^{(k)}) - \sigma c) + \tau\mathcal{A}^{\dagger}b,\ s^{(k+1)} = \sigma^{-1}\Pi_{\mathcal{K}^*}(-z^{(k)}), \quad (22)$$

and $y^{(k+1)}$ can be recovered from (20a)–(20c). Except for $s^{(k+1)}$'s recovery, these rules are different from ADMM's in (14). Consequently, the observable KKT residuals are also different.

PDHG case. Fix $\eta_x, \eta_y > 0$. We further assume $\gamma_U := \sup_{\theta\in U}\sqrt{\eta_x\eta_y}\|\mathcal{A}\|_{\mathrm{op}} < 1$. Let $\mathcal{H} := \mathcal{X}\times\mathcal{Y}$ and the auxiliary state variable $z := (x, y) \in \mathcal{H}$. PDHG's update for conic-LP is (Chambolle & Pock, 2011):

$$x^{(k+1)} = x_\theta^+(z^{(k)}) := \Pi_{\mathcal{K}}[x^{(k)} - \eta_x(c - \mathcal{A}^*y^{(k)})], \qquad (23a)$$

$$y^{(k+1)} = y_\theta^+(z^{(k)}) := y^{(k)} + \eta_y[b - \mathcal{A}(2x^{(k+1)} - x^{(k)})]. \qquad (23b)$$

Let $G_\theta := \begin{bmatrix}\eta_x^{-1}\cdot\mathrm{Id}_{\mathcal{X}} & \mathcal{A}^*\\ \mathcal{A} & \eta_y^{-1}\cdot\mathrm{Id}_{\mathcal{Y}}\end{bmatrix}$. Since $\gamma_U < 1$, $G_\theta \succ_{\mathcal{H}} 0$. Denote $T_\theta(z) := (x_\theta^+(z), y_\theta^+(z))$. From (Chan et al., 2014), T_θ is firmly nonexpansive (*i.e.,* 1/2-averaged) under G_θ-metric.

For Assumption 2.1 (iv), again put $M_U^{\mathcal{A}} := \sup_{\theta\in U}\|\mathcal{A}\|_{\mathrm{op}}$, $M_U^c := \sup_{\theta\in U}\|c\|_{\mathcal{X}}$, both finite by compactness of U, and recall $\|\mathcal{A} - \mathcal{A}'\|_{\mathrm{op}} \le \|\theta - \theta'\|_{\mathcal{D}}$. Fix $z = (x, y) \in \mathcal{H}$ and $\theta, \theta' \in U$. Since $\Pi_{\mathcal{K}}$ is 1-Lipschitz, the x-block obeys

$$\|x_\theta^+(z) - x_{\theta'}^+(z)\|_{\mathcal{X}} \le \eta_x\|c - c'\|_{\mathcal{X}} + \eta_x\|\mathcal{A} - \mathcal{A}'\|_{\mathrm{op}}\|y\|_{\mathcal{Y}} \le \eta_x(1 + \|z\|_{\mathcal{H}})\|\theta - \theta'\|_{\mathcal{D}}. \quad (24)$$

Moreover $\Pi_{\mathcal{K}}(0) = 0$ gives the uniform bound $\|x_{\theta'}^+(z)\|_{\mathcal{X}} \le \|x\|_{\mathcal{X}} + \eta_x(\|c'\|_{\mathcal{X}} + \|\mathcal{A}'\|_{\mathrm{op}}\|y\|_{\mathcal{Y}}) \le (1 + \eta_x M_U^{\mathcal{A}})\|z\|_{\mathcal{H}} + \eta_x M_U^c$, hence $\|2x_{\theta'}^+(z) - x\|_{\mathcal{X}} \le (3 + 2\eta_x M_U^{\mathcal{A}})\|z\|_{\mathcal{H}} + 2\eta_x M_U^c$. For the y-block, inserting $\mathcal{A}(2x_{\theta'}^+(z) - x)$ and using (24),

$$\begin{aligned}
\|y_\theta^+(z) - y_{\theta'}^+(z)\|_{\mathcal{Y}} &\le \eta_y\|b - b'\|_{\mathcal{Y}} + 2\eta_y\|\mathcal{A}\|_{\mathrm{op}}\|x_\theta^+(z) - x_{\theta'}^+(z)\|_{\mathcal{X}}\\
&\quad + \eta_y\|\mathcal{A} - \mathcal{A}'\|_{\mathrm{op}}\|2x_{\theta'}^+(z) - x\|_{\mathcal{X}}\\
&\le \eta_y[1 + 2\eta_x M_U^{\mathcal{A}}(1 + \|z\|_{\mathcal{H}}) + (3 + 2\eta_x M_U^{\mathcal{A}})\|z\|_{\mathcal{H}} + 2\eta_x M_U^c]\cdot\|\theta - \theta'\|_{\mathcal{D}}.
\end{aligned}$$

Adding the two blocks yields (3) with the concrete modulus

$$L_U(r) = [\eta_x + \eta_y(1 + 2\eta_x M_U^{\mathcal{A}} + 2\eta_x M_U^c)] + [\eta_x + \eta_y(3 + 4\eta_x M_U^{\mathcal{A}})]\cdot r, \qquad (25)$$

again an element of Aff_+. As with ADMM, the factor $\|z\|_{\mathcal{H}}$ is unavoidable, though here the argument needs one more step: the *inputs* to $\Pi_{\mathcal{K}}$ separate by $\eta_x\|(\mathcal{A} - \mathcal{A}')^*y\|_{\mathcal{X}} \to \infty$, but $\Pi_{\mathcal{K}}$ is 1-Lipschitz and could in principle absorb that growth. We now show that it does not. Take $\mathcal{K} = \mathbb{R}_+^n$, $b = b'$, $c = c'$, $x = 0$ and $z = (0, ty)$; then $x^+ - x'^+ = \Pi_{\mathbb{R}_+^n}[t\eta_x\mathcal{A}^*y - \eta_x c] - \Pi_{\mathbb{R}_+^n}[t\eta_x\mathcal{A}'^*y - \eta_x c]$, which for large t behaves like $t\eta_x[(\mathcal{A}^*y)_+ - (\mathcal{A}'^*y)_+]$. Were $(\mathcal{A}^*y)_+ = (\mathcal{A}'^*y)_+$ for every $y \in \mathcal{Y}$, then applying this to both y and $-y$ coordinatewise would force $\mathcal{A}^*y = \mathcal{A}'^*y$ for all y, *i.e.,* $\mathcal{A} = \mathcal{A}'$. Hence whenever $\mathcal{A} \ne \mathcal{A}'$ some y makes the x-block grow linearly in t.

It remains to determine m_U and M_U. Take any $z = (u, w) \in \mathcal{X}\times\mathcal{Y}$, $\langle G_\theta z, z\rangle_{\mathcal{H}} = \eta_x^{-1}\|u\|_{\mathcal{X}}^2 + \eta_y^{-1}\|w\|_{\mathcal{Y}}^2 + 2\langle\mathcal{A}u, w\rangle_{\mathcal{Y}}$. By Cauchy–Schwarz and Young's inequality,

$$2\,|\langle\mathcal{A}u, w\rangle_{\mathcal{Y}}| \le \sqrt{\eta_x\eta_y}\|\mathcal{A}\|_{\mathrm{op}}(\eta_x^{-1}\|u\|_{\mathcal{X}}^2 + \eta_y^{-1}\|w\|_{\mathcal{Y}}^2) \le \gamma_U(\eta_x^{-1}\|u\|_{\mathcal{X}}^2 + \eta_y^{-1}\|w\|_{\mathcal{Y}}^2).$$

Since $\min\{\eta_x^{-1}, \eta_y^{-1}\}\|z\|_{\mathcal{H}}^2 \le \eta_x^{-1}\|u\|_{\mathcal{X}}^2 + \eta_y^{-1}\|w\|_{\mathcal{Y}}^2 \le \max\{\eta_x^{-1}, \eta_y^{-1}\}\|z\|_{\mathcal{H}}^2$, we can take $m_U := (1 - \gamma_U)/\max\{\eta_x, \eta_y\}$ and $M_U := (1 + \gamma_U)/\min\{\eta_x, \eta_y\}$.

Proof of Theorem 2.2. Write $\theta := (\mathcal{A}, b, c)$, $\theta' := (\mathcal{A}, b', c')$. For ADMM/sGS-ADMM, Table 1 gives $\delta_\theta z - \delta_{\theta'} z = \mathcal{A}^\dagger(b - b') - \tau\sigma P_{\mathcal{A}}^{\perp}(c - c')$, which is (6). For PDHG, let $\Delta x^+ := \Pi_{\mathcal{K}}(x + \eta_x \mathcal{A}^* y - \eta_x c) - \Pi_{\mathcal{K}}(x + \eta_x \mathcal{A}^* y - \eta_x c')$, so that $\delta_\theta z - \delta_{\theta'} z = (\Delta x^+,\ \eta_y(b - b') - 2\eta_y \mathcal{A}\Delta x^+)$. Since $\Pi_{\mathcal{K}}$ is 1-Lipschitz, $\|\Delta x^+\|_{\mathcal{X}} \le \eta_x \|c - c'\|_{\mathcal{X}}$, giving (5); if $c = c'$ then $\Delta x^+ = 0$ and $\delta_\theta z - \delta_{\theta'} z = (0, \eta_y(b - b'))$, giving (6). □

Fixed points correspond to KKT points. It remains to verify Assumption 2.1 (i). The equivalence $\mathrm{Fix}(T_\theta) \neq \emptyset$ iff $\mathrm{KKT}(\theta) \neq \emptyset$ already follows from the classical fixed-point characterizations, of Douglas–Rachford splitting (Bauschke & Combettes, 2017, Prop. 26.1) and of the PDHG saddle point (Chambolle & Pock, 2011). Beyond that, we need the explicit correspondence in the z-coordinates of Table 1.

