On the Complexity of Finding Fixed Points for Set-Valued Contractions

Emmanouil-Vasileios Vlatakis-Gkaragkounis* Pucheng Xiong†

Abstract

We study the computational complexity of finding fixed points of set-valued contractions. We first formulate a computational problem for Nadler's fixed-point theorem, PROJECTED-NADLER, and prove that it is CLS-complete by showing its equivalence to CONTINUOUS-LOCALOPT. We then establish a stronger converse to Nadler's theorem, which can be applied as a tool to analyze the convergence rate of the set-valued basic iteration procedure. Finally, we reduce the large-margin triplet stationarity problem to PROJECTED-NADLER. Combining the argument for CLS-hardness of exact triplet stationarity problem [YLCP+25] with the CLS-hardness of approximate KKT points of quadratic programs [FGHS25], we obtain CLS-completeness of large-margin triplet stationarity at any prescribed accuracy.

*Department of Computer Sciences, University of Wisconsin-Madison, email: vlatakis@wisc.edu
†Department of Computer Sciences, University of Wisconsin-Madison, email: pxiong79@wisc.edu

Contents

1 Introduction **3**
- 1.1 Our contributions and techniques . . . 4
- 1.2 Organization, acknowledgment, and AI disclosure . . . 6

2 Preliminaries **7**

3 Nadler's Problem and Its Complexity **8**

4 Converse Nadler's Fixed Point Theorem **11**

5 The Complexity of Large-Margin Triplet Loss Minimization **13**

6 Conclusion and Further Directions **15**

A Additional Preliminaries **19**

B Example: Random-Permutation Cyclic Coordinate Descent **20**

C Omitted Proofs in Section 3 **30**
- C.1 Hölder-continuous local search . . . 31
- C.2 The CLS-membership of β-PROJECTED-NADLER . . . 37
- C.3 The CLS-hardness of β-PROJECTED-NADLER . . . 38

D Background and Omitted Proofs for Section 4 **40**
- D.1 Fryszkowski's problem and set-theoretic remetrization . . . 40
- D.2 Proof of equivalence relations . . . 41
- D.3 Outline for proof of Theorem 4.1 . . . 45
- D.4 Topological facts . . . 46
- D.5 Proof of Theorem 4.1 . . . 48
- D.6 Proof of Corollary 4.2 . . . 53

E Omitted Proofs in Section 5 **53**
- E.1 Technical lemmas . . . 53
- E.2 Proof of Theorem 5.1 and Corollary 5.2 . . . 57

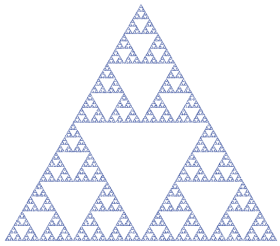


Figure 1: A Hausdorff-metric fixed point: the Sierpiński attractor is the unique fixed point of the Hutchinson operator $\mathcal{H}(S) = w_1(S) \cup w_2(S) \cup w_3(S)$ on the hyperspace of nonempty compact subsets of $\mathbb{R}^2$, illustrating the metric geometry shared by Banach and Nadler contractions.

1 Introduction

Many computations are iterations: gradient descent, power iteration, expectation–maximization, alternating minimization, and coordinate descent all apply an update rule repeatedly in the hope that the orbit settles at a fixed point. For a single-valued map f on a complete metric space, Banach's contraction principle (also known as Banach's fixed point theorem) gives the cleanest possible certificate: if $d(f(x), f(y)) \leq c\, d(x, y)$ for some $c < 1$, then f has a unique fixed point and every orbit converges to it geometrically. The converse of the Banach contraction principle [Bes59; Mey67] shows that this certificate is surprisingly universal: assuming global convergence to the unique fixed point, one can change the metric so that the dynamics become a contraction. Its computational counterpart is equally sharp: finding the fixed point of a succinctly represented contraction map is complete for continuous local search [DP11; DTZ18].

From functions to correspondences. Many algorithms are not intrinsically single-valued. Their next state may depend on a tie, an arrival order, a coordinate permutation, an equilibrium selection, or an adversarially chosen minimizer. Fixing one selection in advance hides the robustness question that matters: does convergence hold for *every* admissible sequence of choices? Random-permutation cyclic coordinate descent is a representative example: once the distribution over permutations is discarded, one epoch sends the current point to the finite set of all possible epoch outputs, and each trajectory of this correspondence is a possibly adaptive sequence of coordinate orders. Correspondences also arise directly in equilibrium computation: best-response maps, feasible-action maps, generalized and quasi-variational inequalities are set-valued, and when the feasible region depends on the current point, replacing the model by a predetermined selection changes the problem [CP82; CPP02; AC24; KS24]. The relevant fixed-point condition is $x \in F(x)$. A useful geometric picture comes from iterated function systems: if $w_1, \ldots, w_m$ are contractions, the Hutchinson operator $\mathcal{H}(S) = \bigcup_{i=1}^{m} w_i(S)$ is a contraction on the hyperspace of nonempty compact sets under the Hausdorff metric, and its fixed point is a self-similar attractor [Hut81; BD85]. Strictly speaking, $\mathcal{H}$ is a single-valued Banach contraction whose *points* are compact sets, but Figure 1 makes the Hausdorff geometry behind Nadler's theorem concrete.

The missing computational counterpart. Nadler's theorem is the contraction principle for correspondences: if (X, d) is complete, each $F(x)$ is nonempty, closed, and bounded, and $H_d(F(x), F(y)) \leq k\, d(x, y)$ for some $k < 1$, then F has a fixed point [Nad69]. The theorem sits naturally between two pillars of computational fixed-point theory. It is set-valued, like Kakutani's

theorem, whose computational formulations lead to PPAD [PVZ23], and contractive, like Banach's theorem, whose computational formulation leads to CLS [DTZ18]. Yet its computational content has not received a corresponding classification. This paper asks:

> *Can Nadler's fixed-point guarantee be captured by a natural total search problem? Conversely, when can robust global convergence of a set-valued iteration be certified as Hausdorff contraction after a change of metric?*

A satisfactory formulation must provide succinct access to the values $F(x)$, make approximate fixedness verifiable, and return checkable witnesses when a claimed contraction or regularity property is false. It must also allow the metric to be part of the instance: the energy metrics and potential ultrametrics that reveal contraction below are not interchangeable with the ambient Euclidean metric.

The projection barrier. The most tempting representation is a nearest-point circuit $\Pi_F(x, z)$ and its diagonal selection $P(x) = \Pi_F(x, x)$. A subtlety arises immediately: Hausdorff–Lipschitz correspondence with convex values does *not* generally make P Lipschitz. Even for exact projections, the correct moving-set estimate is only square-root,

$$\|\Pi_A(z) - \Pi_B(z)\|_2^2 \le d_{\mathrm{H}}(A, B)\big(\mathrm{dist}(z, A) + \mathrm{dist}(z, B)\big),$$

so on a bounded domain a Hausdorff–Lipschitz correspondence yields a Hölder-$1/2$ diagonal projection in general, and the exponent is sharp. Projection-based formulations of computational Kakutani face the same obstacle; any proof that silently promotes the projection to a Lipschitz map loses a square root. Our formulation accommodates this geometry rather than assuming it away: we fix a rational exponent $\beta \in (0, 1]$, with $\beta = 1/2$ as the canonical convex Euclidean case, and treat a claimed β-Hölder bound for the implemented map P as a separately verifiable condition. The price is that P falls outside the Lipschitz regime of standard continuous local search; the key to recovering it is a succinct Kuhn-simplicial interpolant, which needs only $n + 1$ grid values per simplex instead of the 2^n for multilinear interpolation. At the same time, the dynamical heart of Nadler's theorem needs no continuous selector at all: for $r(x) = d(x, P(x))$, exact nearest-point responses give $r(P(x)) \le k\, r(x)$. We do not assume exactness. Instead, the contraction certificate tests this inequality, up to a small additive tolerance, at $(x, P(x), x)$: either r decreases sufficiently along P, making it a local-search potential, or the triple is itself a certificate. This works for arbitrary input circuits, and a failed test reflects a genuine contraction violation whenever the circuits' errors fit within the tolerance.

1.1 Our contributions and techniques

1. A computational formulation of Nadler's theorem. For every fixed rational $\beta \in (0, 1]$, we formulate the total search problem β-PROJECTED-NADLER. The input comprises a complete metric d, a correspondence $F\colon X \rightrightarrows X$ represented by an approximate nearest-point projection circuit $\Pi_F(x, z)$ that returns an approximate closest point to z in $F(x)$, a claimed Hausdorff contraction factor $k \in (0, 1)$, and claimed Hölder/metric regularity bounds. The diagonal selection $P(x) = \Pi_F(x, x)$ induces the residual $r(x) = d(x, P(x))$. The representation parallels computational Kakutani fixed points [PVZ23], but deliberately omits small-volume outputs, which would trivialize

singleton correspondences (Section 3). A valid solution is an approximate fixed point, a projection-based certificate that the Hausdorff contraction fails, or a succinct violation of a regularity bound. Our first results establish that

$$\beta\text{-PROJECTED-NADLER} \equiv_p \beta\text{-HÖLDER-LOCALOPT} \equiv_p \text{CONTINUOUS-LOCALOPT},$$

so β-PROJECTED-NADLER is CLS-complete (Theorem 3.2) relative to the canonical complete problem for CLS [DP11]. The problem β-HÖLDER-LOCALOPT is CONTINUOUS-LOCALOPT with its Lipschitz conditions weakened to β-Hölder conditions. We need it because P only satisfies Hölder condition in general, as discussed in the projection barrier. We show equivalence between CONTINUOUS-LOCALOPT and β-HÖLDER-LOCALOPT through a Hölder-to-Lipschitz interpolation theorem that operates in polynomial dimension, preserves succinctness of representations, and loses only polynomial precision for fixed β, which we mentioned when describing the projection barrier. For hardness, we reduce from CONTINUOUS-LOCALOPT. We encode a continuous local-search instance (f, p) as the singleton correspondence $F_f(x) = \{f(x)\}$ under a complete ultrametric whose nonzero distances track the potential values of their endpoints; a contraction violation then exposes a point where p fails to decrease sufficiently. Since this ultrametric is not topologically equivalent to the Euclidean metric, the metric must be part of the input of our computational formulation of Nadler's theorem.

2. A quantitative converse to Nadler's theorem. Classical converse results for Banach's fixed point theorem show that globally convergent single-valued iterations can be made contractive after a suitable change of metric [Mey67; DTZ18]. For correspondences, the hypothesis must cover *every* branch of the iteration at once, since convergence of one selected trajectory says nothing about the rest. We therefore study correspondences for which every admissible trajectory converges robustly, and locally uniformly, to an endpoint x^*. This is stronger than $x^* \in F(x^*)$ and is necessary because a Nadler contraction can otherwise have many fixed points. Under these hypotheses, for every contraction factor $c \in (0, 1)$ and accuracy scale $\varepsilon > 0$ there is a complete metric $D_{c,\varepsilon}$ that generates the original topology, makes F a Hausdorff c-contraction, and transfers small distances back to the original metric (Theorem 4.1). This strengthens the set-theoretic converses to Fryszkowski's problem [Com17; Luc18], reviewed in Appendix D.1. Those results characterize when some complete contracting metric exists, but the metric they produce need not generate the original topology, and closeness in it need not carry a quantitative guarantee about closeness in the original metric. These are precisely the two properties that an algorithmic convergence analysis needs, and our construction provides both, following the framework of the converse of Banach's theorem developed in [DTZ18]. The construction performs the potential-to-metric construction of our complexity result in reverse. The convergence hypothesis manufactures a *rank*, which is the depth of a point in the nested global images of F. Every admissible step lands in a deeper image, so the rank increases along every branch. We then enlarge the original metric to an equivalent one under which F is Hausdorff-nonexpansive, weight it by c^{rank}, and pass to shortest-chain (geodesic) distances. Each step now contracts by the factor c while the topology is unchanged, and the residual $\text{dist}_{D_{c,\varepsilon}}(x, F(x))$ is comparable to the distance from x^*.

3. Robust convergence of coordinate descent under arbitrary orderings. As an illustration of the converse, we study random-permutation cyclic coordinate descent on strongly convex quadratics, precisely the kind of correspondence motivated above, whose next state depends on

an "adaptive" ordering choice. One coordinate at a time may not be the right object: the one-coordinate correspondence has branches that revisit the same nonoptimal point forever, since a coordinate update fixes an entire hyperplane rather than only the minimizer. The full-epoch correspondence removes this obstruction, since every coordinate is updated once per transition, and we show that it contracts uniformly over *all* $n!$ orderings in the energy norm induced by the Hessian (Proposition B.2). This yields a pathwise convergence theorem, not merely one in expectation, valid for deterministic, adaptive, randomized, or even adversarial sequences of orderings (Corollaries B.4 and B.5). A two-dimensional example shows that the choice of metric is not incidental: the same epoch map can strictly *expand* the Euclidean norm even though every ordering converges to the unique minimizer (Appendix B). In summary, the energy metric makes the epoch correspondence a strict Hausdorff contraction, illustrating the role of non-Euclidean metrics in our quantitative converse.

4. Complexity of Large margin triplet stationarity. Finally, we show that β-PROJECTED-NADLER is not only an abstract characterization but a computational primitive for a concrete, nonconvex learning objective: finding an approximate first-order stationary point of a weighted triplet loss. The triplet loss, introduced in FaceNet [SKP15], is now one of the most prominent contrastive losses. In the large-margin regime, where the margin α is at least the embedding dimension d_{emb}, every hinge function (i.e. $x \mapsto \max\{0, x\}$) agrees on the feasible box with its underlying quadratic, so the objective is globally smooth even though it need not be convex. We reduce this problem to β-PROJECTED-NADLER (Theorem 5.1) using the singleton correspondence $F_{\mathcal{L}}(x) = \{T(x)\}$ induced by the projected-gradient update T, together with an ultrametric built from the normalized loss. This is the construction from our hardness proof, applied to a concrete objective. An approximate Nadler fixed point is then a fixed point of T, hence an exact stationary point; a contraction violation identifies a witness point with insufficient potential decrease, which a standard projected-gradient estimate converts into an approximate stationary point. The remaining regularity outputs are ruled out by explicit Lipschitz bounds on T and on the metric. Yan et al. [YLCP+25] prove CLS-hardness for exact stationarity. We obtain CLS-hardness at a prescribed accuracy $\varepsilon > 0$ by applying their reduction to approximate QUADRATICPROGRAM-KKT [FGHS25] and converting an approximate stationary point into an approximate KKT point with one projected-gradient step. This yields CLS-completeness of LARGE-MARGIN-TRIPLET-FOSP (Corollary 5.2).

1.2 Organization, acknowledgment, and AI disclosure

Section 2 introduces the notions used throughout. Section 3 formulates β-PROJECTED-NADLER, whose CLS-completeness is proved in Appendix C. Section 4 states the new converse to Nadler's theorem; Appendix D reviews the set-theoretic converses surrounding Fryszkowski's problem and gives the proofs, and Appendix B applies the converse to coordinate descent. Section 5 introduces large-margin triplet stationarity, with proofs in Appendix E. Appendix A collects additional background for Appendices B, C, and D.

Acknowledgment and AI disclosure. The authors thank Ioannis Panageas (UC Irvine) for helpful advice at an early stage of this project. We used generative-AI systems (ChatGPT 5.6 sol, ChatGPT 6-Astra, Fable 5.1, and Opus 5.5) for polishing the language and exposition, navigating related literature for Sections 4 and 5 and Appendix B, drawing figures, and auditing proofs. For mathematical work, we used ChatGPT 6-Astra and Fable 5.1 to help finalize the details based on authors'

original proofs, primarily for the proof of Theorem 4.1 in Section D. We also used ChatGPT 6-Astra to help with the algebraic and numerical computations, primarily in Section B. The Lemma D.3 and Proposition B.3 were discovered by ChatGPT 6-Astra as improvements to the authors' existing work. The authors then developed the corresponding proofs with assistance from ChatGPT 6-Astra. The authors verified and take responsibility for every mathematical statement, proof, and citation.

2 Preliminaries

Notation and computational model. For a positive integer n, let $[n] = \{1, \dots, n\}$. We write $\mathbb{R}_{\geq 0}$ for the nonnegative reals, $\langle \cdot, \cdot \rangle$ for the Euclidean inner product, $\| \cdot \|_p$ for the ℓ_p norm, and $\operatorname{diam}_d(A) = \sup_{x,y \in A} d(x,y)$ for a metric d and a set A. Unless stated otherwise, all numerical parameters are rational and represented in binary. Functions and metrics are represented by polynomial-size arithmetic circuits over rational constants and the operations $\{+, -, \times, \min, \max, >\}$, where the comparison gate returns one or zero. Structural properties such as metricity and completeness are treated as promises, while the regularity properties needed by the reductions admit explicit violation outputs. Throughout, we use polynomial-time many-one reductions: a reduction maps each source instance to a target instance in polynomial time and provides a polynomial-time decoder that maps every valid solution of the target instance to a valid solution of the source instance. We write $\mathcal{A} \leq_p \mathcal{B}$ for such a reduction and $\mathcal{A} \equiv_p \mathcal{B}$ when reductions hold in both directions.

Well-behaved arithmetic circuits. To ensure efficient exact evaluation, we use well-behaved arithmetic circuits as defined in [FGHS23, Section 3.1.3]. A multiplication gate is a *true multiplication* if neither input is a rational constant gate. Denoting S as the binary description length of a circuit C, including its rational constants, C is *well-behaved* if every directed path ending at an output contains at most $\log_2 S$ true multiplication gates. Multiplication by an explicit rational constant does not count. This syntactic restriction is checkable in polynomial time and rules out the exponential growth of encoding lengths possible under repeated squaring: on a rational input x, all intermediate values have polynomial encoding length, and $C(x)$ can be evaluated exactly in time $\operatorname{poly}(S, \operatorname{size}(x))$ [FGHS23, Lemma 3.3 and Remark 1].

Metric spaces. A metric on a set X is a function $d \colon X \times X \to \mathbb{R}_{\geq 0}$ such that, for all $x, y, z \in X$, $d(x,y) = 0$ if and only if $x = y$, $d(x,y) = d(y,x)$, and $d(x,z) \leq d(x,y) + d(y,z)$. It is an ultrametric if it satisfies the stronger triangle inequality $d(x,z) \leq \max\{d(x,y), d(y,z)\}$. A sequence (x_t) is d-Cauchy if, for every $\varepsilon > 0$, all sufficiently late pairs satisfy $d(x_s, x_t) < \varepsilon$, and (X,d) is complete if every d-Cauchy sequence converges to a point of X.

Correspondences and Hausdorff distance. A correspondence, or set-valued map, from X to Y is a map $F \colon X \rightrightarrows Y$ assigning a subset $F(x) \subseteq Y$ to every $x \in X$. We write $\mathrm{CB}(X)$ for the family of nonempty closed and bounded subsets of X and $\mathcal{K}(X)$ for the family of nonempty compact subsets. For a nonempty set $A \subseteq X$, the point-to-set distance is $\operatorname{dist}_d(z, A) = \inf_{a \in A} d(z,a)$, and for nonempty closed and bounded $A, B \subseteq X$ the Hausdorff distance induced by d is

$$H_d(A,B) = \max\left\{ \sup_{a \in A} \operatorname{dist}_d(a,B), \sup_{b \in B} \operatorname{dist}_d(b,A) \right\}.$$

We also write d_H when the underlying metric is clear. A correspondence $F\colon X \to \mathcal{K}(X)$ is Hausdorff-continuous if $d(x_t, x) \to 0$ implies $H_d(F(x_t), F(x)) \to 0$, and Hausdorff L-Lipschitz if $H_d(F(x), F(y)) \le L d(x, y)$ for all $x, y \in X$. A point x^* is a fixed point of F if $x^* \in F(x^*)$ and an endpoint if $F(x^*) = \{x^*\}$. Every endpoint is a fixed point, but not conversely.

Contractions and Nadler's theorem. A correspondence $F\colon X \rightrightarrows X$ is a Hausdorff k-contraction if $k \in [0, 1)$ and $H_d(F(x), F(y)) \le k\, d(x, y)$ for all $x, y \in X$.

Theorem 2.1 (Nadler's fixed-point theorem [Nad69])**.** *Let (X, d) be complete and let $F\colon X \to \mathrm{CB}(X)$ be a Hausdorff k-contraction for some $k < 1$. Then F has a fixed point.*

Unlike the single-valued Banach theorem, Nadler's theorem does not in general guarantee uniqueness. This distinction motivates the endpoint assumption in the converse theorem of Section 4.