Lemma A.1 (Fixed points correspond to KKT points)**.** *Let $\theta = (\mathcal{A}, b, c) \in U$. For ADMM and sGS-ADMM under the conditions of Table 1, the map $\Phi : (x, y, s) \mapsto x - \sigma s$ is a bijection from $\mathrm{KKT}(\theta)$ onto $\mathrm{Fix}(T_\theta)$ (the rank condition of Table 1 makes y unique given (x, s)); for PDHG, $(x, y) \in \mathrm{Fix}(T_\theta)$ iff $(x, y, c - \mathcal{A}^* y) \in \mathrm{KKT}(\theta)$. In both cases $\mathrm{Fix}(T_\theta) \neq \emptyset$ iff $\mathrm{KKT}(\theta) \neq \emptyset$.*

Proof. ADMM and sGS-ADMM. By Table 1 the two share the z-map, so one computation suffices. Put $X := \Pi_{\mathcal{K}}(z)$ and $\sigma S := \Pi_{\mathcal{K}^*}(-z)$; Moreau's decomposition makes $z \mapsto (X, S)$ a bijection onto the pairs with $X \in \mathcal{K}$, $S \in \mathcal{K}^*$, $\langle X, S\rangle_{\mathcal{X}} = 0$, with inverse $z = X - \sigma S$. In these terms Table 1 reads $\delta_\theta(z) = -P_{\mathcal{A}}[X - \mathcal{A}^\dagger b] - \tau\sigma P_{\mathcal{A}}^{\perp}[c - S]$, whose two terms lie in the orthogonal subspaces $\mathrm{ran}(\mathcal{A}^*)$ and $\mathrm{ran}(\mathcal{A}^*)^\perp$; as $\tau\sigma \neq 0$, $\delta_\theta(z) = 0$ iff both vanish. The first says $\mathcal{A}X = b$, since $\mathcal{A}\mathcal{A}^\dagger = \mathrm{Id}_{\mathcal{Y}}$ makes it $\mathcal{A}^\dagger(\mathcal{A}X - b) = 0$ with $\mathcal{A}^\dagger$ injective; the second says $c - S \in \mathrm{ran}(\mathcal{A}^*)$, *i.e.,* $\mathcal{A}^* y + S = c$ for some y. These are exactly (2), and y is unique because $\ker(\mathcal{A}^*) = \{0\}$, which gives the stated bijection.

PDHG. Here $z = (x, y)$ and $T_\theta z = (x^+, y + \eta_y[b - \mathcal{A}(2x^+ - x)])$. If z is fixed then $x^+ = x$, and the y-block gives $\mathcal{A}x = b$. With $s := c - \mathcal{A}^* y$, the x-block reads $x = \Pi_{\mathcal{K}}[x - \eta_x s]$, which for a closed convex cone $\mathcal{K}$ and $\eta_x > 0$ holds iff $-\eta_x s \in N_{\mathcal{K}}(x)$, *i.e.,* iff $x \in \mathcal{K}$, $s \in \mathcal{K}^*$ and $\langle x, s\rangle_{\mathcal{X}} = 0$. Thus $(x, y, s) \in \mathrm{KKT}(\theta)$. The converse follows by reversing each step. □

From the z-residual to KKT residuals. The theory measures $\|\delta_\theta(z)\|_{\mathcal{H}}$, while solvers monitor the KKT residuals of (x, y, s). The next lemma connects the two for the maps of Table 1.

Lemma A.2 (KKT-residual bridge)**.** *Let $\theta = (\mathcal{A}, b, c) \in U$. For ADMM/sGS-ADMM under the conditions of Table 1, take any $z \in \mathcal{X}$, put $X := \Pi_{\mathcal{K}}(z)$ and $\sigma S := \Pi_{\mathcal{K}^*}(-z)$, so that $X \in \mathcal{K}$, $S \in \mathcal{K}^*$, $\langle X, S\rangle_{\mathcal{X}} = 0$, and define the projected KKT residual $r_\theta(z) := \|\mathcal{A}X - b\|_{\mathcal{Y}} + \min_{y \in \mathcal{Y}} \|\mathcal{A}^* y + S - c\|_{\mathcal{X}}$. Then*

$$\frac{1}{2}\min\{1/M_U^{\mathcal{A}},\ \tau\sigma\} \cdot r_\theta(z) \ \le\ \|\delta_\theta(z)\|_{\mathcal{X}} \ \le\ \max\{\kappa_U^{-1/2},\ \tau\sigma\} \cdot r_\theta(z). \tag{26}$$

For PDHG, write $z = (x, y)$ and $\delta_\theta(z) = (\delta_x, \delta_y)$, and put $x^+ := x + \delta_x$ and $\tilde{s} := c - \mathcal{A}^* y + \eta_x^{-1}\delta_x$, so that $x^+ \in \mathcal{K}$, $\tilde{s} \in \mathcal{K}^*$, $\langle x^+, \tilde{s}\rangle_{\mathcal{X}} = 0$. Then, with $r_\theta(z) := \|\mathcal{A}x^+ - b\|_{\mathcal{Y}} + \|\mathcal{A}^* y + \tilde{s} - c\|_{\mathcal{X}}$,

$$(\eta_x^{-1} + \eta_y^{-1} + M_U^{\mathcal{A}})^{-1} \cdot r_\theta(z) \ \le\ \|\delta_\theta(z)\|_{\mathcal{H}} \ \le\ (\eta_x + \eta_y + \eta_x\eta_y M_U^{\mathcal{A}}) \cdot r_\theta(z). \tag{27}$$

Proof. For ADMM/sGS-ADMM, the proof of Lemma A.1 gives $\delta_\theta(z) = -P_{\mathcal{A}}[X - \mathcal{A}^\dagger b] - \tau\sigma P_{\mathcal{A}}^{\perp}[c - S]$, with $P_{\mathcal{A}}[X - \mathcal{A}^\dagger b] = \mathcal{A}^\dagger(\mathcal{A}X - b)$ since $P_{\mathcal{A}}\mathcal{A}^\dagger = \mathcal{A}^\dagger$, and the two terms are orthogonal. Since $\mathcal{A}\mathcal{A}^\dagger = \mathrm{Id}_{\mathcal{Y}}$ and $\|\mathcal{A}^\dagger\|_{\mathrm{op}} \le \kappa_U^{-1/2}$, we get $\|\mathcal{A}X - b\|_{\mathcal{Y}}/M_U^{\mathcal{A}} \le \|\mathcal{A}^\dagger(\mathcal{A}X - b)\|_{\mathcal{X}} \le \kappa_U^{-1/2}\|\mathcal{A}X - b\|_{\mathcal{Y}}$, while $\|P_{\mathcal{A}}^{\perp}(c - S)\|_{\mathcal{X}}$ is the distance from $c - S$ to $\mathrm{ran}(\mathcal{A}^*)$, namely $\min_{y \in \mathcal{Y}} \|\mathcal{A}^* y + S - c\|_{\mathcal{X}}$. The upper bound in (26) is the triangle inequality. For the lower bound, orthogonality gives $\|\delta_\theta(z)\|_{\mathcal{X}} \ge \max\{\|\mathcal{A}^\dagger(\mathcal{A}X - b)\|_{\mathcal{X}},\ \tau\sigma\|P_{\mathcal{A}}^{\perp}(c - S)\|_{\mathcal{X}}\}$, and the maximum of two nonnegative numbers is at least half their sum. For PDHG, $x^+ = \Pi_{\mathcal{K}}(w)$ with $w := x - \eta_x(c - \mathcal{A}^* y)$, and Moreau's decomposition of w gives $x^+ \in \mathcal{K}$, $w - x^+ \in -\mathcal{K}^*$ and $\langle x^+, w - x^+\rangle_{\mathcal{X}} = 0$; since $w - x^+ = -\eta_x\tilde{s}$, the pair $(x^+, \tilde{s})$ is exactly cone-feasible and complementary. Its dual residual is $\mathcal{A}^* y + \tilde{s} - c = \eta_x^{-1}\delta_x$, while the y-block $\delta_y = \eta_y[b - \mathcal{A}(2x^+ - x)]$ gives $b - \mathcal{A}x^+ = \eta_y^{-1}\delta_y + \mathcal{A}\delta_x$. Both sides of (27) follow from the triangle inequality applied to these two identities, using $\|\delta_x\|_{\mathcal{X}}, \|\delta_y\|_{\mathcal{Y}} \le \|\delta_\theta(z)\|_{\mathcal{H}} \le \|\delta_x\|_{\mathcal{X}} + \|\delta_y\|_{\mathcal{Y}}$. □

In words, for ADMM/sGS-ADMM the z-residual *is* the projected KKT residual of the recovered pair (X, S), up to constants depending only on U, σ, and τ: the pair is exactly cone-feasible and complementary, its primal infeasibility is the $\mathrm{ran}(\mathcal{A}^*)$-component of $\delta_\theta(z)$, and its minimal dual infeasibility is the orthogonal component. For PDHG the same reading holds at the half-updated pair $(x^+, \tilde{s})$, which one projection computes from the state.

A.3 Additional material for §2.2

Lemma A.3 (Fixed point sets are closed and convex)**.** *Under Assumption 2.1,* $\mathrm{Fix}(T_\theta)$ *is closed and convex for every* $\theta \in U$.

Proof. T_θ is nonexpansive in the Hilbertian metric $\langle\cdot,\cdot\rangle_\theta$, and the fixed point set of a nonexpansive map on a Hilbert space is closed and convex (Bauschke & Combettes, 2017, Corollary 4.24). □

Proof of Proposition 2.4. Fix $\varepsilon > 0$ and $z \in \mathcal{S}_\varepsilon$. Since $T_\varepsilon^{\mathrm{c}}$ is $\alpha_{\theta_\varepsilon^{\mathrm{c}}}$-averaged with respect to $\langle\cdot,\cdot\rangle_{\theta_\varepsilon^{\mathrm{c}}}$, it is nonexpansive in that dynamic metric. Thus, $\forall k \geq 1$, $\|(T_\varepsilon^{\mathrm{c}})^{k+1}(z) - (T_\varepsilon^{\mathrm{c}})^k(z)\|_{\theta_\varepsilon^{\mathrm{c}}} \leq \|T_\varepsilon^{\mathrm{c}}(z) - z\|_{\theta_\varepsilon^{\mathrm{c}}} = \|\delta_\varepsilon^{\mathrm{c}}(z)\|_{\theta_\varepsilon^{\mathrm{c}}}$. On the other hand, $\left|\mathrm{dist}_{\theta_\varepsilon^{\mathrm{c}}}((T_\varepsilon^{\mathrm{c}})^k(z), V_\varepsilon^{\mathrm{c}}) - \mathrm{dist}_{\theta_\varepsilon^{\mathrm{c}}}(z, V_\varepsilon^{\mathrm{c}})\right| \leq \|(T_\varepsilon^{\mathrm{c}})^k(z) - z\|_{\theta_\varepsilon^{\mathrm{c}}}$ by 1-Lipschitzness of the distance function. Thus, $\forall k \geq 1$,

$$\begin{aligned}&\mathrm{dist}_{\theta_\varepsilon^{\mathrm{c}}}((T_\varepsilon^{\mathrm{c}})^k(z), V_\varepsilon^{\mathrm{c}}) \geq \mathrm{dist}_{\theta_\varepsilon^{\mathrm{c}}}(z, V_\varepsilon^{\mathrm{c}}) - \|(T_\varepsilon^{\mathrm{c}})^k(z) - z\|_{\theta_\varepsilon^{\mathrm{c}}}\\ \geq&\mathrm{dist}_{\theta_\varepsilon^{\mathrm{c}}}(z, V_\varepsilon^{\mathrm{c}}) - \sum_{i=0}^{k-1}\|(T_\varepsilon^{\mathrm{c}})^{i+1}(z) - (T_\varepsilon^{\mathrm{c}})^i(z)\|_{\theta_\varepsilon^{\mathrm{c}}} \geq \mathrm{dist}_{\theta_\varepsilon^{\mathrm{c}}}(z, V_\varepsilon^{\mathrm{c}}) - k\|\delta_\varepsilon^{\mathrm{c}}(z)\|_{\theta_\varepsilon^{\mathrm{c}}}.\end{aligned}$$

From Assumption 2.1, for any $w \in \mathcal{H}$, we get $\sqrt{m_U}\cdot\mathrm{dist}_{\mathcal{H}}(w, V_\varepsilon^{\mathrm{c}}) \leq \mathrm{dist}_{\theta_\varepsilon^{\mathrm{c}}}(w, V_\varepsilon^{\mathrm{c}}) \leq \sqrt{M_U}\cdot\mathrm{dist}_{\mathcal{H}}(w, V_\varepsilon^{\mathrm{c}})$, $\|w\|_{\theta_\varepsilon^{\mathrm{c}}} \leq \sqrt{M_U}\cdot\|w\|_{\mathcal{H}}$. Thus, $\forall k \geq 1$, $\sqrt{M_U}\cdot\mathrm{dist}_{\mathcal{H}}((T_\varepsilon^{\mathrm{c}})^k(z), V_\varepsilon^{\mathrm{c}}) \geq \sqrt{m_U}\cdot\mathrm{dist}_{\mathcal{H}}(z, V_\varepsilon^{\mathrm{c}}) - k\sqrt{M_U}\cdot\|\delta_\varepsilon^{\mathrm{c}}(z)\|_{\mathcal{H}}$. From (8) and $k \leq N_\varepsilon$, we have

$$\mathrm{dist}_{\mathcal{H}}((T_\varepsilon^{\mathrm{c}})^k(z), V_\varepsilon^{\mathrm{c}}) \geq (\sqrt{m_U/M_U} - k\rho_\varepsilon)\cdot\mathrm{dist}_{\mathcal{H}}(z, V_\varepsilon^{\mathrm{c}}) \geq \sqrt{m_U/M_U}\beta\cdot\mathrm{dist}_{\mathcal{H}}(z, V_\varepsilon^{\mathrm{c}})$$

for any $z \in \mathcal{S}_\varepsilon$. □

Proof of Proposition 2.5. Fix $\varepsilon > 0$. The union is nonempty and misses $V_\varepsilon^{\mathrm{c}}$, since each $\mathcal{S}_\varepsilon^i$ does, which is Definition 2.3 (i). Any $z \in \cup_{i\in[r]}\mathcal{S}_\varepsilon^i$ lies in some $\mathcal{S}_\varepsilon^i$, so (8) gives $\|\delta_\varepsilon^{\mathrm{c}}(z)\|_{\mathcal{H}} \leq \rho_\varepsilon^i\cdot\mathrm{dist}_{\mathcal{H}}(z, V_\varepsilon^{\mathrm{c}}) \leq \max_{i\in[r]}\rho_\varepsilon^i\cdot\mathrm{dist}_{\mathcal{H}}(z, V_\varepsilon^{\mathrm{c}})$. Finiteness of the index set gives $\max_{i\in[r]}\rho_\varepsilon^i \to 0$ as $\varepsilon \downarrow 0$. Conversely each $\mathcal{S}_\varepsilon^i \subseteq \cup_{j\in[r]}\mathcal{S}_\varepsilon^j$, so the supremum in (8) over the union is at least ρ_ε^i for every i, and the modulus is exactly $\max_{i\in[r]}\rho_\varepsilon^i$. □

Proposition 2.5 does not extend verbatim to countably many families: $\rho_\varepsilon^i \to 0$ for each fixed i does not force $\sup_i \rho_\varepsilon^i \to 0$, as $\rho_\varepsilon^i := \varepsilon^{1/i}$ shows.

Proof of Proposition 2.6. Since $T_\varepsilon^{\mathrm{c}}$ is nonexpansive in its dynamic metric, for any $z, q \in \mathcal{H}$, we have

$$\begin{aligned}&\|\delta_\varepsilon^{\mathrm{c}}(z) - \delta_\varepsilon^{\mathrm{c}}(q)\|_{\mathcal{H}} \leq 1/\sqrt{m_U}\cdot(\|T_\varepsilon^{\mathrm{c}}(z) - T_\varepsilon^{\mathrm{c}}(q)\|_{\theta_\varepsilon^{\mathrm{c}}} + \|z - q\|_{\theta_\varepsilon^{\mathrm{c}}})\\ \leq&2/\sqrt{m_U}\cdot\|z - q\|_{\theta_\varepsilon^{\mathrm{c}}} \leq 2\sqrt{M_U/m_U}\cdot\|z - q\|_{\mathcal{H}}.\end{aligned}$$

Now take $z \in \tilde{\mathcal{S}}_\varepsilon$ and $q \in \mathcal{S}_\varepsilon$ with $\|z - q\|_{\mathcal{H}} \leq a_\varepsilon\cdot\mathrm{dist}_{\mathcal{H}}(q, V_\varepsilon^{\mathrm{c}})$. By 1-Lipschitzness of the distance function, $\mathrm{dist}_{\mathcal{H}}(z, V_\varepsilon^{\mathrm{c}}) \geq \mathrm{dist}_{\mathcal{H}}(q, V_\varepsilon^{\mathrm{c}}) - \|z - q\|_{\mathcal{H}} \geq (1 - a_\varepsilon)\cdot\mathrm{dist}_{\mathcal{H}}(q, V_\varepsilon^{\mathrm{c}}) > 0$, since $a_\varepsilon < 1$ and $q \notin V_\varepsilon^{\mathrm{c}}$; as $\mathcal{S}_\varepsilon \subseteq \tilde{\mathcal{S}}_\varepsilon$, this settles Definition 2.3 (i). Finally,

$$\begin{aligned}&\|\delta_\varepsilon^{\mathrm{c}}(z)\|_{\mathcal{H}} \leq \|\delta_\varepsilon^{\mathrm{c}}(q)\|_{\mathcal{H}} + 2\sqrt{M_U/m_U}\cdot\|z - q\|_{\mathcal{H}}\\ \leq&(\rho_\varepsilon + 2\sqrt{M_U/m_U}\cdot a_\varepsilon)\cdot\mathrm{dist}_{\mathcal{H}}(q, V_\varepsilon^{\mathrm{c}}) \leq \frac{\rho_\varepsilon + 2\sqrt{M_U/m_U}\cdot a_\varepsilon}{1 - a_\varepsilon}\cdot\mathrm{dist}_{\mathcal{H}}(z, V_\varepsilon^{\mathrm{c}}).\end{aligned}$$

The last inequality uses the displayed lower bound on $\mathrm{dist}_{\mathcal{H}}(z, V_\varepsilon^{\mathrm{c}})$, and $\tilde{\rho}_\varepsilon \to 0$ because $\rho_\varepsilon, a_\varepsilon \to 0$. □

Proof of Proposition 2.7. The first statement is classical, and needs more than nonexpansiveness. Since T_θ is nonexpansive in $\langle\cdot,\cdot\rangle_\theta$, the set $\mathrm{cl}(\mathrm{ran}(\mathrm{Id}-T_\theta))$ is convex (Pazy, 1971), hence so is $\mathrm{cl}(\mathrm{ran}(\delta_\theta)) = -\mathrm{cl}(\mathrm{ran}(\mathrm{Id}-T_\theta))$, so the projection defining v_θ is single-valued and $-v_\theta$ is the *minimal displacement vector* of T_θ. Being α_θ-averaged, T_θ is in addition strongly nonexpansive in that metric (Bauschke & Combettes, 2017), and for strongly nonexpansive maps the successive differences $(\mathrm{Id}-T_\theta)z^{(k)} = -\delta_\theta z^{(k)}$ converge strongly to $-v_\theta$, from any $z^{(0)}$ (Bruck & Reich, 1977; Baillon et al., 1978); plain nonexpansiveness yields only convergence of their norms. Monotonicity is immediate from nonexpansiveness of T_θ in $\|\cdot\|_\theta$: since $\delta_\theta z^{(k+1)} = T_\theta(T_\theta z^{(k)}) - T_\theta z^{(k)}$, we get $\|\delta_\theta z^{(k+1)}\|_\theta \le \|T_\theta z^{(k)} - z^{(k)}\|_\theta = \|\delta_\theta z^{(k)}\|_\theta$; the limit is $\|v_\theta\|_\theta$ by the first statement. Note the ambient residual $\|\delta_\theta z^{(k)}\|_{\mathcal{H}}$ need not be monotone, which is why the dynamic metric is the natural one to monitor. For the second statement, the projection definition gives $\langle v_\theta, v_\theta - w\rangle_\theta \le 0, \forall w \in \mathrm{cl}(\mathrm{ran}(\delta_\theta))$. On the other hand, since T_θ is α_θ-averaged under $\langle\cdot,\cdot\rangle_\theta$, $-\delta_\theta$ is $1/(2\alpha_\theta)$-cocoercive: $\dfrac{1}{2\alpha_\theta}\|\delta_\theta(z) - \delta_\theta(z+tv_\theta)\|_\theta^2 \le \langle \delta_\theta(z) - \delta_\theta(z+tv_\theta), tv_\theta\rangle_\theta$. With $\delta_\theta(z) = v_\theta$ the right side is $t\langle v_\theta - \delta_\theta(z+tv_\theta), v_\theta\rangle_\theta \le 0$, by the projection inequality applied to $w := \delta_\theta(z+tv_\theta) \in \mathrm{ran}(\delta_\theta)$; hence $\|v_\theta - \delta_\theta(z+tv_\theta)\|_\theta = 0$. For the last statement, note $\mathrm{Fix}(T_\theta) = \emptyset$ means $0 \notin \mathrm{ran}(\delta_\theta)$. If $v_\theta \in \mathrm{ran}(\delta_\theta)$, then $v_\theta \neq 0$; picking z with $\delta_\theta(z) = v_\theta$ and setting $z_j := z + jv_\theta$, the previous statement gives $\delta_\theta(z_j) = v_\theta$ while $\|z_j\|_{\mathcal{H}} \to \infty$. Otherwise $v_\theta \in \mathrm{cl}(\mathrm{ran}(\delta_\theta)) \setminus \mathrm{ran}(\delta_\theta)$, so there are w_j with $\delta_\theta(w_j) \to v_\theta$; such $\{w_j\}$ must be unbounded, since a bounded one would admit, by finite dimensionality and continuity of δ_θ, a subsequence converging to some w with $\delta_\theta(w) = v_\theta$. Extracting a subsequence with norms tending to ∞ finishes the proof. □