Basic iteration procedure for correspondences. For a continuous self-map $f\colon X \to X$, the iteration $x_{t+1} = f(x_t)$ generates a unique orbit, and any convergent orbit has a fixed-point limit. For a correspondence $F\colon X \rightrightarrows X$, each step instead permits an arbitrary choice $x_{t+1} \in F(x_t)$, so an initial point generates a family of admissible trajectories, whose exact-time reachable sets are $F^{[0]}(x) = \{x\}$ and $F^{[t+1]}(x) = \bigcup_{y \in F^{[t]}(x)} F(y)$. If F has a closed graph, the limit of any convergent admissible trajectory is a fixed point, but this says nothing about the other branches, so the natural notion of global convergence requires $H_d(F^{[t]}(x), \{x^*\}) \to 0$ for every initial point x. Moreover, $x^* \in F(x^*)$ does not make x^* absorbing. This requires the stronger condition that $F(x^*) = \{x^*\}$. A selection such as $P(x) = \Pi_F(x, x)$ restores a deterministic iteration, but describes only one branch of the full correspondence.

Smoothness, projections, and stationarity. Let $K \subseteq \mathbb{R}^q$ be nonempty, closed, and convex. The Euclidean projection $\Pi_K(z) = \arg\min_{u \in K} \|z - u\|_2$ is single-valued and nonexpansive. A differentiable function $\mathcal{L}\colon K \to \mathbb{R}$ is Λ-smooth if $\|\nabla\mathcal{L}(x) - \nabla\mathcal{L}(y)\|_2 \le \Lambda\|x - y\|_2$ for all $x, y \in K$, and its projected-gradient update with step size $\eta > 0$ is $T_\eta(z) = \Pi_K(z - \eta\nabla\mathcal{L}(z))$. The normal cone of K at $z \in K$ is $N_K(z) = \{n \in \mathbb{R}^q : \langle n, u - z\rangle \le 0 \text{ for every } u \in K\}$, and projection optimality gives $y = \Pi_K(x)$ if and only if $x - y \in N_K(y)$. A point $z \in K$ is an ε-first-order stationary point (ε-FOSP) of $\mathcal{L}$ over K if

$$\langle u - z, \nabla\mathcal{L}(z)\rangle \ge -\varepsilon \qquad \text{for every } u \in K.$$

For $\varepsilon = 0$, this is equivalent to $0 \in \nabla\mathcal{L}(z) + N_K(z)$; in particular, a fixed point of the projected-gradient map $z = \Pi_K(z - \eta\nabla\mathcal{L}(z))$ is an exact first-order stationary point.

3 Nadler's Problem and Its Complexity

In this section, we formulate a computational version of Nadler's theorem and characterize its complexity. Each value of the correspondence is represented by a projection circuit, paralleling the projection-based formulation of computational Kakutani fixed points [PVZ23], but with the Hausdorff-Lipschitz condition replaced by the strict contraction required by Nadler's theorem. We intentionally omit the small-volume output used in weak-separation-oracle formulations of Kakutani [PVZ23]: a singleton correspondence $F(x) = \{f(x)\}$ has $\mathrm{vol}(F(x)) = 0$ everywhere, so such an output would solve every singleton instance trivially and invalidate the singleton

construction in our hardness reduction. Our main result is a polynomial-time equivalence between the computational formulation of Nadler's theorem and CONTINUOUS-LOCALOPT, a standard complete problem for CLS. We take $X = [0,1]^n$ as the domain; this choice is inessential, and the usual variants over rational polytopes and high-dimensional boxes do not affect our arguments.

Definition 1 (β-PROJECTED-NADLER)**.** *An instance consists of well-behaved circuits Π, d, where $\Pi(x,y) \in F(x)$ is intended as an (approximate) nearest point projection from y to $F(x)$ and $d : X \times X \to [0, D]$ is intended as the metric, and rational parameters $\varepsilon_1, \varepsilon_2, L_P, L_d, D > 0$, and $k \in (0,1)$. We promise that $0 < \varepsilon_2 < \varepsilon_1(1-k)$ and (X,d) is a complete metric space. Set $P(x) = \Pi(x,x)$ and $r(x) = d(x, P(x))$. A valid output is one of the following:*

(N1) *A point $x \in X$ such that $r(x) \le \varepsilon_1$.*

(N2) *Points $x, y, z \in X$ such that, for $u = \Pi(x,z), v = \Pi(y,u)$ or $u = \Pi(y,z), v = \Pi(x,u)$,*

$$d(u,v) > k\,d(x,y) + \varepsilon_2. \tag{3.1}$$

(N3) *Points $x, y \in X$ such that $\|P(x) - P(y)\|_1 > L_P \|x - y\|_1^{\beta}$.*

(N4) *Points $a, b, a', b' \in X$, with $a \neq b$ and $a' \neq b'$, such that*

$$|d(a,b) - d(a',b')| > L_d\big(\|a - a'\|_1 + \|b - b'\|_1\big). \tag{3.2}$$

Remark 3.1. The numerical promise $0 < \varepsilon_2 < \varepsilon_1(1-k)$ can be checked in polynomial time; therefore it can be considered part of the input format. The other promises (that d is a metric and (X,d) is complete) are what connect a solution to Nadler's theorem, but the totality of our problem does not depend on them. The only promise (other than the numerical one) used by our CLS-membership proof (Lemma C.6) is that $d(x,x) = 0$, and this can actually be avoided (see Remark C.7 for details). Conversely, the instances built in our hardness reduction satisfy every promise.

The circuits evaluate their own rational outputs exactly on rational inputs; approximation refers to their agreement with an underlying precise metric or the nearest-point projection map. For example, an exact Euclidean projection can be irrational even at a rational query and need not be representable by the allowed rational gates.

The contraction certificate. An output of type (N2) can be checked in polynomial time by evaluating Π and d at the rational witness points. At a practical level, we consider the circuit d itself as the metric and Π as an η-approximate nearest-point circuit, which means $\Pi(w,t) \in F(w)$ and $d(t, \Pi(w,t)) \le \operatorname{dist}_d(t, F(w)) + \eta$ for all $w, t \in X$. For $u = \Pi(x,z) \in F(x)$ and $v = \Pi(y,u)$ (the other orientation is symmetric),

$$H_d\big(F(x), F(y)\big) \;\ge\; \operatorname{dist}_d\big(u, F(y)\big) \;\ge\; d(u,v) - \eta \;>\; k\,d(x,y) + \varepsilon_2 - \eta,$$

so whenever $\eta \le \varepsilon_2$ an (N2) output certifies $H_d(F(x), F(y)) > k\,d(x,y)$. In the reverse direction, the same estimate shows that a genuine k-contraction admits no (N2) witness at all.

More generally, these input circuits may only approximate an underlying geometry. Let ρ be a metric on X with respect to which F has nonempty closed and bounded values, and suppose that $|d(a,b) - \rho(a,b)| \le \nu$ for all $a, b \in X$, while every response $q = \Pi(w,t)$ satisfies $\operatorname{dist}_\rho(q, F(w)) \le \tau$

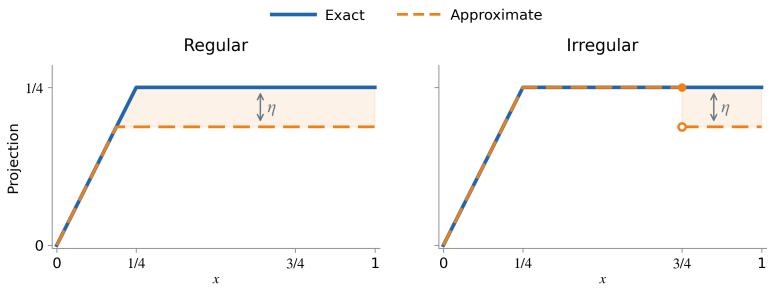


Figure 2: Regular (left) and irregular (right) projection approximations. Blue solid curves represent the exact projection, and orange dashed curves represent its approximations. Both have maximum pointwise error η, but the regular approximation is Lipschitz, whereas the irregular approximation has a jump discontinuity.

and $\rho(t,q) \le \mathrm{dist}_\rho(t, F(w)) + \eta$. The same argument, now absorbing both approximation errors, gives

$$\begin{aligned} H_\rho\big(F(x),F(y)\big) &\ge \mathrm{dist}_\rho\big(u,F(y)\big) - \tau \ge \rho(u,v) - \tau - \eta \\ &\ge d(u,v) - \tau - \eta - \nu > k\,\rho(x,y) + \varepsilon_2 - \tau - \eta - (1+k)\nu, \end{aligned}$$

so (N2) certifies a genuine violation of Hausdorff contraction as long as $\tau + \eta + (1+k)\nu \le \varepsilon_2$. The previous practical consideration is the case where $\rho = d$ and $\tau = \nu = 0$, and exact projection gives $\eta = 0$ as well. Two caveats remain. An approximate response need not satisfy $\Pi(x,u) = u$ for $u \in F(x)$, and because of the positive tolerance ε_2, the witness format does not claim to detect every strict violation of Hausdorff contraction. Moreover, these accuracy assumptions are hypotheses beyond Definition 1: without them, (N2) reflects a contraction violation only up to the additive slack $\tau + \eta + (1+k)\nu - \varepsilon_2$. Promising the bound $\tau + \eta + (1+k)\nu \le \varepsilon_2$ does not affect the complexity of the problem, since every instance produced by our reductions uses exact circuits, but it narrows the admissible inputs: together with $\varepsilon_2 < (1-k)\varepsilon_1$ it forces $\tau + \eta + (1+k)\nu < (1-k)\varepsilon_1$, so the circuits' inaccuracy must be bounded by $(1-k)\varepsilon_1$.

The regularity certificates. The outputs (N3) and (N4) expose failures of the claimed β-Hölder regularity of P and off-diagonal Lipschitz regularity of d, respectively. Beyond continuity failures, they also certify irregular approximation error of the circuits. For example, let $X = [0,1]$ and $F(w) = [0, 1/4]$, whose exact projection is $\Pi_*(w,z) = \min\{z, 1/4\}$, and for a rational $0 < \eta < 1/4$ define $\Pi(w,z) = \max\{0, \min\{z,1/4\} - \eta \mathbf{1}[w > 3/4]\}$. This circuit always returns a point in $F(w)$ and satisfies $|\Pi(w,z) - \Pi_*(w,z)| \le \eta$, but its diagonal map $P(x) = \Pi(x,x)$ jumps by η at $x = 3/4$: for $x = 3/4$ and $y = 3/4 + \eta^2/4$, we have $|P(x) - P(y)| = \eta > \eta/2 = |x-y|^{1/2}$, so (x,y) is a valid (N3) answer for $\beta = 1/2$ and $L_P = 1$, even though the exact diagonal projection is 1-Lipschitz. The regular approximation $P_{\mathrm{reg}}(x) = \min\{x, 1/4 - \eta\}$, by contrast, is 1-Lipschitz with the same maximum pointwise error η (Figure 2). These are properties of the implemented maps; an (N3) answer need not identify a defect of the underlying correspondence.

Residual decrease. The residual decrease follows from condition (N2). At a point x, test the ordered triple $(x, P(x), x)$ in the first orientation, which gives $u = P(x)$ and $v = P(P(x))$. Either this is an (N2) answer, or $r(P(x)) \le k r(x) + \varepsilon_2$, and hence

$$r(x) - r(P(x)) \ge (1-k) r(x) - \varepsilon_2.$$

On the second branch, $r(x) > \varepsilon_1$ implies a strictly positive decrease because $\varepsilon_2 < (1-k)\varepsilon_1$. The decrease is enforced by the certificate, not derived from Hausdorff contraction; when Π is exact and d is the contracting metric, contraction gives $r(P(x)) \leq k\, r(x)$ directly. As k approaches 1, the guaranteed decrease shrinks, so the same target accuracy ε_1 requires a smaller additive tolerance. This is compatible with finite-precision computation: for every rational $k < 1$, the choice $\varepsilon_2 = (1-k)\varepsilon_1/2$ has binary encoding length polynomial in those of k and ε_1, so although ε_2 tends to zero as $k \uparrow 1$, every individual instance uses a strictly positive, finitely represented tolerance.

Continuous local search. To determine the complexity of β-PROJECTED-NADLER, we recall the standard complete problem for CLS [DTZ18].

Definition 2 (CONTINUOUS-LOCALOPT)**.** *An instance consists of well-behaved arithmetic circuits $f\colon [0,1]^3 \to [0,1]^3$ and $p\colon [0,1]^3 \to [0,1]$, together with positive rational parameters δ and λ. A valid output is one of the following.*

(CO1) *A point $x \in [0,1]^3$ satisfying $p(f(x)) \geq p(x) - \delta$.*

(CO2) *Points $x, y \in [0,1]^3$ satisfying $\|f(x) - f(y)\|_1 > \lambda \|x - y\|_1$.*

(CO3) *Points $x, y \in [0,1]^3$ satisfying $|p(x) - p(y)| > \lambda \|x - y\|_1$.*

The class CLS consists of the total search problems that admit a polynomial-time many-one reduction to CONTINUOUS-LOCALOPT. As discussed in [DP11], the choice of $[0,1]^3$ and ℓ_1 norms is inessential: high-dimensional polytopes (e.g., $[0,1]^n$) and other ℓ_p norms can be used without affecting the complexity. Since the β-Hölder condition (N3) could fail to recover the Lipschitz conditions (CO2) or (CO3) in a reduction, we also define β-HÖLDER-LOCALOPT, which replaces the Lipschitz conditions of CONTINUOUS-LOCALOPT by Hölder conditions (e.g., $\|f(x) - f(y)\|_1 > \lambda \|x - y\|_1^{\beta}$), and show that it is CLS-complete for every fixed rational $\beta \in (0,1]$ by proving its polynomial-time equivalence to CONTINUOUS-LOCALOPT. A reduction from CONTINUOUS-LOCALOPT and a reduction to β-HÖLDER-LOCALOPT then yield the following theorem.

Theorem 3.2. *The problem β-PROJECTED-NADLER is CLS-complete for any fixed rational $\beta \in (0,1]$.*

Our hardness reduction follows the proof ideas of the reductions for BANACH in [DTZ18], with modifications that accommodate the different problem formulation. The details and proofs are given in Appendix C.

4 Converse Nadler's Fixed Point Theorem

Classical converses to Banach's contraction principle ask when a self-map can be made contractive by replacing the ambient metric. Fryszkowski posed the corresponding remetrization problem for set-valued maps, and Comăneci and Luchian answered it at a set-theoretic level: under an endpoint assumption, a complete bounded contracting metric exists precisely when a bounded Lyapunov-type rank function decreasing along every admissible successor exists and, equivalently, when the nested global images of the correspondence collapse to the endpoint [Jac00; Com17; Luc18]. Appendix D.1 surveys these results. However, these remetrization results do not begin

with a reference metric, and the constructed metric need neither generate a prescribed topology nor be comparable with any pre-existing distance, so convergence in the remetrized space need not translate into convergence in the metric relevant to an iterative algorithm. This section proves a stronger converse that requires both topological equivalence and quantitative error transfer, in the spirit of the single-valued result of Daskalakis, Tzamos, and Zampetakis [DTZ18]. For a correspondence, the correct analogue of global convergence must control all admissible choices, not merely one selected trajectory.

Let (X, d) be compact and let $F : X \to \mathcal{K}(X)$ be Hausdorff-continuous. Define $F^{[0]}(x) := \{x\}$ and $F^{[t+1]}(x) := \widehat{F}(F^{[t]}(x))$, where $\widehat{F}(A) := \bigcup_{x \in A} F(x)$. Thus $F^{[t]}(x)$ is the set of all states reachable after exactly t admissible updates from x. We say that F is *robustly globally convergent* to an endpoint x^* if

$$\lim_{t \to \infty} H_d\Big(F^{[t]}(x), \{x^*\}\Big) = 0 \qquad \text{for every } x \in X. \tag{4.1}$$

We additionally require local uniformity: there is an open neighborhood U of x^* such that

$$\lim_{t \to \infty} \sup_{x \in U} H_d\Big(F^{[t]}(x), \{x^*\}\Big) = 0. \tag{4.2}$$

Condition (4.1) says that every admissible trajectory converges uniformly over all choices made at each time, which is strictly stronger than the existence of one convergent selection. Under the compactness and Hausdorff-continuity assumptions above, Lemma D.3 shows that (4.1) and (4.2), together with the endpoint condition, are equivalent to *global-image collapse*, $\bigcap_{t \geq 0} \widehat{F}^t(X) = \{x^*\}$, and also to convergence that is uniform over all initial states; the proof appears in Appendix D.2. For related discussions of set-valued topological contraction and endpoints, see [TY95].

Theorem 4.1 (strong converse to Nadler's theorem)**.** *Let (X, d) be a compact metric space and let $F : X \to \mathcal{K}(X)$ be continuous with respect to the Hausdorff metric H_d. Suppose that F has an endpoint x^* and satisfies (4.1) and (4.2). Then, for every contraction factor $c \in (0, 1)$ and every accuracy scale $\varepsilon > 0$, there exists a metric $D_{c,\varepsilon}$ on X with the following properties.*

1. *The metrics $D_{c,\varepsilon}$ and d generate the same topology. In particular, $(X, D_{c,\varepsilon})$ is compact and complete.*

2. *The correspondence is a Hausdorff c-contraction:*

$$H_{D_{c,\varepsilon}}(F(x), F(y)) \leq c D_{c,\varepsilon}(x, y) \qquad \text{for all } x, y \in X. \tag{4.3}$$

3. *Approximation in the constructed metric transfers to the original metric:*

$$D_{c,\varepsilon}(x, x^*) \leq \varepsilon \implies d(x, x^*) \leq 2\varepsilon, \tag{4.4}$$

$$D_{c,\varepsilon}(x, y) \leq \varepsilon \implies \min\{d(x, y), d(x, x^*), d(y, x^*)\} \leq 2\varepsilon. \tag{4.5}$$

The proof, given in Appendix D, follows the structure of [DTZ18, Theorem 1] and Meyers' theorem [Mey67]. The theorem immediately yields convergence and residual bounds analogous to [DTZ18, Corollary 1].

Corollary 4.2 (Global convergence and residual error bound)**.** *Under the assumptions of Theorem 4.1, let $D = D_{c,\varepsilon}$ and define $r_D(x) := \operatorname{dist}_D(x, F(x))$. Then:*

1. *x^* is the unique fixed point of F.*

2. *Every admissible trajectory $x_{t+1} \in F(x_t)$ satisfies*

$$D(x_t, x^*) \le c^t D(x_0, x^*). \tag{4.6}$$

3. *The residual is equivalent, up to constants depending only on c, to the distance from the endpoint:*

$$(1-c)D(x, x^*) \le r_D(x) \le (1+c)D(x, x^*). \tag{4.7}$$

4. *Whenever $r_D(x_0) > 0$, every integer*

$$t \ge \max\left\{0, \frac{\log(r_D(x_0)/((1-c)\varepsilon))}{\log(1/c)}\right\} \tag{4.8}$$

guarantees $d(x_t, x^*) \le 2\varepsilon$.

As an example, Appendix B studies random-permutation cyclic coordinate descent (RPCD) on strongly convex quadratic objectives. With G_π denoting one complete coordinate sweep under permutation π and $\mathfrak{S}_n$ denoting the set of permutations over $[n]$, we prove directly that $\widetilde{F}(x) = \{G_\pi(x) : \pi \in \mathfrak{S}_n\}$ is a strict Hausdorff contraction in the energy metric induced by the quadratic objective. The resulting convergence bound holds pathwise for every sequence of coordinate orderings, including adaptive or adversarial choices, and on a compact invariant energy sublevel set the example also satisfies the hypotheses of Theorem 4.1.