A.4 Additional material for §3

A.4.1 Motivating example

Here $\bar{\mathcal{A}} = [1\ -1]$, so $\bar{\mathcal{A}}\bar{\mathcal{A}}^* = 2$ and $\|\bar{\mathcal{A}}\|_{\mathrm{op}} = \sqrt{2}$. Table 1 asks for $\mathcal{A}\mathcal{A}^* \succeq_{\mathcal{Y}} \kappa_U \cdot \mathrm{Id}$ for ADMM/sGS-ADMM and $\gamma_U < 1$ for PDHG; at $\bar\theta$ these read $2 > 0$ and $\sqrt{2\eta_x\eta_y} < 1$. Both are strict inequalities between continuous functions of θ, so both persist on a small compact $U \ni \bar\theta$, and Assumption 2.1 holds there. Since $\theta_\varepsilon^{\mathrm{c}}$ and $\bar\theta$ differ only in b, U may be taken with $\mathcal{A}$ fixed, and Theorem 2.2 (i) makes the defect state-free: $\delta_\varepsilon^{\mathrm{c}} - \delta_\varepsilon^{\mathrm{p}} \equiv \bar{\mathcal{A}}^\dagger\varepsilon = (\varepsilon/2,\ -\varepsilon/2)$ for ADMM and $(0, 0, \eta_y\varepsilon)$ for PDHG, of size $e_\varepsilon = \varepsilon/\sqrt{2}$, resp. $\eta_y\varepsilon$, in $\|\cdot\|_{\mathcal{H}}$. As $\theta_\varepsilon^{\mathrm{p}} \equiv \bar\theta$, $\delta_\varepsilon^{\mathrm{p}}$ vanishes on $\mathcal{L}$, so (10) holds with $\mu_\varepsilon \equiv 0$ and (11) needs only e_ε divided by $\inf_{z\in\mathcal{S}} \mathrm{dist}_{\mathcal{H}}(z, V_\varepsilon^{\mathrm{c}})$. That infimum is at least c_0: for ADMM, $z = (-(1-t)\sigma, -t\sigma)$ gives $\|z - (\varepsilon, -\sigma)\|_{\mathcal{H}}^2 = ((1-t)\sigma + \varepsilon)^2 + ((1-t)\sigma)^2 \ge 2((1-t)\sigma)^2 = \mathrm{dist}_{\mathcal{H}}(z, F)^2$, and for PDHG, $z = (0, 0, y)$ gives $\|z - (\varepsilon, 0, 1)\|_{\mathcal{H}}^2 = \varepsilon^2 + (1-y)^2 \ge \mathrm{dist}_{\mathcal{H}}(z, F)^2$, while $\mathrm{dist}_{\mathcal{H}}(z, F) \ge c_0$ on $\mathcal{S}$ by construction. Hence $\rho_\varepsilon \le \varepsilon/(\sqrt{2}c_0)$, resp. $\eta_y\varepsilon/c_0$. Replacing the fixed margin by $c_\varepsilon := c_\star\varepsilon^\alpha$ with $c_\star > 0$ and $\alpha \ge 0$ leaves every step above unchanged, $\mathcal{S}_\varepsilon$ staying nonempty once c_ε falls below the length of $\mathcal{L}$, and gives $\rho_\varepsilon = \Theta(\varepsilon^{1-\alpha})$ while $c_\varepsilon \gg \varepsilon$, that is for $\alpha < 1$; for $\alpha \ge 1$ the $O(\varepsilon)$ displacement of $V_\varepsilon^{\mathrm{c}}$ dominates the margin, the infimum saturates at $\Theta(\varepsilon)$ and $\rho_\varepsilon = \Theta(1)$. Either way $\rho_\varepsilon \to 0$ exactly when $\alpha < 1$; Definition 2.8 constrains only ρ_ε, and not how $\mathcal{S}_\varepsilon$ moves with ε.

Which ε-independent states are slow. For a fixed singleton $\mathcal{S}_\varepsilon \equiv \{z\}$, (8) reads $\rho_\varepsilon = \|\delta_\varepsilon^{\mathrm{c}}(z)\|_{\mathcal{H}}/\mathrm{dist}_{\mathcal{H}}(z, V_\varepsilon^{\mathrm{c}}) =: R_\varepsilon(z)$. Only b moves, so $\delta_\varepsilon^{\mathrm{c}} = \delta_{\bar\theta} + \bar{\mathcal{A}}^\dagger\varepsilon$ everywhere and $V_\varepsilon^{\mathrm{c}} = \{F_\varepsilon\}$ with $F_\varepsilon \to F$; hence $R_\varepsilon(z) \to \|\delta_{\bar\theta}(z)\|_{\mathcal{H}}/\|z - F\|_{\mathcal{H}}$ for every fixed $z \neq F$, which vanishes exactly when $\delta_{\bar\theta}(z) = 0$:

$$\{z\} \text{ is an SR family } \iff z \in \mathcal{L} \setminus \{F\}.$$

At F the two scales coincide, $\varepsilon/\sqrt{2}$ against ε, so $R_\varepsilon(F) \equiv 2^{-1/2}$ (η_y for PDHG). The conclusion is uniform off $\mathcal{L}$: on a compact $K \subset \mathcal{H} \setminus \mathcal{L}$, $m := \inf_K \|\delta_{\bar\theta}\|_{\mathcal{H}} > 0$ and $\sup_K \mathrm{dist}_{\mathcal{H}}(\cdot, V_\varepsilon^{\mathrm{c}}) \le M < \infty$, while (4) makes $\delta_\varepsilon^{\mathrm{c}} \to \delta_{\bar\theta}$ uniformly on K, so $\inf_K R_\varepsilon \ge m/(2M) > 0$ for small ε. This bounds the local step, not the whole run, since an orbit may enter the afterimage later; on one ray, however, it also bounds the whole run. For ADMM with $\sigma = \tau = 1$ and $t \le 0$, both coordinates of $z = (t, t-1)$ are nonpositive, so $\Pi_{\mathcal{K}}(z) = 0$ and Table 1 sends the whole ray to $T_{\theta_\varepsilon^{\mathrm{c}}}(t, t-1) = (\varepsilon/2,\ -1 - \varepsilon/2)$, at distance $\varepsilon/\sqrt{2}$ from F_ε. Writing $z = F_\varepsilon + w$ on the quadrant $z_1 \ge 0 \ge z_2$ makes the map affine with linear part $0.5\left[\begin{smallmatrix} 1 & -1 \\ 1 & 1 \end{smallmatrix}\right] = 2^{-1/2}Q$, Q orthogonal; the orbit stays there as long as $\|w\|_{\mathcal{H}} < \varepsilon$, which holds at $k = 1$ and persists under the contraction; so $\mathrm{dist}_{\mathcal{H}}(z^{(k)}, V_\varepsilon^{\mathrm{c}}) = \varepsilon\, 2^{-k/2}$ exactly for every $k \ge 1$, and one step reaches any fixed accuracy $\zeta > 0$ once $\varepsilon \le \sqrt{2}\,\zeta$.

Restart and Halpern acceleration do not remove this plateau. Restart and Halpern-type schemes are popular accelerations for FOM on LP (Applegate et al., 2023; Lu & Yang, 2024), so it is natural to ask whether they shorten the plateau above. For the fixed maps of Table 1, they do not. On $\mathcal{L}$, $T_\varepsilon^{\mathrm{c}} z = z + d_\varepsilon$ with $d_\varepsilon := \bar{\mathcal{A}}^\dagger \varepsilon = (\varepsilon/2, -\varepsilon/2)$ parallel to $\mathcal{L}$, so the plain iteration advances by $\|d_\varepsilon\|_{\mathcal{H}} = \varepsilon/\sqrt{2}$ per step and needs $\Theta(\varepsilon^{-1})$ steps to traverse an $O(1)$ portion of $\mathcal{L}$. Here $\Theta(\varepsilon^{-1})$ counts operator evaluations for a fixed relative reduction of $\mathrm{dist}_{\mathcal{H}}(z, V_\varepsilon^{\mathrm{c}})$; it is not the time to reach a fixed residual tolerance, which the $O(\varepsilon)$ residual on $\mathcal{L}$ meets immediately. Restarts are counted by their operator evaluations. For the Halpern iteration $z^{(k+1)} = \dfrac{1}{k+2} z^{(0)} + \dfrac{k+1}{k+2} T_\varepsilon^{\mathrm{c}} z^{(k)}$, induction gives $z^{(k)} = z^{(0)} + \dfrac{k}{2} d_\varepsilon$ while the orbit stays on $\mathcal{L}$ — half the speed of the plain iteration; applied to the reflected map $2T_\varepsilon^{\mathrm{c}} - \mathrm{Id}$, the same computation gives $z^{(k)} = z^{(0)} + k d_\varepsilon$, recovering the plain speed. Restarting either scheme only resets the anchor $z^{(0)}$ and leaves the displacement per operator evaluation at $O(\varepsilon)$. Hence these accelerations change constants, but not the $\Theta(\varepsilon^{-1})$ plateau; the same computation applies to PDHG with $d_\varepsilon = (0, 0, \eta_y \varepsilon)$. This does not contradict the restart guarantees of Applegate et al. (2023), which accelerate with respect to a fixed instance's sharpness constant: along our family that constant degenerates, as the error-bound reading of ρ_ε in §2.2 records. Solvers that adapt σ, the step sizes, or the scaling fall outside this computation. The constant-step phase itself is a known regime of ADMM on LP/QP (Boley, 2013); the afterimage certificate explains it through a neighboring problem and extends it to non-polyhedral and non-attained limits.

A.4.2 SDP SQUARE-ROOT HÖLDER

The $\Theta(h_\varepsilon^{1/2})$ dislocation below is consistent with the square-root error bounds for LMIs (Sturm, 2000; Ding & Udell, 2023); we derive it directly, since this center has no strictly feasible dual slack and strict complementarity fails at the collapse endpoint. Throughout $\sigma = \tau = 1$, $\mathcal{K} = \mathbb{S}^2_+$ is self-dual and $\bar{\mathcal{A}}\bar{\mathcal{A}}^* = \|A_1\|_{\mathsf{F}}^2 = 3$, so $\mathcal{A}^\dagger b = \dfrac{b}{3} A_1$ and $P_{\mathcal{A}}^{\perp} Y = Y - \dfrac{1}{3}\langle A_1, Y\rangle A_1$. Both fixed sets are the images $Z = X - \sigma S$ of the KKT triples (X, y, S).