5 The Complexity of Large-Margin Triplet Loss Minimization

In this section, we construct a reduction from large-margin triplet-loss stationarity to β-PROJECTED-NADLER over a polynomial-dimensional box. We begin with a formal introduction of the problem. Let d_{emb} be a positive integer, $V \subset \mathbb{R}^{d_{\text{emb}}}$ be a finite set of movable points, and A, B be two fixed pivots with rational embeddings $a_A, a_B \in [0,1]^{d_{\text{emb}}}$. Each movable point $v \in V$ is assigned to an embedding $z_v \in [0,1]^{d_{\text{emb}}}$, and the movable embeddings are collected into a vector $z \in K := [0,1]^q$, where $q = d_{\text{emb}}|V|$. For notational uniformity, set $z_A := a_A$ and $z_B := a_B$. These two vectors are fixed and are not optimization variables. Let $\mathcal{C}$ be a collection of ordered triplets $t = (i_t, j_t, k_t)$ over $V \cup \{A, B\}$, each specifying that the anchor i_t should be closer to j_t than to k_t. Each triplet carries a rational weight $w_t \ge 0$ and a rational margin $\alpha > 0$. Defining $[a]_+ := \max\{a, 0\}$, the weighted triplet loss is

$$\mathcal{L}(z) := \sum_{t \in \mathcal{C}} w_t \left[\|z_{i_t} - z_{j_t}\|_2^2 - \|z_{i_t} - z_{k_t}\|_2^2 + \alpha\right]_+. \tag{5.1}$$

The hinge function $[\,\cdot\,]_+$ for triplet t vanishes precisely when the squared distance from i_t to k_t exceeds the squared distance from i_t to j_t by at least α. We focus on the large-margin regime $\alpha \ge d_{\text{emb}}$, which allows us to treat the loss on the feasible box as a smooth function. Since all embeddings lie in $[0,1]^{d_{\text{emb}}}$, for every $z \in K$ and every triplet t,

$$\Delta_t(z) := \|z_{i_t} - z_{j_t}\|_2^2 - \|z_{i_t} - z_{k_t}\|_2^2 + \alpha \ge \alpha - d_{\text{emb}} \ge 0.$$

Thus $[\Delta_t(z)]_+ = \Delta_t(z)$ on K, and $\mathcal{L}$ agrees there with the quadratic polynomial $\widetilde{\mathcal{L}}(z) := \sum_{t\in\mathcal{C}} w_t \Delta_t(z)$. In the stationarity condition below, $\nabla\mathcal{L}$ is the gradient of $\widetilde{\mathcal{L}}$, including at the boundary points of K. This gradient is canonical, since any two continuously differentiable extensions agreeing with $\mathcal{L}$ on K have identical gradients on its interior and, by continuity, on its boundary. It determines the first-order variation of $\mathcal{L}$ along every feasible direction. The convention also covers $\alpha = d_{\text{emb}}$. When $\alpha = d_{\text{emb}}$, $\Delta_t(z)$ can be zero at some boundary points of K, so the hinge term $[\Delta_t(z)]_+$ is not differentiable at those points as a function on $\mathbb{R}^q$. Nevertheless, this causes no difficulty: the hinge term coincides with the polynomial Δ_t on K, and only the smoothness of $\mathcal{L}$ restricted to K is ever used.

Definition 3 (LARGE-MARGIN-TRIPLET-FOSP)**.** *An instance consists of the sets and embeddings described above, the triplet collection $\mathcal{C}$ and its rational weights, a rational margin $\alpha \geq d_{\text{emb}}$, and a rational accuracy parameter $\varepsilon > 0$. The task is to output a point $z^* \in K$ such that*

$$\langle u - z^*, \nabla\mathcal{L}(z^*)\rangle \geq -\varepsilon \qquad \text{for every } u \in K. \tag{5.2}$$

Such a point is called an ε-first-order stationary point, or ε-FOSP. For $\varepsilon = 0$, condition (5.2) is the usual variational-inequality formulation of first-order stationarity over the box K. If $V = \varnothing$, then $q = 0$ and $K = [0,1]^0$ consists of the empty tuple. Its variational-inequality gap is the empty sum, so the empty tuple is an exact FOSP.

Yan et al. prove CLS-hardness of exact FOSP computation for triplet loss minimization through a reduction from exact QUADRATICPROGRAM-KKT [YLCP+25]. That result does not, by itself, give hardness at a prescribed positive accuracy, so we instead reduce from the approximate QUADRATICPROGRAM-KKT problem, which is also CLS-complete [FGHS25]. The following recalls the definition of the approximate QUADRATICPROGRAM-KKT problem.

Definition 4 (Approximate QUADRATICPROGRAM-KKT)**.** *An instance consists of a symmetric matrix $Q \in \mathbb{Q}^{n\times n}$, a vector $b \in \mathbb{Q}^n$ and a rational accuracy parameter $\varepsilon > 0$. These coefficients define the quadratic program*

$$\min_{x\in[0,1]^n} p(x) = x^\top Q x + b^\top x. \tag{5.3}$$

The goal is to output an ε-KKT point $x^* \in [0,1]^n$ for (5.3). A point $x \in [0,1]^n$ is an ε-KKT point for (5.3) if $x_i > 0$ implies $\frac{\partial p}{\partial x_i}(x) \leq \varepsilon$ and $x_i < 1$ implies $\frac{\partial p}{\partial x_i}(x) \geq -\varepsilon$.

We state the following complexity results for LARGE-MARGIN-TRIPLET-FOSP, with proofs given in full in Appendix E.

Theorem 5.1. *There is a polynomial-time many-one reduction from* LARGE-MARGIN-TRIPLET-FOSP *to* β-PROJECTED-NADLER *with fixed parameter $\beta = 1$.*

Theorem 5.1 places LARGE-MARGIN-TRIPLET-FOSP in CLS, via the CLS-membership of β-PROJECTED-NADLER established earlier. Combined with the argument of [YLCP+25, Theorem 4.1] for CLS-hardness of the exact triplet-loss stationarity problem, this yields the following corollary.

Corollary 5.2. *The problem* LARGE-MARGIN-TRIPLET-FOSP *is* CLS*-complete.*

6 Conclusion and Further Directions

We formulated a computational version of Nadler's fixed-point theorem, showed it is CLS-complete even under the Hölder-regularity relaxation, proved a quantitative converse that certifies robust convergence of a set-valued iteration by an explicit change of metric, and instantiated both directions on random-permutation coordinate descent and large-margin triplet loss. The following questions remain open.

1. **Weak projection oracles.** Our formulation lets the projection circuit be inexact in distance, but its responses always lie in $F(x)$. Under weak separation oracle access of a convex compact domain, as for the computational formulation of Kakutani [PVZ23], a response is only guaranteed to lie in the outer γ-neighborhood $\overline{B}(F(x), \gamma) = \bigcup_{z \in F(x)} \{y : d(y, z) \leq \gamma\}$. Does the CLS-completeness of β-PROJECTED-NADLER survive when feasibility holds only up to such a tolerance γ in the weak separation oracle model?

2. **Beyond Hausdorff contractions.** Generalized, Ćirić-type, and Meir–Keeler-type multivalued contractions weaken $H_d(F(x), F(y)) \leq k\, d(x, y)$ to asymptotic or case-by-case decrease. Fryszkowski's remetrization results suggest that some generality survives set-theoretically [Com17; Luc18], but since our reduction needs *uniform* geometric decrease of the residual, it is open which weakenings preserve CLS-membership and which push toward PPAD or full FIXP.

3. **Set-valued equilibrium computation.** Best-response maps and quasi-variational inequalities with point-dependent feasible regions [CP82; CPP02; AC24; KS24] motivated correspondences in our introduction but were not pursued computationally. Under what proximal or Tikhonov-type regularization does a generalized Nash best-response correspondence become a genuine Hausdorff contraction, and is the resulting equilibrium CLS-complete via β-PROJECTED-NADLER, a contraction-based counterpart to PPAD-completeness in the convex-valued case [PVZ23]?

4. **Beyond triplet loss.** The reduction needs only two facts: large margin forces global smoothness, and one projected-gradient step supplies potential decrease. Quadruplet losses, margin-based pairwise ranking losses, and localized softmax contrastive losses [YLCP+25] share both, so the same singleton-correspondence/potential-ultrametric construction should transfer; the open task is to find the weakest margin condition for each objective.

References

[AC24] Didier Aussel and Parin Chaipunya. *Variational and quasi-variational inequalities under local reproducibility: solution concept and applications*. 2024. arXiv: 2402.02115 [math.OC]. URL: https://arxiv.org/abs/2402.02115 (cit. on pp. 3, 15).

[BD85] Michael F. Barnsley and Stephen Demko. "Iterated Function Systems and the Global Construction of Fractals". In: *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences* 399.1817 (1985), pp. 243–275 (cit. on p. 3).

[Bes59] C. Bessaga. "On the converse of Banach 'fixed-point principle'". In: *Colloquium Mathematicae* 7.1 (1959), pp. 41–43. URL: http://eudml.org/doc/210681 (cit. on p. 3).

[CHL08] Kai-Wei Chang, Cho-Jui Hsieh, and Chih-Jen Lin. "Coordinate Descent Method for Large-scale L2-loss Linear Support Vector Machines". In: *Journal of Machine Learning Research* 9.45 (2008), pp. 1369–1398. URL: https://jmlr.org/papers/v9/chang08a.html (cit. on p. 20).

[CM05] David Cheban and Cristiana Mammana. "Relation between Different Types of Global Attractors of Set-Valued Nonautonomous Dynamical Systems". In: *Set-Valued Analysis* 13.3 (2005), pp. 291–321 (cit. on p. 41).

[Com17] Andrei Comăneci. "On Fryszkowski's Problem". In: *Studia Universitatis Babeş-Bolyai Mathematica* 62.4 (2017), pp. 537–542. URL: https://doi.org/10.24193/subbmath.2017.4.10 (cit. on pp. 5, 11, 15, 40, 41).

[CP82] D. Chan and J. S. Pang. "The Generalized Quasi-Variational Inequality Problem". In: *Mathematics of Operations Research* 7.2 (1982), pp. 211–222. URL: http://www.jstor.org/stable/3689542 (cit. on pp. 3, 15).

[CPP02] E. Cavazzuti, Massimo Pappalardo, and Mauro Passacantando. "Nash Equilibria, Variational Inequalities, and Dynamical Systems". In: *Journal of Optimization Theory and Applications* 114 (Sept. 2002), pp. 491–506 (cit. on pp. 3, 15).

[Dav96] Scott Davies. "Multidimensional Triangulation and Interpolation for Reinforcement Learning". In: *Advances in Neural Information Processing Systems*. Vol. 9. MIT Press, 1996, pp. 1005–1011. URL: https://papers.nips.cc/paper/1229-multidimensional-triangulation-and-interpolation-for-reinforcement-learning (cit. on p. 31).

[DP11] Constantinos Daskalakis and Christos Papadimitriou. "Continuous local search". In: SODA '11. Society for Industrial and Applied Mathematics, 2011, pp. 790–804 (cit. on pp. 3, 5, 11).

[DTZ18] Constantinos Daskalakis, Christos Tzamos, and Manolis Zampetakis. "A Converse to Banach's Fixed Point Theorem and Its CLS-Completeness". In: *Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing*. STOC '18. Association for Computing Machinery, 2018, pp. 44–50. arXiv: 1702.07339 [cs.CC]. URL: https://doi.org/10.1145/3188745.3188968 (cit. on pp. 3–5, 11, 12, 45, 48).

[EKN92] Ludwig Elsner, Israel Koltracht, and Michael Neumann. "Convergence of sequential and asynchronous nonlinear paracontractions". In: *Numerische Mathematik* 62 (1992), pp. 305–319 (cit. on p. 29).

[ESY14] Kousha Etessami, Alistair Stewart, and Mihalis Yannakakis. "A Note on the Complexity of Comparing Succinctly Represented Integers, with an Application to Maximum Probability Parsing". In: *ACM Trans. Comput. Theory* 6.2 (May 2014). URL: https://doi.org/10.1145/2601327 (cit. on p. 36).

[FGHS23] John Fearnley, Paul W. Goldberg, Alexandros Hollender, and Rahul Savani. "The Complexity of Gradient Descent: CLS = PPAD ∩ PLS". In: *Journal of the ACM* 70.1 (2023), 7:1–7:74 (cit. on p. 7).

[FGHS25] John Fearnley, Paul W. Goldberg, Alexandros Hollender, and Rahul Savani. *The Complexity of Computing KKT Solutions of Quadratic Programs*. 2025. arXiv: 2311.13738 [cs.CC]. URL: https://arxiv.org/abs/2311.13738 (cit. on pp. 1, 6, 14, 58).

[GOVW20] Mert Gürbüzbalaban, Asuman Ozdaglar, Nuri Denizcan Vanli, and Stephen J. Wright. "Randomness and Permutations in Coordinate Descent Methods". In: *Mathematical Programming* 181.2 (2020), pp. 349–376 (cit. on p. 20).

[Hut81] John E. Hutchinson. "Fractals and Self Similarity". In: *Indiana University Mathematics Journal* 30.5 (1981), pp. 713–747 (cit. on p. 3).

[Jac00] Jacek Jachymski. "A Short Proof of the Converse to the Contraction Principle and Some Related Results". In: *Topological Methods in Nonlinear Analysis* 15.1 (2000), pp. 179–186. URL: https://doi.org/10.12775/TMNA.2000.014 (cit. on pp. 11, 40).

[KLY25] Donghwa Kim, Jaewook Lee, and Chulhee Yun. "Provable Benefit of Random Permutations over Uniform Sampling in Stochastic Coordinate Descent". In: *Proceedings of the 42nd International Conference on Machine Learning*. Vol. 267. Proceedings of Machine Learning Research. PMLR, 2025, pp. 30480–30547. URL: https://proceedings.mlr.press/v267/kim25x.html (cit. on p. 20).

[KS24] Bruce M. Kapron and Koosha Samieefar. *The Computational Complexity of Variational Inequalities and Applications in Game Theory*. 2024. arXiv: 2411.04392 [cs.CC]. URL: https://arxiv.org/abs/2411.04392 (cit. on pp. 3, 15).

[Luc18] Bogdan-Alexandru Luchian. "Revisiting Fryszkowski's Problem". In: *Journal of Fixed Point Theory and Applications* 20.3 (2018). Article 127, p. 127. URL: https://doi.org/10.1007/s11784-018-0605-7 (cit. on pp. 5, 11, 15, 41, 51).

[LW19] Ching-pei Lee and Stephen J. Wright. "Random Permutations Fix a Worst Case for Cyclic Coordinate Descent". In: *IMA Journal of Numerical Analysis* 39.3 (2019), pp. 1246–1275 (cit. on p. 20).

[Mey67] Philip R. Meyers. "A Converse to Banach's Contraction Theorem". In: *Journal of Research of the National Bureau of Standards, Section B: Mathematics and Mathematical Physics* 71B.2–3 (1967), pp. 73–76 (cit. on pp. 3, 5, 12, 45, 48).

[Nad69] Sam B. Nadler Jr. "Multi-valued Contraction Mappings". In: *Pacific Journal of Mathematics* 30.2 (1969), pp. 475–488. URL: https://doi.org/10.2140/pjm.1969.30.475 (cit. on pp. 3, 8).

[OZ17] Peter Oswald and Weiqi Zhou. "Random Reordering in SOR-Type Methods". In: *Numerische Mathematik* 135 (2017), pp. 1207–1220 (cit. on p. 20).

[PVZ23] Christos H. Papadimitriou, Emmanouil-Vasileios Vlatakis-Gkaragkounis, and Manolis Zampetakis. "The Computational Complexity of Multi-player Concave Games and Kakutani Fixed Points". In: *Proceedings of the 24th ACM Conference on Economics and Computation*. EC '23. Association for Computing Machinery, 2023, p. 1045. arXiv: 2207.07557 [cs.CC]. URL: https://doi.org/10.1145/3580507.3597812 (cit. on pp. 4, 8, 15, 53).

[SKP15] Florian Schroff, Dmitry Kalenichenko, and James Philbin. "FaceNet: A Unified Embedding for Face Recognition and Clustering". In: *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE Computer Society, June 2015, pp. 815–823. URL: https://doi.ieeecomputersociety.org/10.1109/CVPR.2015.7298682 (cit. on p. 6).

[SLY20] Ruoyu Sun, Zhi-Quan Luo, and Yinyu Ye. "On the Efficiency of Random Permutation for ADMM and Coordinate Descent". In: *Mathematics of Operations Research* 45.1 (2020), pp. 233–271 (cit. on p. 20).

[SY21] Ruoyu Sun and Yinyu Ye. "Worst-case complexity of cyclic coordinate descent: $O(n^2)$ gap with randomized version". In: *Mathematical Programming* 185.1–2 (2021), pp. 487–520 (cit. on p. 24).

[Tam18] Matthew K. Tam. "Algorithms based on unions of nonexpansive maps". In: *Optimization Letters* 12 (2018), pp. 1019–1027 (cit. on p. 21).

[TY95] E. Tarafdar and X.-Z. Yuan. "Set-valued topological contractions". In: *Applied Mathematics Letters* 8.6 (1995), pp. 79–81. URL: https://www.sciencedirect.com/science/article/pii/0893965995000899 (cit. on p. 12).

[WL20] Stephen J. Wright and Ching-Pei Lee. "Analyzing Random Permutations for Cyclic Coordinate Descent". In: *Mathematics of Computation* 89.325 (2020), pp. 2217–2248 (cit. on p. 20).

[XZ02] Jinchao Xu and Ludmil Zikatanov. "The method of alternating projections and the method of subspace corrections in Hilbert space". In: *Journal of the American Mathematical Society* 15.3 (2002), pp. 573–597 (cit. on p. 22).

[YLCP+25] Jingming Yan, Yiyuan Luo, Vaggos Chatziafratis, Ioannis Panageas, Parnian Shahkar, and Stelios Stavroulakis. *The Complexity of Finding Local Optima in Contrastive Learning*. 2025. arXiv: 2509.16898 [cs.LG]. URL: https://arxiv.org/abs/2509.16898 (cit. on pp. 1, 6, 14, 15, 58).

A Additional Preliminaries

Here, we introduce additional topological notions and concepts for later proofs.

Topological spaces. A *topology* on a set X is a collection $\tau \subseteq \mathcal{P}(X)$ such that $\varnothing, X \in \tau$, arbitrary unions of sets in τ belong to τ, and finite intersections of sets in τ belong to τ. The pair (X, τ) is called a *topological space*, and the elements of τ are called open sets. A set is closed if its complement is open. A neighborhood of $x \in X$ is a set containing an open set that contains x. For $A \subseteq X$, its closure $\overline{A}$ is the smallest closed set containing A.

A topological space is *Hausdorff* if any two distinct points admit disjoint open neighborhoods. A bijection $f\colon X \to Y$ is a *homeomorphism* if both f and f^{-1} are continuous. In this case, X and Y have the same topological structure.

Metrics and pseudometrics. A *pseudometric* on X is a function $p\colon X \times X \to \mathbb{R}_{\geq 0}$ satisfying $p(x,x) = 0$, symmetry, and the triangle inequality. Unlike a metric, a pseudometric may satisfy $p(x,y) = 0$ for distinct points $x \neq y$. Thus, a metric is precisely a pseudometric that separates distinct points.

Every metric d on X induces a topology whose open sets are unions of open balls $B_d(x,r) := \{y \in X : d(x,y) < r\}$. Two metrics d and D on the same set are *topologically equivalent* if they induce the same topology. Equivalently, the identity map $\mathrm{id}\colon (X,d) \to (X,D)$ is continuous and has a continuous inverse. That is, it is a homeomorphism.

A metric or pseudometric p is said to be continuous with respect to a given topology on X if $(x,y) \mapsto p(x,y)$ is continuous on $X \times X$ with its product topology. When X is metrized by d, the product topology is induced, for example, by

$$d_\times\big((x,y),(x',y')\big) := d(x,x') + d(y,y').$$

If $g\colon X \to (Z,\delta)$ is any map into a metric space, then $p_g(x,y) := \delta(g(x),g(y))$ is a pseudometric on X. This observation will be applied with Z equal to a hyperspace of compact sets.

Continuity and compactness. A map $f\colon X \to Y$ between topological spaces is continuous if $f^{-1}(V)$ is open in X for every open set $V \subseteq Y$. For metric spaces (X,d) and (Y,p), continuity at $x \in X$ is equivalently expressed as follows: for every $\varepsilon > 0$, there exists $\delta > 0$ such that

$$d(x,y) < \delta \quad \Longrightarrow \quad p\big(f(x),f(y)\big) < \varepsilon.$$

It is also equivalent to sequential continuity: $x_n \to x$ implies $f(x_n) \to f(x)$. The map is uniformly continuous if δ may be chosen independently of x.

A topological space is compact if every open cover has a finite subcover. We use the following standard facts. Every compact metric space is sequentially compact and complete; every continuous map from a compact metric space to a metric space is uniformly continuous; and every continuous real-valued function on a compact space attains its minimum and maximum. Moreover, a continuous bijection from a compact space to a Hausdorff space is a homeomorphism. Metric spaces are Hausdorff and regular; in particular, if $x \in U$ and U is open, then there exists an open set V such that $x \in V$ and $\overline{V} \subseteq U$.

Hausdorff distance and Hausdorff continuity. Let (Y,δ) be a metric space, and let $\mathcal{K}(Y)$ denote the family of nonempty compact subsets of Y. For $z \in Y$ and a nonempty set $A \subseteq Y$, define

$\mathrm{dist}_\delta(z, A) := \inf_{a\in A} \delta(z, a)$. For $A, B \in \mathcal{K}(Y)$, recall the notion of Hausdorff distance:

$$H_\delta(A, B) := \max\left\{\sup_{a\in A} \mathrm{dist}_\delta(a, B), \sup_{b\in B} \mathrm{dist}_\delta(b, A)\right\}.$$

The function H_δ is a metric on $\mathcal{K}(Y)$. Compactness also ensures that each point-to-set infimum above is attained. Let (X, d) be another metric space. A compact-valued correspondence $F\colon X \rightrightarrows Y$ is *Hausdorff-continuous* if the ordinary map $x \longmapsto F(x)$ from (X, d) to $(\mathcal{K}(Y), H_\delta)$ is continuous. Equivalently, whenever $x_n \to x$ in d, one has $H_\delta(F(x_n), F(x)) \to 0$. It is *Hausdorff L-Lipschitz* if $H_\delta(F(x), F(y)) \le Ld(x, y)$ for all $x, y \in X$.

For a sequence of compact-valued correspondences F_t, we say that F_t converges uniformly to F on a set $E \subseteq X$ if

$$\sup_{x\in E} H_\delta\big(F_t(x), F(x)\big) \longrightarrow 0.$$

A set $W \subseteq X$ is *forward invariant* under $F\colon X \rightrightarrows X$ if $F(x) \subseteq W$ for every $x \in W$.

B Example: Random-Permutation Cyclic Coordinate Descent

The strong converse proved in the previous section shows that robust convergence of all branches of a set-valued iteration can be certified by a suitable contracting metric. We illustrate this principle using random-permutation cyclic coordinate descent (RPCD), a standard variant of coordinate descent in which a fresh permutation of the coordinates is selected at the beginning of each epoch. Various research studies have been conducted to analyze the convergence behavior of RPCD [CHL08; OZ17; LW19; GOVW20; SLY20; WL20; KLY25]. For strongly convex quadratic objectives, this method is equivalent to applying the Gauss–Seidel method with a randomly chosen coordinate ordering.

The randomness itself will play no role in our analysis. Instead, we discard the probability distribution and retain the set of all possible outputs of one epoch. This produces a finite-valued correspondence. A trajectory of this correspondence represents coordinate descent under an arbitrary and possibly adversarial sequence of permutations.

The quadratic problem and the epoch correspondence. Let $A \in \mathbb{R}^{n\times n}$ be symmetric and positive definite, let $b \in \mathbb{R}^n$, and consider the strongly convex quadratic objective

$$\Phi(x) := \frac{1}{2}x^\top Ax - b^\top x. \tag{B.1}$$

Its unique minimizer is $x^* = A^{-1}b$. Moreover,

$$\Phi(x) - \Phi(x^*) = \frac{1}{2}(x - x^*)^\top A(x - x^*). \tag{B.2}$$

Let e_i denote the ith standard basis vector. Exact minimization of Φ along coordinate i gives the update

$$C_i(x) := x - \frac{e_i^\top(Ax - b)}{A_{ii}}e_i. \tag{B.3}$$

Indeed, the right-hand side is the unique minimizer of $\Phi(x + \tau e_i)$ over $\tau \in \mathbb{R}$.

Let $\mathfrak{S}_n$ be the set of permutations of $[n]$. For a permutation $\pi = (\pi_1, \ldots, \pi_n) \in \mathfrak{S}_n$, define the corresponding one-epoch map by

$$G_\pi := C_{\pi_n} \circ \cdots \circ C_{\pi_1}. \tag{B.4}$$

Thus, if π_t is the permutation selected at epoch t, the usual random-permutation coordinate-descent iteration can be written as

$$\begin{aligned} x^{t,0} &:= x^t, \\ x^{t,j} &:= C_{\pi_{t,j}}\big(x^{t,j-1}\big), \qquad (j \in [n]) \\ x^{t+1} &:= x^{t,n} = G_{\pi_t}(x^t), \end{aligned} \tag{B.5}$$

where $\pi_{t,j}$ denotes the jth coordinate of π_t. We associate with the algorithm the epoch correspondence

$$\widetilde{F}(x) := \{G_\pi(x) : \pi \in \mathfrak{S}_n\}. \tag{B.6}$$

The values of $\widetilde{F}$ are nonempty and finite, and hence compact. An admissible trajectory

$$x^{t+1} \in \widetilde{F}(x^t) \tag{B.7}$$

is exactly an execution of coordinate descent in which an arbitrary permutation may be selected at each epoch. This representation places the iteration within the framework of algorithms generated by finite unions of single-valued maps studied by Tam [Tam18]. Here, every permutation is admissible at each point.

Why an epoch, rather than one coordinate update, is necessary. One might instead consider the one-coordinate correspondence

$$\mathcal{R}(x) := \{C_i(x) : i \in [n]\}. \tag{B.8}$$

This correspondence does not satisfy the all-branches convergence hypothesis of the strong converse. For every coordinate i, every point satisfying

$$e_i^\top (Ax - b) = 0 \tag{B.9}$$

is fixed by the branch C_i. When $n \geq 2$, this affine hyperplane contains points other than x^*. By selecting the same coordinate repeatedly, one therefore obtains an admissible trajectory that remains forever at a nonoptimal point.

This obstruction is specific to the set-valued setting. Randomized coordinate descent may converge probabilistically under weaker coordinate-sampling conditions, but robust convergence of the full correspondence must hold for every admissible sequence of choices. Grouping the updates into complete epochs removes this obstruction because every coordinate is processed once during every transition of $\widetilde{F}$.

Euclidean distance need not contract. Although every complete epoch converges toward x^*, its map need not be contracting under every distance metric. Here we show that $\widetilde{F}$ is not contracting under Euclidean distance. Consider the two-dimensional instance

$$A = \begin{pmatrix} 1 & 2 \\ 2 & 5 \end{pmatrix}, \qquad b = 0. \tag{B.10}$$

The matrix is positive definite because its leading principal minors are positive. The minimizer is $x^* = 0$. For later use, define the linear error-update matrices

$$Q_i := I - \frac{1}{A_{ii}} e_i e_i^\top A. \tag{B.11}$$

Since $Ax^* = b$, equation (B.3) gives $C_i(x) - x^* = Q_i(x - x^*)$. For the matrix in (B.10),

$$Q_1 = \begin{pmatrix} 0 & -2 \\ 0 & 1 \end{pmatrix}, \qquad Q_2 = \begin{pmatrix} 1 & 0 \\ -\frac{2}{5} & 0 \end{pmatrix}. \tag{B.12}$$

For the ordering that updates coordinate one followed by coordinate two,

$$Q_2 Q_1 = \begin{pmatrix} 0 & -2 \\ 0 & \frac{4}{5} \end{pmatrix}. \tag{B.13}$$

Consequently,

$$\left\| G_{(1,2)}(e_2) \right\|_2 = \|Q_2 Q_1 e_2\|_2 = \left\| \begin{pmatrix} -2 \\ \frac{4}{5} \end{pmatrix} \right\|_2 = \frac{\sqrt{116}}{5} > 1 = \|e_2\|_2. \tag{B.14}$$

Because $\widetilde{F}(0) = \{0\}$, it follows that

$$H_{\|\cdot\|_2}\left(\widetilde{F}(e_2), \widetilde{F}(0)\right) > \|e_2 - 0\|_2. \tag{B.15}$$

Thus, the epoch correspondence can expand Euclidean distances even though every admissible sequence of epochs converges to the unique minimizer.

The energy metric and convergence analysis. Now we present a metric where $\widetilde{F}$ is contracting. Define the energy norm and its associated metric by

$$\|z\|_A := \sqrt{z^\top A z}, \qquad d_A(x, y) := \|x - y\|_A. \tag{B.16}$$

Since A is positive definite,

$$\sqrt{\lambda_{\min}(A)}\, \|x - y\|_2 \le d_A(x, y) \le \sqrt{\lambda_{\max}(A)}\, \|x - y\|_2. \tag{B.17}$$

Hence d_A is (metrically) equivalent to Euclidean distance, and $(\mathbb{R}^n, d_A)$ is complete. For a linear map M, write

$$\|M\|_A := \sup_{z \neq 0} \frac{\|Mz\|_A}{\|z\|_A} = \max_{\|z\|_A = 1} \|Mz\|_A. \tag{B.18}$$

For each permutation $\pi \in \mathfrak{S}_n$, define its epoch error matrix by $M_\pi := Q_{\pi_n} \cdots Q_{\pi_1}$, and define the worst-order epoch factor

$$\widetilde{\rho} := \max_{\pi \in \mathfrak{S}_n} \|M_\pi\|_A. \tag{B.19}$$

Now we prove that $\widetilde{\rho}$ is a contracting factor namely $\widetilde{\rho} \in [0, 1)$. Notice that exact coordinate minimization can be considered a special case of successive subspace correction, with each correction performed on $\operatorname{span}\{e_i\}$. Based on this viewpoint, the following energy-norm argument is consistent with the classical analysis of such methods [XZ02]. We include a direct self-contained proof for our setting.

Lemma B.1 (Uniform contraction of complete coordinate sweeps)**.** *The worst-order epoch factor satisfies*

$$0 \le \widetilde{\rho} < 1. \tag{B.20}$$

Proof. Let $\alpha = (e_i^\top Az)/A_{ii}$. For every $z \in \mathbb{R}^n$, direct expansion of (B.11) gives

$$\begin{aligned} \|Q_i z\|_A^2 &= (z - \alpha e_i)^\top A(z - \alpha e_i) \\ &= z^\top Az - 2\alpha e_i^\top Az + \alpha^2 e_i^\top Ae_i \\ &= \|z\|_A^2 - \frac{(e_i^\top Az)^2}{A_{ii}}. \end{aligned} \tag{B.21}$$

Thus, Q_i is nonexpansive in the energy norm. In fact, it is the A-orthogonal projection onto the subspace

$$\mathcal{S}_i := \left\{ z \in \mathbb{R}^n : e_i^\top Az = 0 \right\}. \tag{B.22}$$

Equality in (B.21) holds precisely when $z \in \mathcal{S}_i$, in which case $Q_i z = z$. Fix a permutation π and a nonzero vector z. Set $z_0 := z$ and $z_j := Q_{\pi_j} z_{j-1}$ for $j \in [n]$. From (B.21), we know that $\|M_\pi z\|_A \le \|z\|_A$. Thus, suppose, toward a contradiction, that

$$\|M_\pi z\|_A = \|z\|_A. \tag{B.23}$$

Since the sequence

$$\|z_0\|_A,\ \|z_1\|_A,\ \dots,\ \|z_n\|_A \tag{B.24}$$

is nonincreasing, equality between its first and last terms implies equality at every coordinate update. Equation (B.21) therefore gives

$$e_{\pi_j}^\top Az_{j-1} = 0 \qquad \text{for every } j \in [n]. \tag{B.25}$$

It also follows that $z_j = z_{j-1}$ at every step. Hence $z_j = z$ for every j, and

$$e_{\pi_j}^\top Az = 0 \qquad \text{for every } j \in [n]. \tag{B.26}$$

Because π contains every coordinate exactly once, this implies $Az = 0$. Positive definiteness of A then gives $z = 0$, a contradiction. We have shown that $\|M_\pi z\|_A < \|z\|_A$ for every nonzero z. The A-unit sphere $S_A = \{z : \|z\|_A = 1\}$ is compact, so $\|M_\pi\|_A < 1$. Finally, $\mathfrak{S}_n$ is finite, and therefore the maximum in (B.19) is also strictly smaller than one. □

Proposition B.2 (RPCD as a Nadler contraction)**.** *The correspondence $\widetilde{F}$ is a Hausdorff $\widetilde{\rho}$-contraction under the energy metric:*

$$H_{d_A}\left(\widetilde{F}(x), \widetilde{F}(y)\right) \le \widetilde{\rho} d_A(x, y) \qquad \textit{for all } x, y \in \mathbb{R}^n. \tag{B.27}$$

Moreover, $\widetilde{\rho}$ is the smallest global contraction factor for $\widetilde{F}$ under d_A.

Proof. We start from expanding the numerator of coordinate map C_i:

$$C_i(x) = x - \frac{e_i^\top Ax}{A_{ii}} e_i + \frac{e_i^\top b}{A_{ii}} e_i = Q_i x + c_i, \tag{B.28}$$

where Q_i is defined in (B.11) and $c_i = \frac{e_i^\top b}{A_{ii}} e_i$. Consequently, for any two points $u, v \in \mathbb{R}^n$,

$$C_i(u) - C_i(v) = (Q_i u + c_i) - (Q_i v + c_i) = Q_i(u - v).$$

The constant vector c_i is the same in both updates because both applications use the same coordinate i, matrix A, and vector b; hence, it cancels upon subtraction. Now apply the same permutation $\pi = (\pi_1, \ldots, \pi_n)$ to both x and y. Repeatedly applying the preceding identity yields

$$\begin{aligned} G_\pi(x) - G_\pi(y) &= C_{\pi_n} \circ \cdots \circ C_{\pi_1}(x) - C_{\pi_n} \circ \cdots \circ C_{\pi_1}(y) \\ &= Q_{\pi_n} \left(C_{\pi_{n-1}} \circ \cdots \circ C_{\pi_1}(x) - C_{\pi_{n-1}} \circ \cdots \circ C_{\pi_1}(y) \right) \\ &= \cdots \\ &= Q_{\pi_n} \cdots Q_{\pi_1}(x - y) \\ &= M_\pi (x - y). \end{aligned}$$

Consequently, for every permutation π,

$$d_A\left(G_\pi(x), G_\pi(y)\right) \le \widetilde{\rho} d_A(x, y). \tag{B.29}$$

Fix $u \in \widetilde{F}(x)$. There is a permutation π such that $u = G_\pi(x)$. Pairing u with the point $G_\pi(y) \in \widetilde{F}(y)$ gives

$$\operatorname{dist}_{d_A}\left(u, \widetilde{F}(y)\right) \le \widetilde{\rho} d_A(x, y). \tag{B.30}$$

Taking the supremum over $u \in \widetilde{F}(x)$ gives one directed Hausdorff bound. Interchanging x and y gives the other, proving (B.27). It remains to prove sharpness. Since the set of permutations is finite and the A-unit sphere is compact, there are a permutation π^* and a vector z^* satisfying

$$\|z^*\|_A = 1, \qquad \|M_{\pi^*} z^*\|_A = \widetilde{\rho}. \tag{B.31}$$

Let $x = x^* + z^*$ and $y = x^*$. Every coordinate update fixes x^*, and hence $\widetilde{F}(x^*) = \{x^*\}$. It follows that

$$H_{d_A}\left(\widetilde{F}(x), \widetilde{F}(x^*)\right) = \max_{\pi \in \mathfrak{S}_n} \|M_\pi z^*\|_A = \widetilde{\rho} = \widetilde{\rho}\, d_A(x, x^*). \tag{B.32}$$

Thus no smaller global contraction factor is possible under d_A. □

In fact, we can prove an explicit upper bound estimate for $\widetilde{\rho}$, using the framework in [SY21, Appendix A.1] for cyclic coordinate descent.

Proposition B.3 (Explicit upper bound for $\widetilde{\rho}$)**.** *Let $A \succ 0$ be the Hessian of the quadratic objective Φ, and define*

$$D_A := \operatorname{diag}(A_{11}, \ldots, A_{nn}), \quad B := D_A^{-1/2} A D_A^{-1/2}, \quad \mu := \lambda_{\min}(B), \quad S := \sum_{i<j} \frac{A_{ij}^2}{A_{ii} A_{jj}}. \tag{B.33}$$

Then the factor $\widetilde{\rho}$ satisfies

$$\widetilde{\rho} = \max_{\pi \in \mathfrak{S}_n} \|M_\pi\|_A \le q := \sqrt{\frac{S}{S + \mu}} < 1. \tag{B.34}$$

Proof. Let x^* be the unique minimizer of Φ. We will bound the decrease of the error in energy norm (or "error energy" in short):

$$\|x - x^*\|_A^2 = 2\big(\Phi(x) - \Phi(x^*)\big) \tag{B.35}$$

after every coordinate has been minimized once. The proof uses two identities: the total energy decrease equals the sum of the squared normalized coordinate changes, and the final gradient is the strictly upper triangular part of the normalized Hessian applied to those changes. See (B.55) and (B.53). Together they bound the remaining energy by a fixed multiple of the energy lost.

Let e_i denote the ith standard basis vector and let I be the $n \times n$ identity matrix. Exact minimization along coordinate i changes the current error v into $Q_i v$, where

$$v := x - x^*, \qquad Q_i := I - \frac{1}{A_{ii}} e_i e_i^\top A. \tag{B.36}$$

For a fixed ordering $\pi = (\pi_1, \dots, \pi_n)$, recall that

$$M_\pi = Q_{\pi_n} \cdots Q_{\pi_1}. \tag{B.37}$$

We analyze the updates in (B.37) after rescaling and reordering the coordinates. Define

$$P_\pi := (e_{\pi_1}, \dots, e_{\pi_n}), \qquad B_\pi := P_\pi^\top B P_\pi. \tag{B.38}$$

Multiplication by $P_\pi^\top$ places coordinate π_j in position j. The definitions in (B.33) and (B.38) give

$$B_{ii} = 1, \qquad (B_\pi)_{ii} = 1 \qquad (i = 1, \dots, n). \tag{B.39}$$

Consequently, B_π is the Hessian in coordinates where the updates occur in the order $1, \dots, n$ and every coordinate has unit curvature. For an arbitrary initial error $v \in \mathbb{R}^n$, let

$$z^{(0)} := P_\pi^\top D_A^{1/2} v, \qquad z_j^{(0)} = \sqrt{A_{\pi_j \pi_j}}\, v_{\pi_j}. \tag{B.40}$$

Thus $z^{(0)}$ represents the same initial error in the rescaled and reordered coordinates. The change of variables in (B.40) preserves the error energy:

$$\|z^{(0)}\|_{B_\pi}^2 = \|v\|_A^2. \tag{B.41}$$

Moreover, P_π is orthogonal, so (B.38) implies that B_π and B have the same eigenvalues. Hence

$$\lambda_{\min}(B_\pi) = \mu > 0. \tag{B.42}$$

Define the strictly lower triangular matrix L_π by

$$(L_\pi)_{ij} := \begin{cases} (B_\pi)_{ij}, & i > j, \\ 0, & i \le j. \end{cases} \tag{B.43}$$

By symmetry of B_π, together with (B.39) and (B.43),

$$B_\pi = I + L_\pi + L_\pi^\top. \tag{B.44}$$

The entries in row j of L_π describe how changes to the previously updated coordinates $1, \dots, j-1$ affect the derivative in coordinate j.

We now track each update explicitly. For $j = 1, \dots, n$, let $z^{(j)}$ be the transformed error after coordinates $1, \dots, j$ have been updated, and let s_j be the change made to coordinate j. At update j, we minimize the quadratic objective by changing only coordinate j. For an increment $\alpha \in \mathbb{R}$, (B.39) gives

$$\frac{1}{2}(z^{(j-1)} + \alpha e_j)^\top B_\pi (z^{(j-1)} + \alpha e_j) = \frac{1}{2}\|z^{(j-1)}\|_{B_\pi}^2 + \alpha e_j^\top B_\pi z^{(j-1)} + \frac{\alpha^2}{2}. \tag{B.45}$$

Setting the derivative of (B.45) with respect to α equal to zero gives the unique minimizing increment. Thus,

$$s_j := -e_j^\top B_\pi z^{(j-1)}, \qquad z^{(j)} := z^{(j-1)} + s_j e_j. \tag{B.46}$$

The updates in (B.46) are precisely the original coordinate updates expressed in the new coordinates. Indeed, (B.36) and (B.38) imply

$$P_\pi^\top (D_A^{1/2} Q_{\pi_j} D_A^{-1/2}) P_\pi = I - e_j e_j^\top B_\pi. \tag{B.47}$$

Combining (B.37), (B.40), and (B.47) therefore gives

$$z^{(n)} = P_\pi^\top D_A^{1/2} M_\pi v, \qquad \|z^{(n)}\|_{B_\pi}^2 = \|M_\pi v\|_A^2. \tag{B.48}$$

Now collect the coordinate changes into

$$s := (s_1, \dots, s_n)^\top. \tag{B.49}$$

Since every coordinate is updated exactly once, (B.46) and (B.49) yield

$$z^{(j-1)} = z^{(0)} + \sum_{i<j} s_i e_i, \qquad z^{(n)} = z^{(0)} + s. \tag{B.50}$$

Substituting (B.50) into (B.46) gives

$$s_j + \sum_{i<j} (B_\pi)_{ji} s_i = -(B_\pi z^{(0)})_j. \tag{B.51}$$

By (B.43), the equations in (B.51) can be written as

$$(I + L_\pi) s = -B_\pi z^{(0)}. \tag{B.52}$$

Using (B.50), (B.52), and (B.44), we obtain

$$B_\pi z^{(n)} = B_\pi z^{(0)} + B_\pi s = -(I + L_\pi) s + B_\pi s = L_\pi^\top s. \tag{B.53}$$

Equation (B.53) describes the derivative remaining after all updates. The derivative in coordinate j becomes zero when that coordinate is minimized, but subsequent changes to coordinates $j+1, \dots, n$ can make it nonzero again. Those subsequent contributions form the jth component of $L_\pi^\top s$.