At the center, the primal minimizes X_{22} subject to $2X_{12} = X_{22}$, $X \succeq 0$; positivity gives $X_{22} \geq 0$ while $X = 0$ is feasible, so optimality forces $X_{22} = 0$ and then the constraint forces $X_{12} = 0$; the dual forces $y = 0$, $S = E_{22}$, leaving the ray $V_\varepsilon^{\mathrm{c}} = \{\mathrm{diag}\,(t, -1) : t \geq 0\}$. At the petal, the slack $S(y) = E_{22} + \varepsilon^2 E_{11} - yA_1 = \left[\begin{smallmatrix} \varepsilon^2 & -y \\ -y & 1+y \end{smallmatrix}\right]$ is positive semidefinite iff $y^2 \leq \varepsilon^2(1+y)$; since $b = \varepsilon^3 > 0$, the dual maximizes $\varepsilon^3 y$ at the upper root of that quadratic,

$$y_\varepsilon = \frac{1}{2}(\varepsilon^2 + \varepsilon\sqrt{\varepsilon^2 + 4}) = \varepsilon + \frac{1}{2}\varepsilon^2 + O(\varepsilon^3), \qquad X_\varepsilon = \frac{u_\varepsilon u_\varepsilon^\top}{\sqrt{\varepsilon^2 + 4}}, \quad u_\varepsilon := (y_\varepsilon,\, \varepsilon^2)^\top,$$

where $\det S(y_\varepsilon) = 0$, so $S(y_\varepsilon)$ has kernel spanned by u_ε and complementarity forces $X = \varsigma\, u_\varepsilon u_\varepsilon^\top$, with $\bar{\mathcal{A}}(u_\varepsilon u_\varepsilon^\top) = 2y_\varepsilon\varepsilon^2 - \varepsilon^4 = \varepsilon^3\sqrt{\varepsilon^2+4}$ fixing $\varsigma = (\varepsilon^2+4)^{-1/2}$. The petal is thus the single point $V_\varepsilon^{\mathrm{p}} = \{Z_\varepsilon\}$, $Z_\varepsilon := X_\varepsilon - S(y_\varepsilon)$.

For the dislocation, $\|Z_\varepsilon - \mathrm{diag}\,(t, -1)\|_{\mathsf{F}}^2 = ((Z_\varepsilon)_{11} - t)^2 + 2(Z_\varepsilon)_{12}^2 + ((Z_\varepsilon)_{22} + 1)^2$. By $y_\varepsilon^2 = \varepsilon^2(1 + y_\varepsilon)$, $(Z_\varepsilon)_{11} = \varepsilon^2(\dfrac{1+y_\varepsilon}{\sqrt{\varepsilon^2+4}} - 1) < 0$ for small ε, so the minimum over $t \geq 0$ sits at $t = 0$ and the distance is exact, not just asymptotic; with $(Z_\varepsilon)_{12} = \varepsilon + O(\varepsilon^2)$ and $(Z_\varepsilon)_{22} + 1 = -\varepsilon + O(\varepsilon^2)$, it is $\mathrm{dist}_{\mathcal{H}}(Z_\varepsilon, V_\varepsilon^{\mathrm{c}}) = \sqrt{3}\,\varepsilon + O(\varepsilon^2)$. Finally, $\theta_\varepsilon^{\mathrm{p}}$ and $\bar{\theta}$ differ only in b and C, so Theorem 2.2 (i) makes the defect state-free: $\delta_\varepsilon^{\mathrm{c}} - \delta_\varepsilon^{\mathrm{p}} \equiv \mathcal{A}^\dagger(0 - \varepsilon^3) - P_{\mathcal{A}}^{\perp}(-\varepsilon^2 E_{11}) = -\dfrac{\varepsilon^3}{3} A_1 + \varepsilon^2 E_{11}$, since $\langle A_1, E_{11}\rangle = 0$. Its two pieces are orthogonal with squared norms $\varepsilon^6/3$ and ε^4, so it has size $\varepsilon^2\sqrt{1 + \varepsilon^2/3}$; as $\delta_\varepsilon^{\mathrm{p}}(Z_\varepsilon) = 0$, dividing by the dislocation gives $\rho_\varepsilon = \varepsilon/\sqrt{3} \cdot (1 + O(\varepsilon)) = \Theta(\varepsilon)$.

A.4.3 TWO AFTERIMAGES, TWO EXPONENTS

The data, and hence $\mathcal{A}^\dagger$, $P_{\mathcal{A}}^{\perp}$ and the description of the fixed sets, are those of Example 3.2; only the roles change. The frozen petal is that example's center, so $V_\varepsilon^{\mathrm{p},1} = V_Z = \{Q_t = \mathrm{diag}\,(t, -1) : t \geq 0\}$, a ray with endpoint Q_0 and $\{Q_0\}$ as its only proper face; the center is that example's

petal, so $V_\varepsilon^{\mathrm{c}} = \{Z_\varepsilon\}$ with $Z_\varepsilon = X_\varepsilon - S_\varepsilon$ as computed there, and $\mathrm{dist}_{\mathcal{H}}(Z_\varepsilon, V_Z) = \|Z_\varepsilon - Q_0\|_{\mathsf{F}} = \sqrt{3}\,\varepsilon + O(\varepsilon^2)$, the nearest point being the endpoint since $(Z_\varepsilon)_{11} = -0.5\varepsilon^2 + O(\varepsilon^3) < 0$.

For the second petal $\theta_\varepsilon^{\mathrm{p},2} = (\bar{\mathcal{A}}, 0, C_\varepsilon)$, the primal minimizes $\langle C_\varepsilon, X\rangle = \varepsilon^2 X_{11} + X_{22}$ over $2X_{12} = X_{22}$, $X \succeq 0$; both terms are then nonnegative, so $X = 0$ is the unique optimum, while the dual maximizes $0 \cdot y$ over $S(y) = C_\varepsilon - yA_1 \succeq 0$, *i.e.,* over the whole interval $y^2 \le \varepsilon^2(1+y)$, whose endpoints $y_\pm = 0.5(\varepsilon^2 \pm \varepsilon\sqrt{\varepsilon^2+4}) = \pm\varepsilon + 0.5\varepsilon^2 + O(\varepsilon^3)$ are the roots met in Example 3.2. The images $X - S(y) = q(y) = -C_\varepsilon + yA_1$ therefore fill a segment of length $\|A_1\|_{\mathsf{F}}(y_+ - y_-) = \sqrt{3}\,\varepsilon\sqrt{\varepsilon^2+4} = 2\sqrt{3}\,\varepsilon + O(\varepsilon^3)$. Note y_+ is the center's dual optimum, so $Z_\varepsilon = q(y_+) + X_\varepsilon$: the segment ends where the collapse point sits, up to the $O(\varepsilon^2)$ primal part.

Both petals differ from the center in b and/or c only, so Theorem 2.2 (i) gives state-free defects $\delta_\varepsilon^{\mathrm{c}} - \delta_\varepsilon^{\mathrm{p},1} \equiv \dfrac{\varepsilon^3}{3}A_1 - \varepsilon^2 E_{11}$ and $\delta_\varepsilon^{\mathrm{c}} - \delta_\varepsilon^{\mathrm{p},2} \equiv \dfrac{\varepsilon^3}{3}A_1$, the second because the two share $c = C_\varepsilon$; their norms are $\Delta_\varepsilon^1 = \varepsilon^2\sqrt{1+\varepsilon^2/3}$ (orthogonal pieces, as in Example 3.2) and $\Delta_\varepsilon^2 = \varepsilon^3/\sqrt{3}$. Since $\delta_\varepsilon^{\mathrm{p},1}$ vanishes on V_Z and $\delta_\varepsilon^{\mathrm{p},2}$ on the segment, $\|\delta_\varepsilon^{\mathrm{c}}(z)\|_{\mathcal{H}}$ is *constant* on each carrier, equal to Δ_ε^1, resp. Δ_ε^2, and (8) reduces to dividing by the smallest dislocation.

On $\mathcal{S}_\varepsilon^1$, $\|Q_t - Z_\varepsilon\|_{\mathsf{F}}^2 = (t - (Z_\varepsilon)_{11})^2 + 2(Z_\varepsilon)_{12}^2 + ((Z_\varepsilon)_{22} + 1)^2$ increases in $t \ge 0$, so the infimum sits at $t = \varepsilon^\alpha$ and equals $\sqrt{\varepsilon^{2\alpha} + 3\varepsilon^2} + O(\varepsilon^{\alpha+2} + \varepsilon^3)$; with $\Delta_\varepsilon^1 = \varepsilon^2(1 + O(\varepsilon^2))$ this gives $\rho_\varepsilon^1 = \Theta(\varepsilon^{2-\alpha})$ for $\alpha < 1$, $0.5\varepsilon(1 + O(\varepsilon))$ at $\alpha = 1$, and $\dfrac{1}{\sqrt{3}}\varepsilon(1 + o(1))$ for $\alpha > 1$ — the last two being the saturation, forced by the floor $\mathrm{dist}_{\mathcal{H}}(Z_\varepsilon, V_Z) = \sqrt{3}\varepsilon(1 + O(\varepsilon))$ on the denominator. On $\mathcal{S}_\varepsilon^2$, $q(y) - Z_\varepsilon = (y - y_+)A_1 - X_\varepsilon$ gives $\|q(y) - Z_\varepsilon\|_{\mathsf{F}}^2 = 3(y-y_+)^2 - 2(y-y_+)\langle A_1, X_\varepsilon\rangle + \|X_\varepsilon\|_{\mathsf{F}}^2$, with $\langle A_1, X_\varepsilon\rangle = \bar{\mathcal{A}}X_\varepsilon = \varepsilon^3$ by primal feasibility and $\|X_\varepsilon\|_{\mathsf{F}} = 0.5\varepsilon^2 + O(\varepsilon^3)$. The right side decreases in y on $[y_-, y_+]$, so the infimum sits at $y = (1-\eta)y_+$, where $y - y_+ = -\eta y_+$ and the leading term $3\eta^2 y_+^2 = 3\eta^2\varepsilon^2(1 + O(\varepsilon))$ dominates the other two, which are $O(\varepsilon^4)$; hence $D_\varepsilon^2 = \sqrt{3}\,\eta\varepsilon + O(\varepsilon^2) > 0$, which also gives $\mathcal{S}_\varepsilon^2 \cap V_\varepsilon^{\mathrm{c}} = \emptyset$, and $\rho_\varepsilon^2 = \Delta_\varepsilon^2/D_\varepsilon^2 = \dfrac{\varepsilon^2}{3\eta}(1 + O(\varepsilon))$. Both regions take $v_\varepsilon^{\mathrm{p},1} = v_\varepsilon^{\mathrm{p},2} = 0$ and $\mu_\varepsilon \equiv 0$ in Definition 2.8, each sitting inside its own petal's fixed-point set. Figure 6 follows one orbit down the resulting staircase (observed in the run, not implied by the theorems); its ε-independent seed sits at the far end $t = t_1$, an $O(1)$ distance from $V_\varepsilon^{\mathrm{c}}$, so it outlasts $N_\varepsilon = \Theta(\varepsilon^{\alpha-2})$, the worst case over $\mathcal{S}_\varepsilon^1$, which is approached only at the near end $t = \varepsilon^\alpha$.

A.4.4 Escape to the horizon

Write $\bar{\mathcal{A}} = [-1 \;\; 0]$, $\mathcal{A}_\varepsilon = [-1 \;\; \varepsilon]$, $c = (0, 1)$, $a = 1/\varepsilon$, and recall the PDHG step of Table 1, $x^+ = \Pi_{\mathcal{K}}[x - \eta_x(c - \mathcal{A}^* y)]$, $T_\theta z = (x^+,\, y + \eta_y[b - \mathcal{A}(2x^+ - x)])$. The limit parameter is strongly infeasible: $\bar{\mathcal{A}}\mathbb{R}_+^2 = (-\infty, 0]$ is closed and $\mathrm{dist}(1, (-\infty, 0]) = 1 > 0$, so $\bar{\theta}$ admits no KKT point and $\mathrm{Fix}T_{\bar{\theta}} = \emptyset$.