Next, we relate the coordinate changes to the energy decrease. Using (B.46) and (B.39), we obtain

$$\begin{aligned} \|z^{(j)}\|_{B_\pi}^2 &= \|z^{(j-1)}\|_{B_\pi}^2 + 2 s_j e_j^\top B_\pi z^{(j-1)} + s_j^2 \\ &= \|z^{(j-1)}\|_{B_\pi}^2 - s_j^2. \end{aligned} \tag{B.54}$$

Thus (B.54) shows that s_j^2 is exactly the energy lost at update j. Summing (B.54) over $j = 1, \dots, n$ and using (B.49) yields

$$\|z^{(0)}\|_{B_\pi}^2 - \|z^{(n)}\|_{B_\pi}^2 = \sum_{j=1}^{n} s_j^2 = \|s\|_2^2, \tag{B.55}$$

where $\|\cdot\|_2$ denotes the Euclidean norm.

To bound the remaining energy, first observe that (B.43), (B.38), and (B.33) give

$$\|L_\pi\|_F^2 := \sum_{i,j} (L_\pi)_{ij}^2 = \sum_{i>j} (B_\pi)_{ij}^2 = \sum_{i<j} B_{ij}^2 = S. \tag{B.56}$$

Here $\|\cdot\|_F$ denotes the Frobenius norm. The penultimate equality in (B.56) holds because every symmetric off-diagonal pair contributes exactly one entry below the diagonal, regardless of the permutation. Thus (B.56) identifies S as the total squared size of these interactions.

Using (B.53), we have

$$\|z^{(n)}\|_{B_\pi}^2 = (L_\pi^\top s)^\top B_\pi^{-1} (L_\pi^\top s) \le \frac{1}{\mu} \|L_\pi^\top s\|_2^2 \le \frac{\|L_\pi\|_F^2}{\mu} \|s\|_2^2 = \frac{S}{\mu} \|s\|_2^2. \tag{B.57}$$

The first inequality in (B.57) follows from (B.42), the second follows from Cauchy–Schwarz, and the final equality uses (B.56). Substituting (B.55) into (B.57) gives

$$\|z^{(n)}\|_{B_\pi}^2 \le \frac{S}{\mu} \left(\|z^{(0)}\|_{B_\pi}^2 - \|z^{(n)}\|_{B_\pi}^2 \right). \tag{B.58}$$

Rearranging (B.58) and applying (B.41) and (B.48) yields

$$\|M_\pi v\|_A^2 \le \frac{S}{S+\mu} \|v\|_A^2. \tag{B.59}$$

Dividing (B.59) by $\|v\|_A^2$ and taking the supremum over nonzero initial errors gives

$$\|M_\pi\|_A^2 = \sup_{v \neq 0} \frac{\|M_\pi v\|_A^2}{\|v\|_A^2} \le \frac{S}{S+\mu}. \tag{B.60}$$

Since (B.60) holds for every permutation, taking square roots and maximizing over π proves

$$\widetilde{\rho} = \max_{\pi \in \mathfrak{S}_n} \|M_\pi\|_A \le \sqrt{\frac{S}{S+\mu}} = q < 1. \tag{B.61}$$

The strict inequality in (B.61) follows from (B.42) and the nonnegativity of the summands defining S in (B.33). In particular, if $S = 0$, then $B = I$ and one complete sweep reaches the minimizer exactly. □

Corollary B.4 (Uniform convergence over coordinate orderings). *For every $x^0 \in \mathbb{R}^n$ and every $t \ge 0$,*

$$H_{d_A}\left(\widetilde{F}^{[t]}(x^0), \{x^*\} \right) \le \widetilde{\rho}^t d_A(x^0, x^*). \tag{B.62}$$

Consequently, every trajectory $x^{t+1} = G_{\pi_t}(x^t)$, under an arbitrary sequence of coordinate orderings $(\pi_t)_{t\geq 0}$, satisfies

$$\Phi(x^t) - \Phi(x^*) \leq \widetilde{\rho}^{2t}\big(\Phi(x^0) - \Phi(x^*)\big), \tag{B.63}$$

and

$$\|x^t - x^*\|_2 \leq \sqrt{\frac{\lambda_{\max}(A)}{\lambda_{\min}(A)}}\,\widetilde{\rho}^t \|x^0 - x^*\|_2. \tag{B.64}$$

Proof. Since every coordinate update fixes the minimizer x^*, we have

$$G_\pi(x^*) = x^* \qquad \text{for every } \pi \in \mathfrak{S}_n, \tag{B.65}$$

and therefore $\widetilde{F}(x^*) = \{x^*\}$. Let $x^{t+1} = G_{\pi_t}(x^t)$ for an arbitrary sequence of permutations. Because $x^{t+1} \in \widetilde{F}(x^t)$, Proposition B.2 gives

$$d_A(x^{t+1}, x^*) \leq H_{d_A}\left(\widetilde{F}(x^t), \widetilde{F}(x^*)\right) \leq \widetilde{\rho}\, d_A(x^t, x^*). \tag{B.66}$$

Iterating (B.66) yields

$$d_A(x^t, x^*) \leq \widetilde{\rho}^t d_A(x^0, x^*). \tag{B.67}$$

Every point $y \in \widetilde{F}^{[t]}(x^0)$ is the endpoint of some admissible sequence of t coordinate orderings. Hence (B.67) holds uniformly over all such y. Since the Hausdorff distance from a nonempty set to a singleton is the largest distance to that singleton, we obtain

$$H_{d_A}\left(\widetilde{F}^{[t]}(x^0), \{x^*\}\right) = \sup_{y \in \widetilde{F}^{[t]}(x^0)} d_A(y, x^*) \leq \widetilde{\rho}^t d_A(x^0, x^*). \tag{B.68}$$

This proves (B.62). The quadratic objective satisfies

$$\Phi(x) - \Phi(x^*) = \frac{1}{2} d_A(x, x^*)^2. \tag{B.69}$$

Therefore, by (B.67), we have

$$\Phi(x^t) - \Phi(x^*) = \frac{1}{2} d_A(x^t, x^*)^2 \leq \frac{1}{2}\widetilde{\rho}^{2t} d_A(x^0, x^*)^2 = \widetilde{\rho}^{2t}\big(\Phi(x^0) - \Phi(x^*)\big), \tag{B.70}$$

which proves (B.63). Finally, positive definiteness of A gives

$$\sqrt{\lambda_{\min}(A)}\,\|z\|_2 \leq \|z\|_A \leq \sqrt{\lambda_{\max}(A)}\,\|z\|_2. \tag{B.71}$$

Combining these inequalities with (B.67), we obtain

$$\|x^t - x^*\|_2 \leq \frac{1}{\sqrt{\lambda_{\min}(A)}} d_A(x^t, x^*) \leq \frac{\widetilde{\rho}^t}{\sqrt{\lambda_{\min}(A)}} d_A(x^0, x^*) \leq \sqrt{\frac{\lambda_{\max}(A)}{\lambda_{\min}(A)}}\,\widetilde{\rho}^t \|x^0 - x^*\|_2. \tag{B.72}$$

This is precisely (B.64). □

Qualitative convergence under arbitrary sequences of complete sweeps also follows from the common-fixed-point theorem for paracontractions [EKN92, Theorem 1]. Indeed, the coordinate maps are orthogonal projections in the energy inner product, their common fixed point is x^*, and every coordinate is updated infinitely often. The estimate above additionally provides a geometric rate that is uniform over all sequences of complete sweeps.

Corollary B.5 (Residual error bound)**.** *Define the epoch residual by*

$$r_A(x) := \operatorname{dist}_{d_A}\left(x, \widetilde{F}(x)\right) = \min_{\pi \in \mathfrak{S}_n} d_A\left(x, G_\pi(x)\right). \tag{B.73}$$

Then, for every $x \in \mathbb{R}^n$,

$$(1-\widetilde{\rho})\, d_A(x, x^*) \le r_A(x) \le (1+\widetilde{\rho})\, d_A(x, x^*). \tag{B.74}$$

In particular, $r_A(x) = 0$ if and only if $x = x^*$, so x^* is the unique fixed point of $\widetilde{F}$. Moreover, every admissible trajectory satisfies

$$\|x^t - x^*\|_2 \le \frac{\widetilde{\rho}^t}{(1-\widetilde{\rho})\sqrt{\lambda_{\min}(A)}} r_A(x^0). \tag{B.75}$$

Proof. Fix $x \in \mathbb{R}^n$ and $u \in \widetilde{F}(x)$. Since $\widetilde{F}(x^*) = \{x^*\}$, Proposition B.2 implies

$$d_A(u, x^*) \le H_{d_A}\left(\widetilde{F}(x), \widetilde{F}(x^*)\right) \le \widetilde{\rho}\, d_A(x, x^*). \tag{B.76}$$

The reverse triangle inequality and (B.76) give

$$d_A(x, u) \ge d_A(x, x^*) - d_A(u, x^*) \ge (1-\widetilde{\rho})\, d_A(x, x^*). \tag{B.77}$$

Similarly, the triangle inequality gives

$$d_A(x, u) \le d_A(x, x^*) + d_A(u, x^*) \le (1+\widetilde{\rho})\, d_A(x, x^*). \tag{B.78}$$

Taking the infimum over $u \in \widetilde{F}(x)$ in (B.77) and (B.78) yields

$$(1-\widetilde{\rho})\, d_A(x, x^*) \le r_A(x) \le (1+\widetilde{\rho})\, d_A(x, x^*). \tag{B.79}$$

Because $\widetilde{\rho} < 1$, the lower bound shows that $r_A(x) = 0$ only if $x = x^*$. Conversely, $\widetilde{F}(x^*) = \{x^*\}$ implies $r_A(x^*) = 0$. Thus,

$$r_A(x) = 0 \quad \Longleftrightarrow \quad x = x^*. \tag{B.80}$$

In particular, if x is a fixed point of $\widetilde{F}$, then $r_A(x) = 0$, and hence $x = x^*$. Therefore x^* is the unique fixed point. Now let $(x^t)_{t \ge 0}$ be any admissible trajectory. By Corollary B.4,

$$d_A(x^t, x^*) \le \widetilde{\rho}^t d_A(x^0, x^*). \tag{B.81}$$

The lower residual bound at x^0 gives

$$d_A(x^0, x^*) \le \frac{r_A(x^0)}{1-\widetilde{\rho}}. \tag{B.82}$$

Consequently, we have

$$d_A(x^t, x^*) \le \frac{\widetilde{\rho}^t}{1-\widetilde{\rho}} r_A(x^0). \tag{B.83}$$

Finally, using (B.72) from Corollary B.4, we get

$$\|x^t - x^*\|_2 \le \frac{d_A(x^t, x^*)}{\sqrt{\lambda_{\min}(A)}}. \tag{B.84}$$

And we conclude that

$$\|x^t - x^*\|_2 \le \frac{\widetilde{\rho}^t}{(1-\widetilde{\rho})\sqrt{\lambda_{\min}(A)}} r_A(x^0), \tag{B.85}$$

as claimed. □

Relation to the strong converse (Theorem 4.1). For any $R > 0$, consider the ball under d_A:

$$X_R := \{x \in \mathbb{R}^n : d_A(x, x^*) \le R\}. \tag{B.86}$$

Every coordinate update never increases d_A-distance relative to x^*, so for $x \in X_R$ and arbitrary permutation π, $d_A(G_\pi(x), x^*) \le d_A(x, x^*) \le R$ and X_R is forward invariant under every branch of $\widetilde{F}$. The set X_R is compact in the Euclidean topology, and $\widetilde{F}$ is Hausdorff-continuous by Proposition B.2. Equation (B.62) gives uniform convergence of all reachable sets to the endpoint x^*, since as $t \to \infty$,

$$\sup_{x \in X_R} H_{d_A}\big(\widetilde{F}^{[t]}(x), \{x^*\}\big) \le \sup_{x \in X_R} \widetilde{\rho}^t d_A(x, x^*) \le \widetilde{\rho}^t R \longrightarrow 0.$$

Thus, the correspondence restricted to X_R satisfies the hypotheses of the strong converse.

Remark B.6. Although the epoch correspondence is motivated by random-permutation cyclic coordinate descent, the convergence result above is not probabilistic. The distribution over permutations is discarded, and every permutation is instead treated as an admissible branch of the correspondence $\widetilde{F}$. Consequently, the estimate

$$d_A(x^t, x^*) \le \widetilde{\rho}^t d_A(x^0, x^*)$$

holds pathwise for every deterministic, adaptive, or even adversarial sequence of coordinate orderings. The quantity

$$\widetilde{\rho} = \max_{\pi \in \mathfrak{S}_n} \|M_\pi\|_A$$

is a sharp one-epoch Hausdorff contraction factor of the full correspondence under the energy metric, and it yields the convergence estimate with factor $\widetilde{\rho}^{2t}$. Thus, the result should be interpreted as a certificate for robust convergence (or as a topological perspective for the convergence), rather than as an improvement of the convergence rate for RPCD. Since it maximizes over all permutations and repeatedly applies the worst one-step factor, the bound may be more conservative than previous asymptotic analysis.

C Omitted Proofs in Section 3

This section proves the reductions among CONTINUOUS-LOCALOPT, β-HÖLDER-LOCALOPT, and β-PROJECTED-NADLER. Figure 3 shows the organization of the constructions.

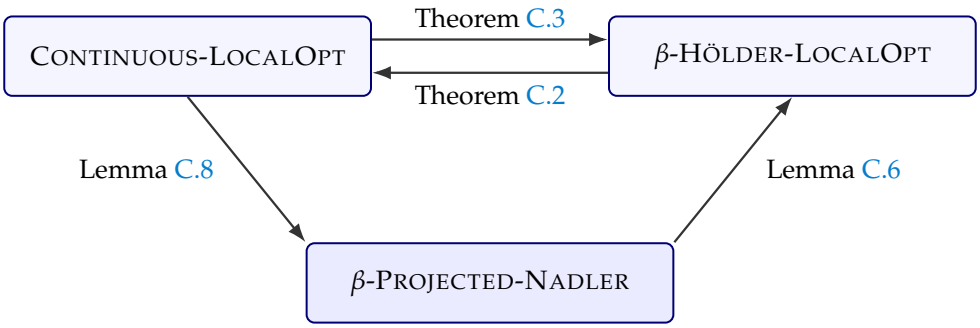


Figure 3: Polynomial-time many-one reductions among CONTINUOUS-LOCALOPT, β-PROJECTED-NADLER, and β-HÖLDER-LOCALOPT, for every fixed rational $\beta \in (0, 1]$. Each arrow points from the source problem to the target problem and is labeled by the theorem or lemma establishing the reduction.

C.1 Hölder-continuous local search

We introduce a Hölder-continuous analogue of CONTINUOUS-LOCALOPT. Throughout this subsection, fix a rational constant $\beta \in (0, 1]$. The exponent β is not part of the input. We show the three-dimensional formulation first and explicitly use their n-dimensional variants in the constructions. The norm is ℓ_1 throughout.

Definition 5 (β-HÖLDER-LOCALOPT)**.** *An instance consists of well-behaved arithmetic circuits $f\colon [0,1]^3 \to [0,1]^3$ and $p\colon [0,1]^3 \to [0,1]$, together with positive rational parameters δ and λ. A valid output is one of the following.*

(HO1) *A point $x \in [0,1]^3$ satisfying $p(f(x)) \geq p(x) - \delta$.*

(HO2) *Points $x, y \in [0,1]^3$ satisfying $\|f(x) - f(y)\|_1 > \lambda \|x - y\|_1^\beta$.*

(HO3) *Points $x, y \in [0,1]^3$ satisfying $|p(x) - p(y)| > \lambda \|x - y\|_1^\beta$.*

We claim that the problem is total. If either asserted Hölder bound fails, then an output of type (HO2) or (HO3) exists. Otherwise, suppose for contradiction that no output of type (HO1) exists. Starting from an arbitrary $x_0 \in [0,1]^3$, define $x_{t+1} := f(x_t)$. Then $p(x_{t+1}) < p(x_t) - \delta$ for every t, and hence $p(x_T) < p(x_0) - T\delta$. Choosing an integer $T > 1/\delta$ gives $p(x_T) < 0$, contradicting the fact that p takes values in $[0,1]$.

We next give the interpolation lemma used in the membership reduction. Here, we use the simplex-based interpolation method from [Dav96, Section 1.2]. However, we record the following implementation and quantitative properties because, in addition to the usual interpolation facts, our reduction requires an explicit polynomial-size arithmetic circuit and a Lipschitz bound with respect to the ℓ_1 norm. Figure 4 gives a two-dimensional example of the construction.

Lemma C.1 (Efficient simplex-based interpolation)**.** *Let $g : [0,1]^n \to [0,1]^r$ be represented by a well-behaved circuit, and let $h = 2^{-m}$ for an integer $m \geq 1$. One can construct a well-behaved circuit $I_h g : [0,1]^n \to [0,1]^r$, in time polynomial in n, r, m and the description length of g, with the following properties.*

(i) At each input x, $I_h g(x)$ is a convex combination of the values of g at $n+1$ grid vertices, each at ℓ_1-distance at most nh from x. These vertices and weights are computable within the same bound.

(ii) For all x, y, one has $\|I_h g(x) - I_h g(y)\|_1 \le (r/h)\|x - y\|_1$.

(iii) If every vertex v used at x satisfies $\|g(x) - g(v)\|_1 \le \lambda \|x - v\|_1^\beta$, then $\|g(x) - I_h g(x)\|_1 \le \lambda (nh)^\beta$.

Proof. We start by introducing the basic construction for the interpolation.

Construction and part (i). Partition $[0,1]^n$ into grid cubes of side length h. For an input x, choose a cube containing x as follows: in each coordinate, a positive grid boundary belongs to the interval immediately to its left, and zero belongs to the first interval. Write c for the lower corner of this cube and set $t_i = (x_i - c_i)/h$. Thus $t_i \in [0,1]$ records the position of x_i within its chosen interval.

Choose a permutation π satisfying $t_{\pi(1)} \ge \cdots \ge t_{\pi(n)}$, breaking ties by increasing index. Starting from c, increase one coordinate at a time by h, in this order, to obtain the vertices

$$v_0 = c, \qquad v_j = c + h \sum_{\ell=1}^{j} e_{\pi(\ell)} \quad (1 \le j \le n). \tag{C.1}$$

The e_j denotes the j-th standard basis vector. Their convex hull is the selected simplex. Define the weights

$$w_0 = 1 - t_{\pi(1)}, \qquad w_j = t_{\pi(j)} - t_{\pi(j+1)} \ (1 \le j < n), \qquad w_n = t_{\pi(n)}. \tag{C.2}$$

The ordering makes these weights nonnegative, and their sum telescopes to one. Moreover, coordinate $\pi(\ell)$ has been increased precisely at the vertices $v_\ell, \dots, v_n$, whose total weight is $\sum_{j=\ell}^{n} w_j = t_{\pi(\ell)}$. Consequently, $x = \sum_{j=0}^{n} w_j v_j$.

Set $I_h g(x) = \sum_{j=0}^{n} w_j g(v_j)$. This is a convex combination of $n+1$ values in $[0,1]^r$, so it also belongs to $[0,1]^r$. Every v_j lies in the same cube as x, and hence $\|x - v_j\|_1 \le nh$. This proves the geometric claims in part (i); the computational claim is proved below.

Continuity of $I_h g$. Within a fixed simplex, the vertices are fixed and the weights are affine functions of x. Thus the interpolation formula is affine there. It remains to check that the formulas agree when x lies on a boundary between simplices or between grid cubes.

First consider different simplices in the same cube. Their formulas can meet only where some normalized coordinates are equal. Suppose $t_{\pi(j)} = t_{\pi(j+1)}$. Exchanging these two coordinates in the order changes only v_j: all earlier vertices are unchanged, and after both coordinates have been increased, the later vertices are unchanged as well. The weight of the changed vertex is $w_j = t_{\pi(j)} - t_{\pi(j+1)} = 0$. Therefore the interpolated value is unchanged. Any two orderings of tied coordinates are related by such adjacent exchanges, so all formulas within the cube agree on their common boundaries.

Next consider two cubes sharing a face perpendicular to coordinate i. On this face, the normalized coordinate is $t_i = 1$ in the cube on the left and $t_i = 0$ in the cube on the right; the other normalized coordinates agree. By the preceding tie argument, we may place coordinate i first in the left cube's ordering and last in the right cube's ordering. In the left cube, only v_0 lies outside the shared face, and its weight is $w_0 = 1 - t_i = 0$. In the right cube, only v_n lies outside that face, and its weight is $w_n = t_i = 0$. After discarding these zero-weight vertices, both formulas start at the lower corner of the shared face and increase the remaining coordinates in exactly the same order. Their weights are also the same, since they are formed from the same ordered coordinates. Thus the two formulas agree on the shared face.

Repeating this argument handles intersections of several grid faces. The finitely many affine formulas therefore agree on all overlaps of the closed simplices, and together define a continuous function on $[0,1]^n$. Notice that this argument does not require g itself to be continuous.

Part (ii): the Lipschitz bound. We first prove the bound when x and y belong to the same closed simplex. By the boundary agreement just proved, both values can be computed using this simplex's vertices $v_0, \dots, v_n$ and coordinate order π, even if the boundary convention selects another simplex. Collecting the coefficients of the vertex values in the weighted sum gives

$$I_h g(x) = g(v_0) + \sum_{j=1}^{n} \frac{x_{\pi(j)} - c_{\pi(j)}}{h} \big(g(v_j) - g(v_{j-1})\big). \tag{C.3}$$

Each coordinate of $g(v_j) - g(v_{j-1})$ lies in $[-1,1]$, so $\|g(v_j) - g(v_{j-1})\|_1 \le r$. Subtracting the two affine formulas and applying the triangle inequality yields

$$\begin{aligned} \|I_h g(x) - I_h g(y)\|_1 &\le \frac{1}{h} \sum_{j=1}^{n} |x_{\pi(j)} - y_{\pi(j)}| \, \|g(v_j) - g(v_{j-1})\|_1 \\ &\le \frac{r}{h} \sum_{j=1}^{n} |x_{\pi(j)} - y_{\pi(j)}| = \frac{r}{h} \|x - y\|_1. \end{aligned}$$

For arbitrary x, y, subdivide the straight segment from x to y into finitely many subsegments, each contained in a closed simplex. Such a subdivision exists because there are finitely many simplices and the intersection of a segment with a convex simplex is an interval, a point, or the empty set. Write the subdivision points as $z_k = x + \theta_k(y - x)$, where $0 = \theta_0 < \cdots < \theta_M = 1$. Continuity ensures that the formula from either neighboring simplex gives the same value at a subdivision point. Applying the preceding bound to each subsegment gives

$$\begin{aligned} \|I_h g(x) - I_h g(y)\|_1 &\le \sum_{k=1}^{M} \|I_h g(z_k) - I_h g(z_{k-1})\|_1 \\ &\le \frac{r}{h} \sum_{k=1}^{M} \|z_k - z_{k-1}\|_1 = \frac{r}{h} \|x - y\|_1. \end{aligned}$$

The final equality follows from $z_k - z_{k-1} = (\theta_k - \theta_{k-1})(y - x)$ and $\sum_{k=1}^{M} (\theta_k - \theta_{k-1}) = 1$. This proves part (ii), including at simplex boundaries.

Part (iii): the approximation bound. Fix x and suppose that the stated comparisons hold at its selected vertices. Since the weights are nonnegative and sum to one,

$$\|g(x) - I_h g(x)\|_1 \le \sum_{j=0}^{n} w_j \|g(x) - g(v_j)\|_1 \le \lambda \sum_{j=0}^{n} w_j \|x - v_j\|_1^{\beta} \le \lambda (nh)^{\beta}.$$

This proves part (iii).

Polynomial-time circuit construction. Let S be the description length of the circuit for g. The circuit computes only the simplex selected by its input; it does not enumerate the grid or all possible coordinate orderings. To locate the cube, process each coordinate x_i by m successive bisections. Initialize $u_0 = x_i$ and, for $\ell = 1, \dots, m$, compute

$$b_\ell = \mathbf{1}[u_{\ell-1} > 1/2], \qquad u_\ell = 2u_{\ell-1} - b_\ell.$$

Here $\mathbf{1}[\cdot]$ denotes the indicator of the stated condition. The bit b_ℓ selects a half-interval, and u_ℓ gives the rescaled position within that half. Inductively, $u_\ell \in [0,1]$ and $x_i = \sum_{a=1}^{\ell} 2^{-a} b_a + 2^{-\ell} u_\ell$. Thus $c_i = \sum_{\ell=1}^{m} 2^{-\ell} b_\ell$ and $t_i = u_m$. The strict comparison assigns a midpoint to its left half, exactly implementing the chosen grid-boundary convention. This stage uses $O(nm)$ gates.

To determine the coordinate order, compute the rank of each t_i:

$$\rho_i = 1 + \sum_{k<i} \mathbf{1}[t_k \geq t_i] + \sum_{k>i} \mathbf{1}[t_k > t_i].$$

This counts the coordinates preceding i in decreasing order, with smaller indices preceding larger ones in a tie. Hence the ranks are exactly $1, \ldots, n$. All ranks can be computed using $O(n^2)$ gates. Non-strict comparisons and equality tests are also available using a constant number of strict comparisons and arithmetic gates. The sorted coordinates and the vertices can now be obtained without explicitly outputting the permutation:

$$t_{\pi(j)} = \sum_{i=1}^{n} t_i \mathbf{1}[\rho_i = j], \qquad (v_j)_i = c_i + h\mathbf{1}[\rho_i \leq j].$$

The first formula applies for $1 \leq j \leq n$, and the second for $0 \leq j \leq n$. These formulas and (C.2) compute all vertices and weights using $O(n^2)$ additional gates. Finally, evaluate $n+1$ copies of g in parallel at these vertices and form the weighted sum, coordinate by coordinate. The resulting circuit has $O(nm + n^2 + (n+1)S + nr)$ gates. All new rational constants have $O(m + \log n)$-bit encodings, so its description and construction time are polynomial in n, r, m, S. In particular, the cost depends on m, but not on the 2^{mn} grid cubes.

For completeness, consider the well-behavedness condition. Cube selection and rank computation use no true multiplications. Selecting the sorted coordinates adds one layer of true multiplications, and forming the weighted sum adds one more. Since the copies of g are parallel, the true-multiplication depth is at most that of g plus two. Under the logarithmic true-multiplication depth convention for well-behaved circuits, padding the circuit by a constant factor with redundant additions of zero therefore preserves well-behavedness and the polynomial size bound. For a rational input of encoding length B, the vertices and weights have encoding length $\mathrm{poly}(n, m, B)$. The well-behaved copies of g can therefore be evaluated exactly, and the entire computation takes time in $\mathrm{poly}(n, r, m, S, B)$. This proves the computational claim in part (i) and completes the construction. □

Now we prove the CLS-completeness of β-HÖLDER-LOCALOPT. In the proof, we assume both CONTINUOUS-LOCALOPT and β-HÖLDER-LOCALOPT are in the polynomial dimensional box $[0,1]^n$ for generality.

Theorem C.2. *For every fixed rational $\beta \in (0,1]$, there is a polynomial-time reduction from β-HÖLDER-LOCALOPT to CONTINUOUS-LOCALOPT.*

Proof. Let (f, p, δ, λ) be an instance of β-HÖLDER-LOCALOPT on $[0,1]^n$, with f and p represented by well-behaved circuits. If $\delta \geq 1$, the point 0 satisfies (HO1), since $p(f(0)) \geq 0 \geq p(0) - \delta$. Hence assume $0 < \delta < 1$.

Let $\bar{\lambda} = \max\{1, \lambda\}$ and choose a grid length $h = 2^{-m}$, with $m \geq 1$, such that $nh \leq 1$ and

$$3\bar{\lambda}^2 (nh)^{\beta^2} \leq \delta/2. \tag{C.4}$$

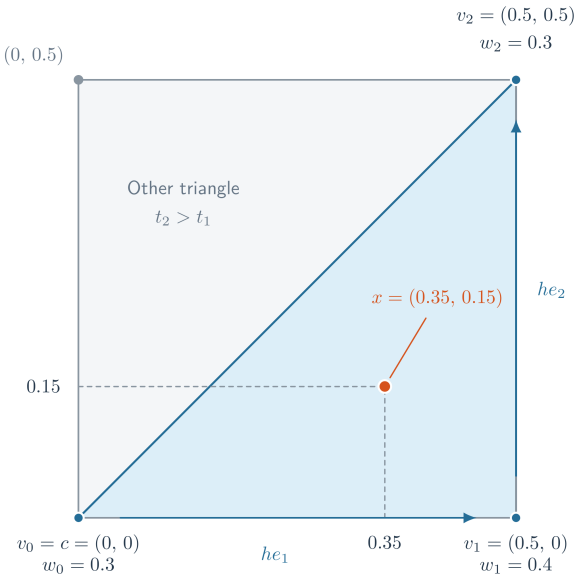


Figure 4: Simplex-based interpolation in two dimensions. A grid square with lower corner $c = (0,0)$ and side length $h = 1/2$ is divided into two triangles. The orange point $x = (0.35, 0.15)$ lies in the blue triangle with vertices $v_0 = c$, $v_1 = c + he_1$, and $v_2 = c + h(e_1 + e_2)$, where e_1, e_2 are the standard basis vectors. Its normalized coordinates are $(t_1, t_2) = (x - c)/h = (0.7, 0.3)$, giving weights $w_0 = 1 - t_1 = 0.3$, $w_1 = t_1 - t_2 = 0.4$, and $w_2 = t_2 = 0.3$. These weights express $x = 0.3v_0 + 0.4v_1 + 0.3v_2$ and define the interpolated value $I_h g(x) = 0.3g(v_0) + 0.4g(v_1) + 0.3g(v_2)$.

The second inequality reserves an error allowance of $\delta/2$ for transferring a solution from the interpolated functions back to f and p. We verify below that this grid length can be chosen in polynomial time.

Set $\widetilde{f} = I_h f$ and $\widetilde{p} = I_h p$. By Lemma C.1, these functions have the required ranges, and their Lipschitz constants are at most n/h and $1/h$, respectively. Construct the CONTINUOUS-LOCALOPT instance $(\widetilde{f}, \widetilde{p}, \delta/2, n/h)$. Neither (CO2) nor (CO3) is a valid answer to this instance, because both interpolants satisfy the stated Lipschitz bound, regardless of the regularity of f and p. Thus every target answer is a point x satisfying $\widetilde{p}(\widetilde{f}(x)) \geq \widetilde{p}(x) - \delta/2$. Fix such an x and write $z = \widetilde{f}(x)$.

For a point u, let $V(u)$ denote the $n + 1$ grid vertices selected by the interpolation construction at u. The vertices depend only on u and the grid, not on the function being interpolated. We check the following inequalities:

1. $\|f(x) - f(v)\|_1 \leq \lambda \|x - v\|_1^\beta$ for every $v \in V(x)$,
2. $|p(x) - p(v)| \leq \lambda \|x - v\|_1^\beta$ for every $v \in V(x)$,
3. $|p(z) - p(v)| \leq \lambda \|z - v\|_1^\beta$ for every $v \in V(z)$,
4. $|p(f(x)) - p(z)| \leq \lambda \|f(x) - z\|_1^\beta$.

If the first comparison fails, return (x, v) as an (HO2) answer. Any other failed comparison gives an (HO3) answer, using the corresponding pair (x, v), (z, v), or $(f(x), z)$.

The last comparison is needed in addition to the vertex comparisons: it controls the change in the original potential p when its argument is changed from $f(x)$ to the interpolated update z. We now assume that every comparison passes and show that x satisfies (HO1). Importantly, we assume only these finitely many inequalities, not a global Hölder condition.

Let $\eta = \lambda(nh)^\beta$. By Lemma C.1(iii), the comparisons for f give $\|f(x) - z\|_1 \le \eta$. By the same lemma, the second family gives $|p(x) - \widetilde{p}(x)| \le \eta$, and the third gives $|p(z) - \widetilde{p}(z)| \le \eta$. Finally, the last comparison gives $|p(f(x)) - p(z)| \le \lambda\eta^\beta$. These are the three potential errors that must be accounted for: interpolation at x, interpolation at z, and changing the update from $f(x)$ to z.

Because $nh \le 1$ and $\beta^2 \le \beta$, we have $(nh)^\beta \le (nh)^{\beta^2}$. Also, $\lambda \le \bar{\lambda}^2$ and $\lambda^{1+\beta} \le \bar{\lambda}^2$, since $\bar{\lambda} \ge \max\{1, \lambda\}$ and $1 + \beta \le 2$. Consequently, the total error satisfies

$$\begin{aligned} &|p(x) - \widetilde{p}(x)| + |p(z) - \widetilde{p}(z)| + |p(f(x)) - p(z)| \\ &\quad \le 2\eta + \lambda\eta^\beta = 2\lambda(nh)^\beta + \lambda^{1+\beta}(nh)^{\beta^2} \le 3\bar{\lambda}^2(nh)^{\beta^2} \le \delta/2. \end{aligned} \tag{C.5}$$

We can therefore transfer the target inequality back to the original functions one term at a time:

$$\begin{aligned} p(f(x)) &\ge p(z) - \lambda\eta^\beta \\ &\ge \widetilde{p}(z) - \eta - \lambda\eta^\beta \\ &\ge \widetilde{p}(x) - \delta/2 - \eta - \lambda\eta^\beta \\ &\ge p(x) - \delta/2 - 2\eta - \lambda\eta^\beta \\ &\ge p(x) - \delta. \end{aligned}$$

The third line uses the target answer and $z = \widetilde{f}(x)$; the last line uses (C.5). Thus x is an (HO1) answer.

Now we show that the grid length can be chosen in polynomial time. Write the fixed exponent as $\beta = a/b$, where a, b are fixed positive integers. A grid length satisfying (C.4) can be found by increasing m until both inequalities hold. This requires only $m = O_\beta(\log n + \log\bar{\lambda} + \log(1/\delta))$. Inequality (C.4) is equivalent to the rational comparison

$$(3\bar{\lambda}^2)^{b^2}(n2^{-m})^{a^2} \le (\delta/2)^{b^2}.$$

Its exponents are constants, so it can be checked exactly in polynomial time by [ESY14, Proposition 3.7]. Lemma C.1 then constructs both interpolant circuits in polynomial time. The parameter $n/h = n2^m$ also has polynomial encoding length, namely $O(m + \log n)$ bits.

The decoder uses only the vertices in $V(x)$ and $V(z)$, hence $O(n)$ circuit evaluations and comparisons; it never searches the whole grid. For any nonnegative rational numbers A, B, a comparison $A \le \lambda B^\beta$ is equivalent to $A^b \le \lambda^b B^a$, so all the Hölder tests can likewise be performed exactly using fixed integer powers. Well-behavedness and Lemma C.1 ensure that $f(x)$, z, the selected vertices, and all evaluated values have encoding lengths polynomial in the instance size and the encoding length of x. Thus the instance construction and the solution decoder are both polynomial-time. Every target solution yields an answer of type (HO1), (HO2), or (HO3), proving membership in CLS. □

Theorem C.3. *For every fixed rational $\beta \in (0, 1]$, there is a polynomial-time reduction from* CONTINUOUS-LOCALOPT *to* β-HÖLDER-LOCALOPT.

Proof. Let $\mathcal{I}_{\mathrm{CO}} = (f, p, \delta, \lambda)$ be an instance of CONTINUOUS-LOCALOPT. We construct the β-HÖLDER-LOCALOPT instance $\mathcal{I}_{\mathrm{HO}} := (f, p, \delta, n\lambda)$. This construction clearly has polynomial size. An output of type (HO1) is exactly an output of type (CO1). Now suppose that (x, y) is an output of type (HO2), and set $t := \|x - y\|_1$. Since $x, y \in [0,1]^n$, one has $0 \le t \le n$. For $t > 0$, $t^{1-\beta} \le n^{1-\beta} \le n$, and hence $t \le nt^{\beta}$. Therefore,

$$\|f(x) - f(y)\|_1 > n\lambda t^{\beta} \ge \lambda t,$$

so (x, y) is an output of type (CO2). The case $t = 0$ cannot satisfy the strict (HO2) inequality. The same argument shows that every output of type (HO3) is an output of type (CO3).

Thus every solution of the constructed β-HÖLDER-LOCALOPT instance can be decoded in polynomial time into a solution of the original CONTINUOUS-LOCALOPT instance. Since CONTINUOUS-LOCALOPT is CLS-complete, β-HÖLDER-LOCALOPT is CLS-hard. □

By Theorem C.2 and C.3, we can conclude the CLS-completeness of β-HÖLDER-LOCALOPT:

Theorem C.4. *For every fixed rational $\beta \in (0, 1]$, the problem β-HÖLDER-LOCALOPT is CLS-complete.*

Remark C.5 (The exponent as part of the input). Our reductions fix β independently of the input. The same proof allows β to be supplied with the instance provided that $1/\beta$ and both its numerator and denominator are polynomially bounded in the input length. If a binary-encoded exponent is allowed to be exponentially small, the grid depth m required by Equation (C.4) may be exponential, so the proof above would no longer establish a polynomial-time reduction.

C.2 The CLS-membership of β-PROJECTED-NADLER

Again, the reduction in this (and the following) section works for any polynomial dimension n.

Lemma C.6. *For every fixed rational $\beta \in (0, 1]$, β-PROJECTED-NADLER admits a polynomial-time reduction to β-HÖLDER-LOCALOPT.*

Proof. Consider a well-formed instance of β-PROJECTED-NADLER. Let $\Delta = (1 - k)\varepsilon_1 - \varepsilon_2 > 0$ and set

$$f_H(x) = P(x), \qquad p_H(x) = \frac{r(x)}{D}, \qquad \delta_H = \frac{\Delta}{D}, \qquad \lambda_H = \max\left\{L_P, \frac{L_d(n + L_P)}{D}\right\}. \tag{C.6}$$

The construction ensures $f_H(X) \subseteq X$ and $p_H(X) \subseteq [0,1]$. A constant number of circuit compositions is used, so the circuits remain well-behaved after polynomial padding. All parameters have polynomial rational encoding length. We decode each possible answer.

Output (HO1). Suppose that $p_H(P(x)) \ge p_H(x) - \delta_H$. If $r(x) \le \varepsilon_1$, return x as an (N1) answer. Otherwise, since $p_H = r/D$ and $k < 1$, we obtain

$$r(P(x)) \ge r(x) - D\delta_H = k\, r(x) + (1 - k)(r(x) - \varepsilon_1) + \varepsilon_2 > k\, r(x) + \varepsilon_2.$$

This inequality gives an (N2) witness: choose $y := P(x)$ and $z := x$. Then $u = \Pi(x, z) = y$ and $v = \Pi(y, u) = P(y)$, so $d(u, v) = r(P(x)) > k\, d(x, y) + \varepsilon_2$. Thus (x, y, z) is a valid (N2) answer.

Output (HO2). Since $\lambda_H \ge L_P$, a Hölder violation for $f_H = P$ immediately gives (N3).