Its drift is nevertheless attained, and on an explicit ray. On $\mathcal{R}_{\bar{\theta}}$, where $x = 0$ and $y \ge 0$, we have $c - \bar{\mathcal{A}}^* y = (y, 1)$, so both entries of $x - \eta_x(y, 1)$ are nonpositive, $x^+ = 0 = x$, and $y^+ - y = \eta_y$; hence $\delta_{\bar{\theta}} \equiv (0, 0, \eta_y)$ there — the primal iterate stays at the origin while the dual moves at constant speed, which reflects the LP dual growing without bound. Seeding at $z^{(0)} = 0 \in \mathcal{R}_{\bar{\theta}}$, the orbit stays on the ray with $\delta_{\bar{\theta}}(z^{(k)}) \equiv (0, 0, \eta_y)$, so Proposition 2.7 forces $v_{\bar{\theta}} = (0, 0, \eta_y)$. Conversely $\delta_{\bar{\theta}}(z) = v_{\bar{\theta}}$ forces $x^+ = x$, whose second coordinate $x_2 = \max\{x_2 - \eta_x, 0\}$ gives $x_2 = 0$; the third residual coordinate is then $\eta_y[1 - \bar{\mathcal{A}}x] = \eta_y(1 + x_1)$, so $x_1 = 0$, and the first coordinate $0 = \max\{-\eta_x y, 0\}$ gives $y \ge 0$. The attainment set is therefore exactly $\mathcal{R}_{\bar{\theta}}$, so $\mathrm{tag}(\bar{\theta}) = (+, \mathrm{fin})$ and $V_\varepsilon^{\mathrm{p},1} = \{z : \delta_{\bar{\theta}}(z) = v_{\bar{\theta}}\} = \mathcal{R}_{\bar{\theta}}$: the petal keeps a carrier in the generalized sense even though $\mathrm{Fix}T_{\bar{\theta}}$ is empty. Being piecewise affine with polyhedral pieces, these maps have closed $\mathrm{ran}\delta_\theta$ on an LP, so an $\inf$ tag cannot occur here.

The center, by contrast, is solvable. Feasibility $-x_1 + \varepsilon x_2 = 1$, $x \ge 0$ forces $x_2 \ge a$, so $\langle c, x\rangle = x_2$ is minimized uniquely at $x = (0, a)$, while the dual maximizes y over $c - \mathcal{A}_\varepsilon^* y = (y,\, 1 - \varepsilon y) \ge 0$, *i.e.,* over $[0, a]$, uniquely at $y = a$. Hence $V_\varepsilon^{\mathrm{c}} = \{(0, a, a)\}$, a singleton, and $D_\varepsilon = \sqrt{2}\,a$.

The two moduli follow. For $z = (0, 0, y)$ with $0 \le y \le a$ the vector $c - \mathcal{A}_\varepsilon^* y = (y,\, 1 - \varepsilon y)$ is again nonnegative, so $x^+ = 0 = x$ and $\delta_\varepsilon^{\mathrm{c}}(z) = (0, 0, \eta_y) = \delta_{\bar\theta}(z) = v_{\bar\theta}$: on $\mathcal{S}_\varepsilon^1$ the center–petal defect vanishes identically, giving $\mu_\varepsilon \equiv 0$, and the residual there is the ε-independent $\|v_{\bar\theta}\|_\theta = \sqrt{\eta_y}$. Since $\|z - (0, a, a)\|_{\mathcal{H}}^2 = a^2 + (y - a)^2$ decreases in y, the infimum over $\mathcal{S}_\varepsilon^1$ sits at $y = D_\varepsilon^\gamma$, whence $\rho_\varepsilon^1 = \eta_y / \sqrt{a^2 + (a - D_\varepsilon^\gamma)^2} = (\eta_y \varepsilon / \sqrt{2})(1 + O(\varepsilon^{1-\gamma}))$. For the second petal the same KKT computation with $b = 1 - \xi_\varepsilon$ gives $V_\varepsilon^{\mathrm{p},2} = \{(0, (1 - \xi_\varepsilon)a,\, a)\}$ whenever $0 < \xi_\varepsilon < 1$; as it differs from the center in b only, Theorem 2.2 (ii) makes the defect state-free and equal to $(0, 0, \eta_y \xi_\varepsilon)$, while the dislocation is $\xi_\varepsilon a$. Both are linear in ξ_ε, which therefore cancels: $\rho_\varepsilon^2 = \eta_y \varepsilon$ exactly, with ξ_ε setting the position of $\mathcal{S}_\varepsilon^2$ and nothing else.

A.4.5 SOCP FACE DROP

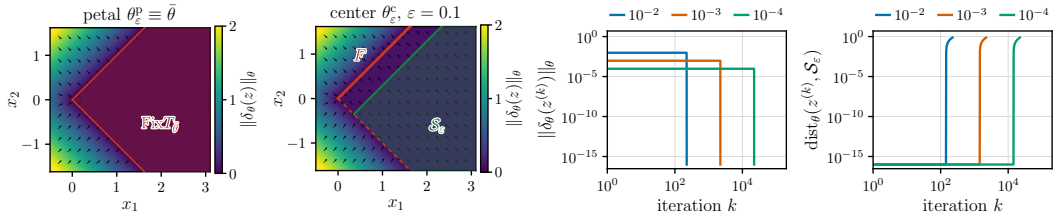

Figure 8: Example A.4 under PDHG; layout as in Figure 4, on the invariant plane $\{x_3 = y = 0\}$. The wedge $V_\varepsilon^{\mathrm{p}}$ (solid, then dashed) stays dark although only its edge F is fixed; from one ε-independent seed in $\mathcal{S}_\varepsilon$ (shaded) the residual sits at $\sqrt{2\eta_x}\,\varepsilon$.

Example A.4 (SOCP face drop). Take $\mathcal{K} = \mathcal{Q}^3$, $\bar{\mathcal{A}} := [0 \;\; 0 \;\; 1]$, $b := 0$, and move only the cost: $\bar\theta := (\bar{\mathcal{A}}, 0, 0)$, $\theta_\varepsilon^{\mathrm{c}} := (\bar{\mathcal{A}}, 0, \varepsilon(1, -1, 0))$, $\theta_\varepsilon^{\mathrm{p}} \equiv \bar\theta$ and $\|\theta_\varepsilon^{\mathrm{c}} - \bar\theta\|_{\mathcal{D}} = \sqrt{2}\,\varepsilon$. Both tags equal $(0, \mathrm{fin})$. The limit cost vanishes, so under PDHG on $\mathcal{H} = \mathbb{R}^4$ the petal is the whole wedge $V_\varepsilon^{\mathrm{p}} = \mathrm{Fix}T_{\bar\theta} = \{(x_1, x_2, 0, 0) : x_1 \ge |x_2|\}$, while the tilt selects one edge, $V_\varepsilon^{\mathrm{c}} = F := \{(r, r, 0, 0) : r \ge 0\}$, for every $\varepsilon > 0$ (Appendix A.4.5). Theorem 2.9 applies, for any $c_0 > 0$, to any nonempty bounded $\mathcal{S}_\varepsilon \equiv \mathcal{S} \subset \{z \in V_\varepsilon^{\mathrm{p}} : \mathrm{dist}_{\mathcal{H}}(z, F) \ge c_0\}$, on which the defect is exactly $-\eta_x \varepsilon(1, -1, 0, 0)$, giving $\rho_\varepsilon \le \sqrt{2}\eta_x \varepsilon / c_0 = \Theta(\|\theta_\varepsilon^{\mathrm{c}} - \bar\theta\|_{\mathcal{D}})$; see Figure 8.

At $\varepsilon = 0$ the objective vanishes, so the optimal set is the *entire* feasible slice $\mathcal{Q}^3 \cap \{x_3 = 0\} = \{x_1 \ge |x_2|\}$; moving only c leaves that slice fixed, so the tilt merely selects one of its two extreme rays, while the dual optimum $y = 0$ stays unique. Hence $V_\varepsilon^{\mathrm{c}} = F \subsetneq V_\varepsilon^{\mathrm{p}}$ exactly for every $\varepsilon > 0$, with no limit taken. On $\mathcal{S}$ the constant is exact: $\mathrm{dist}_{\mathcal{H}}((x_1, x_2, 0, 0), F) = |x_1 - x_2| / \sqrt{2} \ge c_0$ forces $x_1 - x_2 \ge \sqrt{2} c_0$, so for small ε the point $x - \eta_x \varepsilon(1, -1, 0)$ still lies in $\mathcal{Q}^3$, the projection is inactive, and $\delta_\varepsilon^{\mathrm{c}} \equiv -\eta_x \varepsilon(1, -1, 0, 0)$ has norm exactly $\sqrt{2}\eta_x \varepsilon$; dividing by c_0 gives $\rho_\varepsilon \le \sqrt{2}\eta_x \varepsilon / c_0$.

A.4.6 SOCP ZERO GAP, NOT ATTAINED

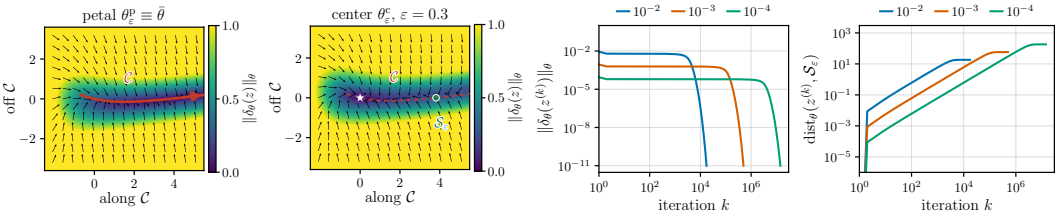

Figure 9: Example A.5 under ADMM; layout as in Figure 4, on the plane through $V_\varepsilon^{\mathrm{c}}$ spanned by the chord to the seed and $\mathcal{C}$'s transverse direction, so both marked points are exactly in-slice. *Left two:* the field has *no* zero — $\mathcal{C}$ merely darkens and runs off the frame (arrow); then a zero appears at $V_\varepsilon^{\mathrm{c}}$ (star), with the seed $\mathcal{S}_\varepsilon$ (dot) further out at $\lambda = 2$. *Right two:* one orbit per $\varepsilon = 10^{-2}, 10^{-3}, 10^{-4}$.