Output (HO3). Suppose $|p_H(x) - p_H(y)| > \lambda_H \|x - y\|_1^\beta$. If $x = P(x)$, the definition of residual gives $r(x) = d(x, x) = 0$, so return x as (N1). The same holds if $y = P(y)$ (returning y as (N1)). Otherwise, both $(x, P(x))$ and $(y, P(y))$ are off-diagonal. Let $t = \|x - y\|_1 > 0$. If $\|P(x) - P(y)\|_1 > L_P t^\beta$, then we obtain a (N3) answer (x, y). If not, take $t \le n$ and $t \le n t^\beta$ to obtain

$$\begin{aligned} |d(x, P(x)) - d(y, P(y))| &= D|p_H(x) - p_H(y)| \\ &> D\lambda_H t^\beta \ge L_d(n + L_P)t^\beta \\ &\ge L_d\big(t + \|P(x) - P(y)\|_1\big). \end{aligned} \tag{C.7}$$

Thus $(x, P(x), y, P(y))$ is an (N4) answer.

Clearly, every part of the reduction is computable in polynomial time and respects the polynomial witness size. Together with Theorem C.4, this proves the CLS-membership. □

Remark C.7 (Bypassing the metric promise). The proof above uses the promise that d is a metric only through $d(x, x) = 0$ in the (HO3) decoder. Without it, suppose $x = P(x)$. If $r(x) \le \varepsilon_1$, then x is an (N1) answer. Otherwise, take all three points of (N2) equal to x. In this case, $u = \Pi(x, x) = x$ and $v = \Pi(x, u) = x$, so $d(u, v) = d(x, x) = r(x)$, and $(1 - k)\, r(x) > (1 - k)\, \varepsilon_1 > \varepsilon_2$ give $d(u, v) > k\, d(x, x) + \varepsilon_2$. Hence, (x, x, x) is an (N2) answer. The same applies to y. No other step uses symmetry, the triangle inequality, metric space completeness, or $\Pi(x, y) \in F(x)$. The ranges $P(X) \subseteq X$ and $d(X, X) \subseteq [0, D]$ can be enforced syntactically by coordinate-wise clipping with min and max gates, which preserves well-behavedness and 1-Lipschitzness. Therefore, subject to the numerical condition $0 < \varepsilon_2 < (1 - k)\varepsilon_1$ and the (enforced) range conditions $P(X) \subseteq X$ and $d(X, X) \subseteq [0, D]$, the membership reduction and the totality of β-PROJECTED-NADLER hold for arbitrary well-behaved circuits.

C.3 The CLS-hardness of β-PROJECTED-NADLER

For the reverse reduction, we encode the local-search potential directly into the metric and use the update circuit as a singleton correspondence.

Lemma C.8. *For every fixed rational $\beta \in (0, 1]$,* CONTINUOUS-LOCALOPT *admits a polynomial-time reduction to* β-PROJECTED-NADLER.

Proof. Let (f, p, δ, λ) be a CONTINUOUS-LOCALOPT instance on $X = [0, 1]^n$. If $\delta \ge 1$, use a fixed target instance of the singleton form below and decode every answer as the origin. Assume $0 < \delta < 1$. Use the following ultrametric construction:

$$d_p(x, y) = \begin{cases} 0, & x = y, \\ 1 + \max\{p(x), p(y)\}, & x \ne y. \end{cases} \tag{C.8}$$

Set $d = d_p$ and $\Pi(x, z) = f(x)$, independently of z, and choose

$$D = 2, \qquad k = 1 - \frac{\delta}{2}, \qquad \varepsilon_1 = \frac{1}{2}, \qquad \varepsilon_2 = \frac{\delta}{8}, \qquad L_P = n\lambda, \qquad L_d = \lambda. \tag{C.9}$$

In particular, $0 < \varepsilon_2 = \delta/8 < \delta/4 = (1 - k)\varepsilon_1$. All circuits and parameters have polynomial description length.

For completeness, d_p is a genuine complete ultrametric. For pairwise distinct x, y, z, the maximum of $d_p(x,z)$ and $d_p(z,y)$ is at least both $1 + p(x)$ and $1 + p(y)$, hence at least $d_p(x,y)$. The cases with repeated points are immediate. Every nonzero distance is within $[1,2]$, so every d_p-Cauchy sequence is eventually constant. Consequently, the singleton correspondence $F_f(x) = \{f(x)\}$ has nonempty closed bounded values, and Π is its exact projection in d_p. The contraction parameter remains a claim that can be violated. We now decode all four output types.

Output (N1). Since $P = f$ and every nonzero d_p-distance is at least one, $d_p(x, f(x)) \le 1/2$ implies $x = f(x)$. Thus $p(f(x)) = p(x)$ and x satisfies (CO1).

Output (N2). The projection circuit ignores its second argument, so the answer gives $d_p(f(x), f(y)) > k d_p(x,y) + \varepsilon_2$. In particular, $x \neq y$ and $f(x) \neq f(y)$. If neither endpoint satisfied (CO1), then $p(f(x)) < p(x) - \delta$ and $p(f(y)) < p(y) - \delta$, which yields

$$\begin{aligned} d_p(f(x), f(y)) &= 1 + \max\{p(f(x)), p(f(y))\} \\ &< 1 + \max\{p(x), p(y)\} - \delta \\ &= d_p(x,y) - \delta \\ &\le (1 - \delta/2) d_p(x,y) = k d_p(x,y). \end{aligned} \tag{C.10}$$

The last inequality uses $d_p(x,y) \le 2$. This contradicts the buffered certificate. Evaluate the two source potential inequalities and return an endpoint satisfying (CO1).

Output (N3). Let $t = \|x - y\|_1 \in (0, n]$; then we have $t \le n t^\beta$. Therefore $\|f(x) - f(y)\|_1 > n\lambda t^\beta \ge \lambda t$, so the same pair (x,y) satisfies (CO2).

Output (N4). Suppose that the returned points satisfy $a \neq b$, $a' \neq b'$, and

$$|d_p(a,b) - d_p(a',b')| > \lambda\big(\|a - a'\|_1 + \|b - b'\|_1\big).$$

Since the two arguments in each metric evaluation are distinct, both evaluations use the formula $d_p(u,v) = 1 + \max\{p(u), p(v)\}$. The common additive constant 1 therefore cancels upon subtraction, giving

$$\begin{aligned} |d_p(a,b) - d_p(a',b')| &= \big|\max\{p(a), p(b)\} - \max\{p(a'), p(b')\}\big| \\ &\le \max\big\{|p(a) - p(a')|,\ |p(b) - p(b')|\big\}. \end{aligned}$$

Here we use the fact that the change in a maximum is at most the larger change in its two arguments.

We claim that at least one of the pairs (a, a') and (b, b') violates the λ-Lipschitz bound for p. Otherwise, the preceding estimate would imply

$$|d_p(a,b) - d_p(a',b')| \le \lambda \max\{\|a - a'\|_1, \|b - b'\|_1\} \le \lambda\big(\|a - a'\|_1 + \|b - b'\|_1\big),$$

contradicting the returned (N4) certificate. The decoder, therefore, checks these two pairs and returns one satisfying $|p(u) - p(v)| > \lambda\|u - v\|_1$ as an (CO3) answer.

Every decoder uses polynomially many rational operations and respects the polynomial witness size. This proves the reduction. □

Now, the proof of Theorem 3.2 naturally follows from Lemma C.8 and C.6.

D Background and Omitted Proofs for Section 4

D.1 Fryszkowski's problem and set-theoretic remetrization

Classical converses to Banach's contraction principle ask when a self-map can be made contractive by replacing the ambient metric. Fryszkowski proposed the corresponding remetrization problem for set-valued maps, as recorded by Jachymski [Jac00]. For a nonempty set Ω, write $2^{\Omega}_{*} := 2^{\Omega} \setminus \{\emptyset\}$. Given a correspondence $F : \Omega \to 2^{\Omega}_{*}$, define its induced action on nonempty subsets by $\widehat{F}(A) := \bigcup_{x \in A} F(x)$, and write $\widehat{F}^{n}$ for its n-fold iterate.

Definition 6 (Fryszkowski's problem)**.** *Given $\alpha \in (0,1)$, a nonempty set Ω, and a correspondence $F : \Omega \to 2^{\Omega}_{*}$, determine necessary and/or sufficient conditions for the existence of a complete metric d on Ω under which F is a Nadler α-contraction:*

$$H_d(F(x), F(y)) \leq \alpha d(x, y) \qquad \text{for all } x, y \in \Omega. \tag{D.1}$$

Here H_d denotes the generalized Hausdorff distance, which may take the value $+\infty$ on arbitrary subsets. This is a set-theoretic remetrization problem: no metric or topology on Ω is fixed in advance, and the goal is to identify conditions on the dynamics of F that permit some complete contracting metric.

The results below concern correspondences with a distinguished *endpoint* z, meaning that $F(z) = \{z\}$. This is stronger than the usual fixed-point condition $z \in F(z)$. The stronger condition is natural in a converse theorem that requires every branch of the correspondence to move toward the same terminal state: once the process reaches z, no admissible update may leave it.

To answer Fryszkowski's problem, Andrei Comăneci developed a Lyapunov-type formulation that serves as the starting point for his partial answers to this problem [Com17]. The following characterization applies to every $\alpha \in (0,1)$ and to complete metrics that are not required to be bounded. It replaces the search for a metric by the search for a scalar rank function that vanishes exactly at the endpoint and decreases geometrically along every admissible successor.

Proposition D.1 (Comăneci's rank-function characterization [Com17])**.** *Let $\alpha \in (0,1)$, let Ω be nonempty, and suppose that $F : \Omega \to 2^{\Omega}_{*}$ has an endpoint z. The following statements are equivalent.*

1. *There exists a complete metric d on Ω such that F is a Hausdorff α-contraction with respect to d.*
2. *There exists a function $\phi : \Omega \to [0, \infty)$ such that $\phi^{-1}(0) = \{z\}$ and*

$$\sup_{u \in F(x)} \phi(u) \leq \alpha \phi(x) \qquad \text{for every } x \in \Omega. \tag{D.2}$$

The supremum in (D.2) is important: the rank must decrease for every $u \in F(x)$, rather than only along one selected trajectory. Thus, Comăneci's characterization isolates the dynamical content of the remetrization problem in a single Lyapunov-type condition that controls all branches of the set-valued iteration.

Comăneci then used this characterization to obtain the first of the two partial answers highlighted in his paper: a criterion for the stronger case in which the witnessing metric is required to be bounded. The relevant condition is expressed in terms of the descending sequence of global images

$$\Omega = \widehat{F}^{0}(\Omega) \supseteq \widehat{F}^{1}(\Omega) \supseteq \widehat{F}^{2}(\Omega) \supseteq \cdots. \tag{D.3}$$

The set $\widehat{F}^n(\Omega)$ contains all states that can be reached after n admissible updates from some initial state in Ω. Hence, the intersection of these sets consists of the states that remain globally reachable at arbitrarily large depths.

Comăneci proved that, for $\alpha \in (0, 1/2)$, collapse of these global images to the endpoint is equivalent to the existence of both a bounded rank function and a complete bounded contracting metric. Later, Luchian removed the restriction $\alpha < 1/2$ and established the same characterization for the full contraction range $\alpha \in (0, 1)$ [Luc18].

Theorem D.2 (Comăneci–Luchian bounded remetrization [Com17; Luc18])**.** *Let $\alpha \in (0, 1)$, let Ω be nonempty, and let $F : \Omega \to 2^{\Omega}_{*}$ have an endpoint z. The following statements are equivalent.*

1. *The nested global images collapse to the endpoint:*
$$\bigcap_{n \geq 0} \widehat{F}^n(\Omega) = \{z\}. \tag{D.4}$$
2. *There exists a bounded function $\phi : \Omega \to [0, \infty)$ such that $\phi^{-1}(0) = \{z\}$ and*
$$\sup_{u \in F(x)} \phi(u) \leq \alpha \phi(x) \qquad \text{for every } x \in \Omega. \tag{D.5}$$
3. *There exists a complete bounded metric d on Ω such that F is a Hausdorff α-contraction with respect to d.*

Comăneci proved the equivalence for $\alpha \in (0, 1/2)$, and Luchian extended it to every $\alpha \in (0, 1)$.

The theorem gives a complete set-theoretic criterion for bounded remetrization under the endpoint assumption: the endpoint must be the only state that survives every level of the global iteration. Comăneci's second partial answer treats finite spaces. He showed that, when Ω is finite and $\{z\}$ is the unique nonempty subset fixed by $\widehat{F}$, a complete metric making F an α-contraction exists for every $\alpha \in (0, 1)$ [Com17]. Luchian further studied the stronger equality form of contractivity, namely Nadler α-similarities, under an additional non-overlap assumption [Luc18].

The Comăneci–Luchian results answer Fryszkowski's question at a purely mathematical level. They do not begin with a reference metric, and the constructed metric is not required to generate a prescribed topology or to be quantitatively comparable with any pre-existing notion of distance. Consequently, approximation or convergence in the remetrized space need not translate into approximation or convergence in the metric relevant to an iterative algorithm. Section 4 strengthens the remetrization goal by requiring both topological equivalence and quantitative error transfer.

D.2 Proof of equivalence relations

We prove the equivalence between global-image collapse, uniform convergence over all initial states, and the two convergence conditions (4.1) and (4.2) together with the endpoint condition. This result need not be novel and can be covered by more general results (e.g., see [CM05, Theorem 4.5]). Here, we give an independent and self-contained proof.

Lemma D.3 (Equivalent convergence conditions on compact spaces)**.** *Let (X, d) be a nonempty compact metric space, let $F : X \to \mathcal{K}(X)$ be Hausdorff-continuous, and fix $p \in X$. For $A \subseteq X$, define $\widehat{F}(A) := \bigcup_{x \in A} F(x)$, and write $F^{[t]}(x) := \widehat{F}^t(\{x\})$, where $\widehat{F}^0$ is the identity. Then the following conditions are equivalent:*

(i) The global reachable sets collapse to $\{p\}$:

$$\bigcap_{t\geq 0} \widehat{F}^t(X) = \{p\}.$$

(ii) The reachable sets converge to $\{p\}$ uniformly over all initial states:

$$\lim_{t\to\infty} \sup_{x\in X} H_d\big(F^{[t]}(x), \{p\}\big) = 0.$$

(iii) The point p is an endpoint, meaning $F(p) = \{p\}$, and

$$\lim_{t\to\infty} H_d\big(F^{[t]}(x), \{p\}\big) = 0 \qquad \textit{for every } x \in X.$$

Moreover, there is an open neighborhood U of p such that

$$\lim_{t\to\infty} \sup_{x\in U} H_d\big(F^{[t]}(x), \{p\}\big) = 0.$$

Proof. We first establish the compactness and continuity properties used below, and then explain why the global reachable sets form a nested sequence.

Compactness and continuity. For every nonempty compact set $A \subseteq X$, the set $\widehat{F}(A)$ is nonempty and compact. To see this, consider

$$\Gamma_A := \{(x,u) \in A \times X : u \in F(x)\}.$$

Suppose that $(x_j, u_j) \in \Gamma_A$ converges to $(x,u) \in A \times X$. Since $u_j \in F(x_j)$,

$$\operatorname{dist}_d(u, F(x)) \leq d(u,u_j) + H_d(F(x_j), F(x)) \longrightarrow 0.$$

The set $F(x)$ is closed, so $u \in F(x)$. Thus Γ_A is closed in the compact space $A \times X$, and hence compact. Its projection onto the second coordinate is precisely $\widehat{F}(A)$, proving the claim. Moreover, compactness of X implies that F is uniformly Hausdorff-continuous. Define

$$\omega(r) := \sup\{H_d(F(x), F(y)) : x, y \in X,\ d(x,y) \leq r\}.$$

Then $\omega(r) \to 0$ as $r \downarrow 0$, and

$$H_d\big(\widehat{F}(A), \widehat{F}(B)\big) \leq \omega\big(H_d(A,B)\big) \qquad \text{for all } A, B \in \mathcal{K}(X).$$

Indeed, for every $x \in A$, compactness of B provides $y \in B$ with $d(x,y) \leq H_d(A,B)$. Consequently, every point of $F(x)$ has distance at most $\omega(H_d(A,B))$ from $F(y) \subseteq \widehat{F}(B)$. Exchanging the roles of A and B gives the displayed Hausdorff bound. Therefore, $\widehat{F}$ is continuous on $\mathcal{K}(X)$. Since $x \mapsto \{x\}$ is continuous, every finite iterate $x \mapsto F^{[t]}(x)$ is also Hausdorff-continuous.

Nestedness of the global reachable sets. For each integer $t \geq 0$, let

$$K_t := \widehat{F}^t(X).$$

The preceding argument shows that every K_t is nonempty and compact. We now prove that

$$X = K_0 \supseteq K_1 \supseteq K_2 \supseteq \cdots.$$

First, $\widehat{F}$ preserves set inclusion: if $A \subseteq B$, then

$$\widehat{F}(A) = \bigcup_{x \in A} F(x) \subseteq \bigcup_{x \in B} F(x) = \widehat{F}(B).$$

Also, $F(x) \subseteq X$ for every $x \in X$, so $K_1 = \widehat{F}(X) \subseteq X = K_0$. Now suppose that $K_t \subseteq K_{t-1}$ for some $t \geq 1$. Applying inclusion preservation gives

$$K_{t+1} = \widehat{F}(K_t) \subseteq \widehat{F}(K_{t-1}) = K_t.$$

The desired nestedness follows by induction.

Equivalently, a point reachable after $t+1$ updates from some point of X is also reachable after t updates: discard the first update and regard the resulting state, which still belongs to X, as the initial point. This argument uses the fact that the initial state in the definition of K_t may be any point of X. Since $\widehat{F}$ preserves unions, we also have

$$K_t = \bigcup_{x \in X} F^{[t]}(x).$$

For every nonempty compact set $A \subseteq X$, $H_d(A, \{p\}) = \sup_{a \in A} d(a, p)$. Hence

$$H_d(K_t, \{p\}) = \sup_{x \in X} H_d(F^{[t]}(x), \{p\}).$$

Thus condition (ii) is equivalent to $H_d(K_t, \{p\}) \to 0$.

(i) $\Rightarrow$ (ii). Assume that $\bigcap_{t \geq 0} K_t = \{p\}$. Suppose, for a contradiction, that $H_d(K_t, \{p\})$ does not converge to zero. Then there exist $\varepsilon > 0$, integers $t_j \to \infty$, and points $z_j \in K_{t_j}$ such that

$$d(z_j, p) \geq \varepsilon \qquad \text{for every } j.$$

By compactness of X, we may pass to a subsequence and assume that $z_j \to z \in X$.

Fix any integer $m \geq 0$. For all sufficiently large j, we have $t_j \geq m$, and nestedness gives $K_{t_j} \subseteq K_m$. Thus $z_j \in K_m$ eventually. Since K_m is closed, its limit z also belongs to K_m. As m was arbitrary,

$$z \in \bigcap_{m \geq 0} K_m = \{p\}.$$

However, continuity of the distance gives $d(z, p) \geq \varepsilon$, a contradiction. Therefore $H_d(K_t, \{p\}) \to 0$, proving (ii).

(ii) $\Rightarrow$ (i) **and** (iii)**.** Assume (ii). The sets K_t are nested, nonempty, and compact, so their intersection is nonempty. If z belongs to this intersection, then, for every t,

$$d(z,p) \le H_d(K_t, \{p\}) \longrightarrow 0.$$

Hence $z = p$, proving $\bigcap_{t\ge 0} K_t = \{p\}$ and therefore (i).

We next show that p is an endpoint. Since $p \in K_t$ for every t, inclusion preservation yields

$$F(p) = \widehat{F}(\{p\}) \subseteq \widehat{F}(K_t) = K_{t+1} \qquad \text{for every } t \ge 0.$$

Consequently,

$$F(p) \subseteq \bigcap_{t\ge 0} K_{t+1} = \{p\}.$$

Nonemptiness of $F(p)$ implies $F(p) = \{p\}$. Finally, global uniform convergence implies pointwise convergence and gives the local uniform convergence required in (iii) by taking $U = X$.