Example A.5 (SOCP zero gap, not attained). Let $\mathcal{K} = \mathcal{Q}^3$, $\bar{\mathcal{A}} := [0 \;\; 0 \;\; 1]$, $b := 1$, $u := (0, 1, 0)$, and move the cost only, freezing the petal at the limit: $\bar\theta = \theta_\varepsilon^{\mathrm{p}} := (\bar{\mathcal{A}},\, 1,\, \bar c)$ with $\bar c := (1, -1, 0)$, $\theta_\varepsilon^{\mathrm{c}} := (\bar{\mathcal{A}},\, 1,\, \bar c + \varepsilon u)$ and $\|\theta_\varepsilon^{\mathrm{c}} - \bar\theta\|_{\mathcal{D}} = \varepsilon$. The limit minimizes $x_1 - x_2$ over $x_3 = 1$, $x \in \mathcal{Q}^3$: the

value is 0, attained by the dual at $y = 0$ but by no feasible x, since $x_1 - x_2 \geq \sqrt{x_2^2 + 1} - x_2 > 0$. So $\bar{\theta} \notin \mathcal{D}_{\mathrm{KKT}}$ and $\mathrm{Fix} T_{\bar{\theta}} = \emptyset$ while $v_{\bar{\theta}} = 0$: $\mathrm{tag}(\bar{\theta}) = (0, \inf)$, a tag the maps of Table 1 cannot exhibit on an LP. Writing $r_t := \sqrt{t^2 + 1}$, the images Z_t of $x_t = (r_t, t, 1)$, $s_t = (1, -t/r_t, -1/r_t)$ trace a curve $\mathcal{C}$ carrying $\|\delta_{\bar{\theta}}(Z_t)\|_{\mathcal{H}} = 1 - t/r_t = \Theta(t^{-2})$, which tends to 0 without reaching it. Since $V_\varepsilon^{\mathrm{p}} = \emptyset$, $\mu_\varepsilon > 0$ is forced and $\mathcal{S}_\varepsilon$ is carried by a residual sublevel set. The petal's residual $\Theta(t^{-2})$ and the state-free defect $\delta_\varepsilon^{\mathrm{c}} - \delta_\varepsilon^{\mathrm{p}} \equiv -\varepsilon u$ match at $t \asymp \varepsilon^{-1/2}$, which is where the center's unique fixed point sits: $V_\varepsilon^{\mathrm{c}} = \{Z_{t_\varepsilon^\star}\}$ lies *on* $\mathcal{C}$ with $t_\varepsilon^\star = 2^{-1/2}\varepsilon^{-1/2}(1 + O(\varepsilon))$, escaping as $D_\varepsilon = \varepsilon^{-1/2}(1 + O(\varepsilon))$. Seeding there, at $\mathcal{S}_\varepsilon := \{Z_{t_\varepsilon}\}$ with $t_\varepsilon := \lambda\varepsilon^{-1/2}$ and any fixed $\lambda > 0$, gives $\mu_\varepsilon = \Theta(\varepsilon)$ and $\rho_\varepsilon = \Theta(\varepsilon^{3/2})$, so $N_\varepsilon = \Theta(\varepsilon^{-3/2})$; see Figure 9. Theorem 2.11 applies, though its own carrier sits inside $\mathbb{B}_{\mathcal{H}}(0, D_\varepsilon^\gamma)$ whereas $\mathcal{S}_\varepsilon$ sits *at* the horizon; confined to the ball the same construction yields only $\Theta(\varepsilon^{\gamma+1/2})$.

Here $\bar{\mathcal{A}}\bar{\mathcal{A}}^* = 1$, $\bar{\mathcal{A}}^\dagger b = (0, 0, 1)$, $P_{\mathcal{A}} = \mathrm{diag}\,(0, 0, 1)$, $P_{\mathcal{A}}^\perp = \mathrm{diag}\,(1, 1, 0)$, and $\sigma = \tau = 1$ makes the metric Euclidean. Dual feasibility $\bar{c} - \bar{\mathcal{A}}^* y = (1, -1, -y) \in \mathcal{Q}^3$ reads $1 \geq \sqrt{1 + y^2}$ and forces $y = 0$; with the primal value unattained, $\mathrm{KKT}(\bar{\theta}) = \emptyset$ and Assumption 2.1 (i) gives $\mathrm{Fix} T_{\bar{\theta}} = \emptyset$.

The triples are almost-KKT: $x_t, s_t \in \partial\mathcal{Q}^3$, $\langle x_t, s_t \rangle = 0$ and $\bar{\mathcal{A}} x_t = 1$, so $\Pi_{\mathcal{K}}(Z_t) = x_t$ and only the dual residual survives, $\delta_{\bar{\theta}}(Z_t) = (1 - t/r_t)u$. Hence $\inf_z \|\delta_{\bar{\theta}}(z)\|_{\mathcal{H}} = 0 = v_{\bar{\theta}}$, while $\delta_{\bar{\theta}}(z) = 0$ would produce a KKT triple: $v_{\bar{\theta}} \notin \mathrm{ran}\delta_{\bar{\theta}}$ and $\mathrm{tag}(\bar{\theta}) = (0, \inf)$. None of the Table 1 maps can do this on an LP: they are piecewise affine with polyhedral pieces, so $\mathrm{ran}\delta_\theta$ is closed and the minimal displacement is attained.

The center is solvable: minimizing $x_1 - (1 - \varepsilon)x_2$ forces $x_2/r_{x_2} = 1 - \varepsilon$, giving the unique optima $x_\varepsilon^\star = (\kappa_\varepsilon^{-1}, (1 - \varepsilon)\kappa_\varepsilon^{-1}, 1)$ with $\kappa_\varepsilon := \sqrt{2\varepsilon - \varepsilon^2}$ and $y_\varepsilon^\star = \kappa_\varepsilon$. Both lie on $\mathcal{C}$, so $V_\varepsilon^{\mathrm{c}} = \{Z_{t_\varepsilon^\star}\}$ with $t_\varepsilon^\star = (1 - \varepsilon)/\kappa_\varepsilon$ and $D_\varepsilon = \sqrt{2}\, t_\varepsilon^\star(1 + o(1))$.

For the moduli, $t_\varepsilon = \lambda\varepsilon^{-1/2}$ gives $\mu_\varepsilon = 1 - \lambda/\sqrt{\lambda^2 + \varepsilon} = \varepsilon/(2\lambda^2) + O(\varepsilon^2)$, which is Definition 2.8 with $v_\varepsilon^{\mathrm{p}} = 0$; since $\delta_\varepsilon^{\mathrm{c}} = \delta_\varepsilon^{\mathrm{p}} - \varepsilon u$ on $\mathcal{C}$, the center residual is $|\mu_\varepsilon - \varepsilon|$, a *signed* cancellation that is exact to leading order at $\lambda = 2^{-1/2}$. Along $\mathcal{C}$, $\dfrac{d}{dt} Z_t = (1, 1, 0) + O(t^{-2})$, so $\|Z_t - Z_{t'}\|_{\mathcal{H}} = \sqrt{2}\,|t - t'|\,(1 + O(\min(t, t')^{-2}))$ and the dislocation is $\sqrt{2}\,|\lambda - \lambda_\varepsilon^\star|\,\varepsilon^{-1/2}(1 + O(\varepsilon))$ with $\lambda_\varepsilon^\star := t_\varepsilon^\star \varepsilon^{1/2} = 2^{-1/2}(1 - \dfrac{3}{4}\varepsilon + O(\varepsilon^2))$. Dividing cancels $\left|\lambda - 2^{-1/2}\right|$ and leaves $\rho_\varepsilon = (\lambda + 2^{-1/2})(\sqrt{2}\lambda^2)^{-1}\varepsilon^{3/2}(1 + o(1))$; at $\lambda = 2^{-1/2}$ both leading terms vanish and one further order returns the same formula, so it is continuous there. A *fixed* λ is admissible for all small ε, whereas an interval of λ's straddling $2^{-1/2}$ would eventually contain the moving $\lambda_\varepsilon^\star$ and meet $V_\varepsilon^{\mathrm{c}}$.

A.4.7 SDP WITH A NON-CLOSED IMAGE

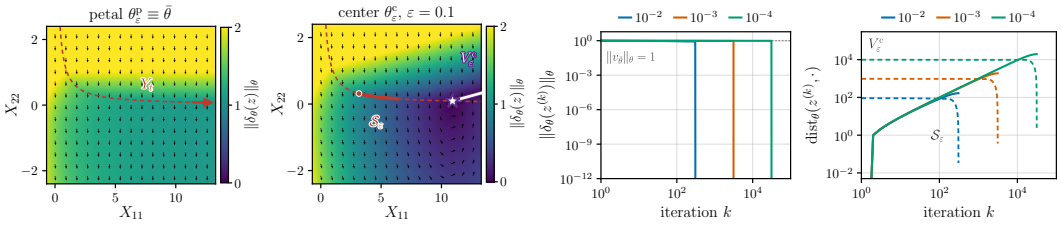

Figure 10: Example A.6 under ADMM; layout as in Figure 4, on the plane $\{X_{12} = 1\}$, which contains the boundary curve Y_t and the center's optimal curve exactly; the horizontal axis is compressed, the tip escaping as ε^{-1} while the transverse structure stays $O(1)$. *Left two:* the color never darkens below $\|v_{\bar{\theta}}\|_{\mathcal{H}} = 1$ and Y_t (dashed) runs off the frame (arrow); then the zero set appears as the unbounded curve $V_\varepsilon^{\mathrm{c}}$ (white), its tip (star) on Y_t, with $\mathcal{S}_\varepsilon$ (solid red, seed dot at its near end) the arc at $t \asymp \varepsilon^{-1/2}$. *Right two:* one orbit per ε, seeded at the arc's near endpoint; solid curves give $\mathrm{dist}_\theta(z^{(k)}, \mathcal{S}_\varepsilon)$ and dashed ones $\mathrm{dist}_\theta(z^{(k)}, V_\varepsilon^{\mathrm{c}})$.

Example A.6 (SDP with a non-closed image). Let $\mathcal{X} = \mathbb{S}^2$, $\mathcal{K} = \mathbb{S}_+^2$, $\mathcal{Y} = \mathbb{R}^2$, $b := (1, -1)$, and move the operator only, freezing the petal at the limit: $\mathcal{A}_\varepsilon X := (X_{12},\ X_{22} - \varepsilon X_{11})$, $\bar{\theta} = \theta_\varepsilon^{\mathrm{p}} := (\mathcal{A}_0,\ b,\ 0)$ and $\theta_\varepsilon^{\mathrm{c}} := (\mathcal{A}_\varepsilon,\ b,\ 0)$, so $\|\theta_\varepsilon^{\mathrm{c}} - \bar{\theta}\|_{\mathcal{D}} = \varepsilon$. The cost vanishes, so the optimal

set is the feasible set. The operator stays nondegenerate, $\mathcal{A}_\varepsilon\mathcal{A}_\varepsilon^* = \operatorname{diag}\left(0.5,\ 1+\varepsilon^2\right) \succeq_{\mathcal{Y}} 0.5\mathrm{Id}$ for every $\varepsilon \geq 0$; what fails is its interaction with the cone boundary, $\mathcal{A}_0(\mathbb{S}^2_+) = \{(u,v) : v > 0\} \cup \{(0,0)\}$, which is not closed. Since b has v-coordinate -1, the nearest image points are $(1,v)$ with $v \downarrow 0$, so $\operatorname{dist}(b, \mathcal{A}_0(\mathbb{S}^2_+)) = 1$ is realized by no $X \succeq 0$: $\mathrm{Fix}T_{\bar\theta} = \emptyset$, $v_{\bar\theta} = -E_{22} \notin \operatorname{ran}\delta_{\bar\theta}$ with $\|v_{\bar\theta}\|_{\mathcal{H}} = 1$, and $\operatorname{tag}(\bar\theta) = (+,\inf)$, the last cell. The drift is approached along the rank-one *boundary curve* $Y_t := \left[\begin{smallmatrix} t & 1 \\ 1 & 1/t \end{smallmatrix}\right] \in \partial\mathbb{S}^2_+$, where $\delta_{\bar\theta}(Y_t) = -(1+\frac{1}{t})E_{22}$ *exactly*. The center is solvable, $\operatorname{tag}(\theta^{\mathrm{c}}_\varepsilon) = (0,\mathrm{fin})$, with optimal set the unbounded ray $V^{\mathrm{c}}_\varepsilon = \{X_p := \left[\begin{smallmatrix} p & 1 \\ 1 & \varepsilon p - 1 \end{smallmatrix}\right] : p \geq r_\varepsilon\}$, $r_\varepsilon := 0.5(1+\sqrt{1+4\varepsilon})/\varepsilon = \varepsilon^{-1} + 1 + O(\varepsilon)$, whose tip $X_{r_\varepsilon} = Y_{r_\varepsilon}$ sits on the boundary curve and escapes, $D_\varepsilon = r_\varepsilon + r_\varepsilon^{-1}$.

On the boundary curve every residual is the scaled primal infeasibility, $\|\delta^{\mathrm{c}}_\varepsilon(Y_t)\|_{\mathcal{H}} = |g_\varepsilon(t)|\,/\sqrt{1+\varepsilon^2}$ with $g_\varepsilon(t) := 1 + 1/t - \varepsilon t$: the limit's own $1/t$ and the perturbation's εt are of the same size when $t \asymp \varepsilon^{-1/2}$, and cancel at $t = \varepsilon^{-1/2}$, where $g_\varepsilon = 1$ exactly. Fixing $0 < c_0 < c_1$, the *diverging* arc $\mathcal{S}_\varepsilon := \{Y_t : c_0\varepsilon^{-1/2} \leq t \leq c_1\varepsilon^{-1/2}\}$ has $\mu_\varepsilon = \varepsilon^{1/2}/c_0 = \Theta(\varepsilon^{1/2})$ and $\rho_\varepsilon = \varepsilon(1+\mu_\varepsilon + O(\varepsilon)) = \Theta(\varepsilon)$, hence $N_\varepsilon = \Theta(\varepsilon^{-1})$, the order Theorem 2.11 displays. The carrier itself *diverges*, at rate $\varepsilon^{-1/2}$ while staying far short of the tip at ε^{-1}, and is the first region here that is positive-dimensional *and* escapes entirely to infinity. Its center residual converges to $\|v_{\bar\theta}\|_{\mathcal{H}} = 1$ uniformly on the arc, $\sup_{z\in\mathcal{S}_\varepsilon} |\|\delta^{\mathrm{c}}_\varepsilon(z)\|_{\mathcal{H}} - 1| = O(\varepsilon^{1/2})$, while $\mu_\varepsilon = \sup_{z\in\mathcal{S}_\varepsilon}\|\delta_{\bar\theta}(z) - v_{\bar\theta}\|_{\mathcal{H}} = \varepsilon^{1/2}/c_0$ records the petal's own gap; and the perturbation moves $\mathcal{A}$, so no U with $\mathcal{A}$ fixed contains the families; Theorem 2.2 is therefore unavailable and the defect is state dependent. See Figure 10.

Throughout $\sigma = \tau = 1$, so $\|\cdot\|_\theta = \|\cdot\|_{\mathsf{F}}$ and $\mathcal{H} = \mathcal{X} = \mathbb{S}^2$; with $c = 0$ and $\Pi_+ := \Pi_{\mathbb{S}^2_+}$, $\Pi_- := \mathrm{Id} - \Pi_+$, Table 1 reads $\delta_\theta(Z) = -P_{\mathcal{A}}(\Pi_+(Z) - \mathcal{A}^\dagger b) - P^\perp_{\mathcal{A}}(\Pi_-(Z))$. Since full row rank is an open condition, $\mathcal{A}_0\mathcal{A}_0^* = \operatorname{diag}(0.5, 1) \succ_{\mathcal{Y}} 0$ lets us fix a compact $U \ni \bar\theta$ on which $\mathcal{A}\mathcal{A}^* \succeq_{\mathcal{Y}} \frac{1}{4}\mathrm{Id}$, and Table 1 then gives Assumption 2.1 throughout U, the limit included.

The image and the tag. For $X \succeq 0$ we have $X_{22} \geq 0$, and $X_{22} = 0$ forces $X_{12} = 0$, while every $v > 0$ is hit by $X = \left[\begin{smallmatrix} u^2/v & u \\ u & v \end{smallmatrix}\right]$; that is the stated image, whose closure contains $(1,0)$ although the set does not. On the boundary curve, $\det Y_t = 0$ and $t > 0$ give $Y_t \succeq 0$, so $\Pi_+(Y_t) = Y_t$ and the map collapses to the scaled primal infeasibility,

$$\delta_\theta(Y_t) = -P_{\mathcal{A}}(Y_t - \mathcal{A}^\dagger b) = -\mathcal{A}^\dagger(\mathcal{A}Y_t - b), \tag{28}$$

which at $\bar\theta$ is $-(1+\frac{1}{t})E_{22} \to -E_{22}$. Conversely $\|\mathcal{A}_0^\dagger(u,v)\|^2_{\mathsf{F}} = 2u^2 + v^2 \geq \|(u,v)\|_2^2$, the pseudo-inverse being expansive here, and the two terms of the map lie in the orthogonal subspaces $\operatorname{ran}\mathcal{A}_0^*$ and $\ker\mathcal{A}_0$; dropping the second and writing $X := \Pi_+(Z)$ gives, for *every* $Z \in \mathcal{H}$,

$$\|\delta_{\bar\theta}(Z)\|_{\mathsf{F}} \ \geq\ \|\mathcal{A}_0^\dagger(\mathcal{A}_0X - b)\|_{\mathsf{F}} \ \geq\ \|\mathcal{A}_0X - b\|_2 \ \geq\ \operatorname{dist}(b, \mathcal{A}_0(\mathbb{S}^2_+)) = 1. \tag{29}$$

So the infimum is exactly 1. Now $\mathrm{cl}(\operatorname{ran}\delta_{\bar\theta})$ is convex by Proposition 2.7, so its minimum-norm element is unique and equals $-E_{22}$; equality in (29) would need an image point at distance exactly 1, so $v_{\bar\theta} \notin \operatorname{ran}\delta_{\bar\theta}$ and $\operatorname{tag}(\bar\theta) = (+,\inf)$.

The center. Solving $\mathcal{A}_\varepsilon X = b$ forces $X_{12} = 1$ and $X_{22} = \varepsilon X_{11} - 1$, so with $p := X_{11}$ the feasible points are exactly the X_p, and $X_p \succeq 0$ iff $p \geq r_\varepsilon$. Taking $y = 0$, $S = 0$ makes each of them a KKT point, and it is the only dual slack: $S = -\mathcal{A}^*_\varepsilon y \succeq 0$ needs $\varepsilon y_2 \geq 0$ and $-y_2 \geq 0$ at once, forcing $y_2 = 0$, after which $\det S = -y_1^2/4 \geq 0$ forces $y_1 = 0$. Hence $V^{\mathrm{c}}_\varepsilon$ is as stated and $\operatorname{tag}(\theta^{\mathrm{c}}_\varepsilon) = (0,\mathrm{fin})$. At $p = r_\varepsilon$ the determinant vanishes, so $X_{r_\varepsilon} = Y_{r_\varepsilon}$, and $\|X_p\|_{\mathsf{F}}$ increases on $p \geq r_\varepsilon$, making the tip the nearest fixed point to the origin, with $D_\varepsilon = r_\varepsilon + r_\varepsilon^{-1}$.

The two moduli. By (28) and $v_{\bar\theta} = -E_{22}$, $\|\delta_{\bar\theta}(Y_t) - v_{\bar\theta}\|_{\mathcal{H}} = 1/t$, largest on $\mathcal{S}_\varepsilon$ at the near endpoint: $\mu_\varepsilon = \varepsilon^{1/2}/c_0$, and $V^{\mathrm{p}}_\varepsilon = \emptyset$ leaves no smaller choice available. For the center, $\mathcal{A}_\varepsilon Y_t - b = (0,\ g_\varepsilon(t))$ and $\mathcal{A}^\dagger_\varepsilon(0,w)$ has norm $|w|\,/\sqrt{1+\varepsilon^2}$, which is the stated residual; writing $t = c\varepsilon^{-1/2}$ with $c \in [c_0, c_1]$ gives $g_\varepsilon = 1 + \varepsilon^{1/2}(c^{-1} - c)$, within $O(\varepsilon^{1/2})$ of 1 *uniformly* in c, which is the uniform rise. The nearest fixed point to Y_t is again the tip, since $r_\varepsilon \sim \varepsilon^{-1}$ dwarfs $c_1\varepsilon^{-1/2}$; so $\mathcal{S}_\varepsilon \cap V^{\mathrm{c}}_\varepsilon = \emptyset$ and $r_\varepsilon - t = \varepsilon^{-1}(1 - c\varepsilon^{1/2} + O(\varepsilon))$. The $-c\varepsilon^{1/2}$ of the denominator cancels that of the numerator, leaving $\|\delta^{\mathrm{c}}_\varepsilon(Y_t)\|_{\mathcal{H}}/\operatorname{dist}_{\mathcal{H}}(Y_t, V^{\mathrm{c}}_\varepsilon) = \varepsilon(1 + \varepsilon^{1/2}/c + O(\varepsilon))$ uniformly in c; the supremum is therefore pinned

from both sides, with no monotonicity needed, at $\rho_\varepsilon = \varepsilon(1 + \mu_\varepsilon + O(\varepsilon))$, the surviving correction contributed by the near end. Theorem 2.11 applies, its conclusion $O(D_\varepsilon^{-1} + D_\varepsilon^{\gamma-1}\|\theta_\varepsilon^{\mathrm{c}} - \bar{\theta}\|_{\mathcal{D}}) = O(\varepsilon)$ being the order realized here, and the arc meets the theorem's near-minimizer test once $\nu_\varepsilon := \mu_\varepsilon$ is chosen, since $\sup_{\mathcal{S}_\varepsilon}\|Z\|_{\mathsf{F}} \asymp \varepsilon^{-1/2}$ sits inside $\mathbb{B}_{\mathcal{H}}(0, D_\varepsilon^{\gamma})$ for every fixed $\gamma > 0.5$. We claim only that the arc *matches* the displayed order.