(iii) $\Rightarrow$ (ii)**.** Assume (iii), and let U be the open neighborhood of p on which convergence is uniform. For each $x \in X$, pointwise convergence provides an integer $n_x \ge 0$ such that

$$F^{[n_x]}(x) \subseteq U.$$

Since $F^{[n_x]}(x)$ is compact and U is open, there exists $r_x > 0$ such that

$$\{z \in X : \mathrm{dist}_d(z, F^{[n_x]}(x)) < r_x\} \subseteq U.$$

Hausdorff continuity of the finite iterate $y \mapsto F^{[n_x]}(y)$ therefore provides an open neighborhood V_x of x such that

$$F^{[n_x]}(y) \subseteq U \qquad \text{for every } y \in V_x.$$

The sets V_x cover X. By compactness, choose a finite subcover $V_{x_1}, \dots, V_{x_m}$, and write

$$n_i := n_{x_i}, \qquad M := \max_{1\le i\le m} n_i.$$

Thus every initial state has all its reachable states inside U at one of the finitely many times $n_1, \dots, n_m$. Fix $\varepsilon > 0$. Uniform convergence on U provides an integer $T \ge 0$ such that

$$\sup_{z\in U} H_d(F^{[s]}(z), \{p\}) < \frac{\varepsilon}{2} \qquad \text{for every } s \ge T.$$

Now let $y \in X$ and $t \ge M + T$. Choose i with $y \in V_{x_i}$. Then $F^{[n_i]}(y) \subseteq U$ and $t - n_i \ge T$. Splitting the iteration at time n_i gives

$$F^{[t]}(y) = \bigcup_{z\in F^{[n_i]}(y)} F^{[t-n_i]}(z).$$

Every set in this union starts from a point of U and has evolved for at least T further updates. Therefore

$$H_d(F^{[t]}(y), \{p\}) \le \sup_{z\in U} H_d(F^{[t-n_i]}(z), \{p\}) < \frac{\varepsilon}{2}.$$

Since M and T do not depend on y, it follows that

$$\sup_{y\in X} H_d(F^{[t]}(y), \{p\}) \le \frac{\varepsilon}{2} < \varepsilon \qquad \text{for every } t \ge M + T.$$

This proves global uniform convergence and hence (ii). $\square$

D.3 Outline for proof of Theorem 4.1

We first give an outline for our proof. The proof adapts the remetrization scheme of Meyers and of Daskalakis, Tzamos, and Zampetakis [Mey67; DTZ18] to a set-valued iteration. In the single-valued setting, one follows a single orbit of the update map. Here, every part of the construction must instead control the full reachable sets and every admissible successor. The proof is organized into five steps. Step 1 turns the convergence assumptions into a nested global filtration of the state space. Step 2 constructs an equivalent metric under which the correspondence is nonexpansive. Steps 3 and 4 convert progress through the filtration into strict Hausdorff contraction. Step 5 shows that small distance in the constructed metric still yields a useful approximation guarantee in the original metric.

Step 1: Uniform collapse of the global reachable sets. We first upgrade the pointwise convergence of all branches to a uniform statement about the global images of the correspondence. The local uniformity assumption is used to construct an open neighborhood W of the endpoint x^* that is forward invariant under F. Hausdorff continuity implies that the property of entering W after finitely many steps is stable under small perturbations of the initial point, and compactness then turns the resulting point-dependent entrance times into one common entrance time for all of X. Consequently, the sets $K_t := \widehat{F}^t(X)$ form a nested sequence of nonempty compact sets satisfying $H_d(K_t, \{x^*\}) \to 0$ and $\bigcap_{t\geq 0} K_t = \{x^*\}$. These sets provide the dynamical layers used later to measure how far a point has progressed toward the endpoint.

Step 2: A Hausdorff-nonexpansive equivalent metric. Starting from d, we recursively enlarge the metric to incorporate the Hausdorff distance between the outputs of F measured using the metric constructed at the previous stage, and then take the monotone supremum $\overline{d}$ of these auxiliary metrics. This is the set-valued analogue of the nonexpansive closure in the converse of Daskalakis, Tzamos, and Zampetakis [DTZ18, Theorem 1]. Lemma D.5 allows the supremum to pass through the Hausdorff construction, giving $H_{\overline{d}}(F(x), F(y)) \leq \overline{d}(x, y)$. Thus, F is nonexpansive with respect to $\overline{d}$. The uniform collapse obtained in Step 1 is then used to show that the infinite metric closure does not change the topology. Hence $\overline{d}$ is topologically equivalent to d, while $\overline{d} \geq d$ preserves quantitative control over the original notion of distance.

Step 3: Dynamical rank and geodesic distance. Choose N so that the global image K_N lies inside the d-ball of radius ε around x^*. For $x \neq x^*$, define the depth $\tau(x)$ to be the largest index for which $x \in K_{\tau(x)}$, and shift this depth by setting $\ell(x) := \tau(x) - N$. Since $\widehat{F}(K_t) = K_{t+1}$, every admissible transition $u \in F(x)$ increases the rank by at least one. We encode this progress in the rank-weighted function $\rho(x, y) := c^{\min\{\ell(x),\ell(y)\}}\overline{d}(x, y)$. The function ρ need not satisfy the triangle inequality, so we replace it by its geodesic distance D.

We prove that D separates distinct points by obtaining a positive lower bound on the total ρ-length of every connecting sequence. We consider separately sequences that avoid a suitable deeper set K_r and those that enter it. For the latter, we bound the length accumulated up to the first entry. The upper bound $D \leq c^{-N}\overline{d}$, together with compactness, then implies that D generates the same topology as $\overline{d}$ and d.

Step 4: Establish contraction under the metric D. The nonexpansiveness established in Step 2 implies that, for every $u \in F(x)$, one can choose $v \in F(y)$ such that $\overline{d}(u, v) \leq \overline{d}(x, y)$. Since every point in $F(x)$ has rank at least one greater than that of x, the rank weighting strengthens this estimate to $\rho(u, v) \leq c\rho(x, y)$.

We apply this estimate successively to any finite sequence $x = x_0, \dots, x_m = y$ considered in the definition of $D(x, y)$. Starting from an arbitrary $u_0 \in F(x)$, we choose $u_i \in F(x_i)$ so that $\rho(u_{i-1}, u_i) \le c\rho(x_{i-1}, x_i)$ for every i. Summing these inequalities and taking the infimum over all such sequences shows that every point of $F(x)$ has D-distance at most $cD(x, y)$ from $F(y)$. Interchanging x and y gives the same estimate in the opposite direction. Therefore,

$$H_D(F(x), F(y)) \le cD(x, y).$$

Step 5: Transfer approximation guarantees to the original metric. Let $S := K_N$. By construction, S is forward invariant and is contained in $B_d(x^*, \varepsilon)$. Every point outside S has negative shifted rank. Consequently, before a finite sequence of intermediate points first reaches S, each term $\rho(z_{i-1}, z_i)$ in the definition of D is no smaller than $\overline{d}(z_{i-1}, z_i)$. Thus, a small value of $D(x, x^*)$ is possible only if x already belongs to S or is close to S under $\overline{d}$.

The same reasoning applies to two points x and y. Any finite sequence used in the definition of $D(x, y)$ either avoids S, in which case its total ρ-value is at least $\overline{d}(x, y)$, or reaches S, in which case its initial portion is at least the $\overline{d}$-distance from one endpoint to S. Hence, a small value of $D(x, y)$ forces either x and y to be close to each other or one of them to be close to S.

Finally, because $S \subseteq B_d(x^*, \varepsilon)$ and $\overline{d} \ge d$, closeness to S implies closeness to x^* in the original metric. This yields the endpoint and pairwise transfer bounds in (4.4)–(4.5), ensuring that approximation under the constructed metric also gives a quantitative guarantee under d.

D.4 Topological facts

Lemma D.4 (Continuity properties of compact-valued correspondences)**.** *Let (X, d) be a compact metric space, and let $F\colon X \to \mathcal{K}(X)$ be Hausdorff-continuous. For every $A \in \mathcal{K}(X)$, define $\widehat{F}(A) := \bigcup_{x \in A} F(x)$. Then the following statements hold.*

(i) The set $\widehat{F}(A)$ is nonempty and compact for every $A \in \mathcal{K}(X)$. Moreover, the induced map $\widehat{F}\colon \mathcal{K}(X) \to \mathcal{K}(X)$ is continuous with respect to the Hausdorff metric. Consequently, the map $x \mapsto F^{[t]}(x)$ is Hausdorff-continuous for every $t \ge 0$.

(ii) Let $G\colon X \to \mathcal{K}(X)$ be Hausdorff-continuous, and let $V \subseteq X$ be open. Then the set $\{x \in X : G(x) \subseteq V\}$ is open in X.

Proof. We first prove part (i). Fix $A \in \mathcal{K}(X)$. Since both A and every value $F(x)$ are nonempty, the union $\widehat{F}(A)$ is nonempty. Because X is compact, it is enough to show that $\widehat{F}(A)$ is closed.

Let $(y_n)_{n \ge 1}$ be a sequence in $\widehat{F}(A)$ converging to some $y \in X$. For each n, choose $x_n \in A$ such that $y_n \in F(x_n)$. Since A is compact, there is a subsequence $(x_{n_j})_{j \ge 1}$ converging to some $x \in A$. Hausdorff continuity of F gives $H_d(F(x_{n_j}), F(x)) \to 0$. Since $y_{n_j} \in F(x_{n_j})$, we have

$$\operatorname{dist}_d(y, F(x)) \le d(y, y_{n_j}) + \operatorname{dist}_d(y_{n_j}, F(x)) \le d(y, y_{n_j}) + H_d(F(x_{n_j}), F(x)). \tag{D.6}$$

The right-hand side converges to zero, and hence $\operatorname{dist}_d(y, F(x)) = 0$. Because $F(x)$ is compact, and therefore closed, it follows that $y \in F(x)$. Since $x \in A$, we conclude that $y \in \widehat{F}(A)$. Thus $\widehat{F}(A)$ is closed in the compact space X, so it is compact.

We next prove that $\widehat{F}$ is Hausdorff-continuous. Since F is a continuous map from the compact space X into the metric space $(\mathcal{K}(X), H_d)$, it is uniformly continuous by the Heine–Cantor theorem. Thus, for every $\varepsilon > 0$, there exists $\delta > 0$ such that $d(a,b) < \delta$ implies $H_d(F(a), F(b)) < \varepsilon/2$. Let $A, B \in \mathcal{K}(X)$ satisfy $H_d(A,B) < \delta$. Fix $u \in \widehat{F}(A)$. There is some $a \in A$ such that $u \in F(a)$. Because B is compact, there exists $b \in B$ satisfying $d(a,b) = \operatorname{dist}_d(a,B)$. In particular, $d(a,b) < \delta$. Therefore,

$$\operatorname{dist}_d(u, \widehat{F}(B)) \le \operatorname{dist}_d(u, F(b)) \le H_d(F(a), F(b)) < \frac{\varepsilon}{2}. \tag{D.7}$$

Taking the supremum over $u \in \widehat{F}(A)$ bounds the directed Hausdorff distance from $\widehat{F}(A)$ to $\widehat{F}(B)$ by $\varepsilon/2$. Interchanging A and B gives the same bound in the opposite direction. Hence $H_d(\widehat{F}(A), \widehat{F}(B)) \le \varepsilon/2 < \varepsilon$. This proves that $\widehat{F}$ is Hausdorff-continuous.

Finally, define $\iota\colon X \to \mathcal{K}(X)$ by $\iota(x) := \{x\}$. This map is an isometry because $H_d(\{x\}, \{y\}) = d(x,y)$. By the definition of the reachable sets, $F^{[t]}(x) = \widehat{F}^{\,t}(\iota(x))$. Since $\widehat{F}$ and ι are continuous, the map $x \mapsto F^{[t]}(x)$ is Hausdorff-continuous for every $t \ge 0$.

We now prove part (ii). Let $E := \{x \in X : G(x) \subseteq V\}$, and fix $x_0 \in E$. We show that x_0 is an interior point of E. If $V = X$, then $E = X$, so the conclusion is immediate. Suppose therefore that $V \neq X$. Since $G(x_0)$ is compact, $X \setminus V$ is closed, and these two sets are disjoint, their distance $\eta := \operatorname{dist}_d(G(x_0), X \setminus V)$ is strictly positive.

By Hausdorff continuity of G at x_0, there exists a neighborhood U of x_0 such that $H_d(G(x), G(x_0)) < \eta/2$ for every $x \in U$. Fix $x \in U$ and $y \in G(x)$. Then $\operatorname{dist}_d(y, G(x_0)) < \eta/2$. If $y \notin V$, then $y \in X \setminus V$, and the definition of η would imply $\operatorname{dist}_d(y, G(x_0)) \ge \eta$, which is a contradiction. Therefore $y \in V$. Since this holds for every $y \in G(x)$, we have $G(x) \subseteq V$. Thus, $U \subseteq E$, so x_0 is an interior point of E. Since $x_0 \in E$ was arbitrary, the set E is open. □

Lemma D.5 (Monotone Hausdorff-limit identity)**.** *Let (X, τ) be a compact topological space, and let $(d_m)_{m\ge0}$ be metrics on X with a common finite upper bound. Assume that $d_m \le d_{m+1}$ pointwise and that every $d_m : X \times X \to \mathbb{R}$ is continuous for the product topology $\tau \times \tau$. Suppose that*

$$\overline{d}(x,y) := \sup_{m\ge0} d_m(x,y)$$

is finite for every $x, y \in X$. Then, for every pair of nonempty compact sets $A, B \subseteq X$,

$$H_{\overline{d}}(A,B) = \sup_{m\ge0} H_{d_m}(A,B).$$

Proof. Continuity of d_0 makes the identity map from (X,τ) to (X, d_0) continuous. It is a bijection from a compact space to a Hausdorff space, and hence a homeomorphism. Thus τ is metrizable, and every τ-compact subset is sequentially compact. All compactness and continuity arguments below use this common topology.

We first prove the corresponding identity for point-to-set distances. Fix $a \in X$ and a nonempty compact set $B \subseteq X$. Set $\delta_m := \operatorname{dist}_{d_m}(a,B)$ and $\delta := \operatorname{dist}_{\overline{d}}(a,B)$. Since $d_m \le \overline{d}$ pointwise, one has $\delta_m \le \delta$ for every m. Hence, $\sup_{m\ge0} \delta_m \le \delta$.

We then prove the reverse inequality. Because $d_m \le d_{m+1}$, the sequence $(\delta_m)_{m\ge0}$ is nondecreasing. Let $L := \sup_{m\ge0} \delta_m$, so that $\delta_m \to L$ as $m \to \infty$. For every m, the function $b \mapsto d_m(a,b)$ is continuous on the compact set B. Therefore, there exists $b_m \in B$ such that $d_m(a, b_m) = \delta_m$.

By compactness of B, the sequence $(b_m)_{m\geq 0}$ has a subsequence $(b_{m_j})_{j\geq 1}$ converging to some $b^* \in B$. Fix an arbitrary $r \geq 0$. Since $m_j \to \infty$, one has $m_j \geq r$ for all sufficiently large j. Monotonicity of the metrics then gives $d_{m_j}(a, b_{m_j}) \geq d_r(a, b_{m_j})$. Consequently,

$$L = \lim_{j\to\infty} d_{m_j}(a, b_{m_j}) \geq \lim_{j\to\infty} d_r(a, b_{m_j}) = d_r(a, b^*), \tag{D.8}$$

where the final equality follows from continuity of d_r.

Since r was arbitrary, it follows that $L \geq \sup_{r\geq 0} d_r(a, b^*) = \overline{d}(a, b^*)$. Because $b^* \in B$, we also have $\overline{d}(a, b^*) \geq \mathrm{dist}_{\overline{d}}(a, B) = \delta$. Therefore $L \geq \delta$. Together with the previously proved inequality $L \leq \delta$, this gives

$$\mathrm{dist}_{\overline{d}}(a, B) = \sup_{m\geq 0} \mathrm{dist}_{d_m}(a, B).$$

For a metric ρ, write $h_\rho(A, B) := \sup_{a\in A} \mathrm{dist}_\rho(a, B)$ for the directed Hausdorff distance from A to B. Applying the point-to-set identity to each $a \in A$ gives

$$h_{\overline{d}}(A, B) = \sup_{a\in A, m\geq 0} \mathrm{dist}_{d_m}(a, B) = \sup_{m\geq 0} h_{d_m}(A, B). \tag{D.9}$$

The interchange of the two suprema is valid because both sides are the supremum of $\mathrm{dist}_{d_m}(a, B)$ over $(a, m) \in A \times \mathbb{N}_0$. By the same argument, $h_{\overline{d}}(B, A) = \sup_{m\geq 0} h_{d_m}(B, A)$.

Finally, using the definition of the Hausdorff distance,

$$H_{\overline{d}}(A, B) = \max\left\{\sup_{m\geq 0} h_{d_m}(A, B), \sup_{m\geq 0} h_{d_m}(B, A)\right\} = \sup_{m\geq 0} H_{d_m}(A, B). \tag{D.10}$$

Indeed, for any two real sequences (α_m) and (β_m),

$$\max\left\{\sup_m \alpha_m, \sup_m \beta_m\right\} = \sup_m \max\{\alpha_m, \beta_m\}.$$

This proves the identity. □

D.5 Proof of Theorem 4.1

For readability, write $D = D_{c,\varepsilon}$. The construction has three components: a Hausdorff-nonexpansive closure of the original metric, a dynamical rank that increases under every admissible transition, and the geodesic distance of the resulting rank-weighted distance.

Step 1: Uniform collapse of the global reachable sets. By Lemma D.4, we know that every finite iterate $x \mapsto F^{[t]}(x)$ is Hausdorff-continuous.

We first construct a forward-invariant neighborhood of x^* analogous to the construction by Daskalakis and Meyers [Mey67; DTZ18]. Choose an open set V whose closure is contained in U, and choose $r_0 > 0$ such that $B_d(x^*, r_0) \subseteq V$. By (4.2), there is $T_0 \geq 1$ such that $\widehat{F}^{T_0}(U) \subseteq B_d(x^*, r_0) \subseteq V$. Define

$$W := \bigcap_{j=0}^{T_0-1} \left\{x \in X : F^{[j]}(x) \subseteq V\right\}. \tag{D.11}$$

By Lemma D.4, each set $\{x : F^{[j]}(x) \subseteq V\}$ is open, so W is open as a finite intersection of open sets. Also, $x^* \in W$ since $F^{[j]}(x^*) = \{x^*\} \subseteq V$ for every j. Observe first that $W \subseteq V \subseteq U$, since the $j = 0$ term in (D.11) gives $\{x\} \subseteq V$. Let $x \in W$ and $u \in F(x)$. Because $\{u\} \subseteq F(x)$ and $\widehat{F}$ is monotone with respect to inclusion,

$$F^{[j]}(u) = \widehat{F}^j(\{u\}) \subseteq \widehat{F}^j(F(x)) = F^{[j+1]}(x) \qquad \text{for every } j \geq 0.$$

For $0 \leq j \leq T_0 - 2$, the right-hand side lies in V by the definition of W. For $j = T_0 - 1$, we use $x \in U$ and the choice of T_0:

$$F^{[T_0]}(x) \subseteq \widehat{F}^{T_0}(U) \subseteq B_d(x^*, r_0) \subseteq V.$$

Hence $F^{[j]}(u) \subseteq V$ for all $0 \leq j \leq T_0 - 1$, which implies $u \in W$. This proves $\widehat{F}(W) \subseteq W$.

For every $x \in X$, robust global convergence gives an integer m_x such that $F^{[m_x]}(x) \subseteq W$. By applying (ii) of Lemma D.4 to $G = F^{[m_x]}$ with the open set W in the role of V, the inclusion $F^{[m_x]}(x') \subseteq W$ holds for all x' in a neighborhood O_x of x. Compactness provides a finite subcover $O_{x_1}, \ldots, O_{x_s}$. Let $M := \max_i m_{x_i}$. Since W is forward invariant, $\widehat{F}^M(X) \subseteq W$. It follows from the local uniform convergence on $W \subseteq U$ that, for $K_t := \widehat{F}^t(X)$,

$$H_d(K_t, \{x^*\}) \longrightarrow 0. \tag{D.12}$$

The sets K_t are nonempty and compact, satisfy $K_{t+1} \subseteq K_t$, and obey $\bigcap_{t \geq 0} K_t = \{x^*\}$.

Step 2: A Hausdorff-nonexpansive equivalent metric. Define a sequence of metrics by $d_0 := d$ and

$$d_{m+1}(x, y) := \max\left\{d(x, y), H_{d_m}(F(x), F(y))\right\}, \qquad \overline{d}(x, y) := \sup_{m \geq 0} d_m(x, y). \tag{D.13}$$

The pullback $(x, y) \mapsto H_{d_m}(F(x), F(y))$ is a pseudometric, so taking maximum with d creates a metric. Moreover, $d_m \leq d_{m+1}$ and $d_m(x, y) \leq \operatorname{diam}_d(X)$ for all m, x, y. Hence $\overline{d}$ is a finite metric satisfying $\overline{d} \geq d$.

We show by induction on m that each d_m is continuous on $X \times X$ (with the product topology of d). This holds for $d_0 = d$. Assume d_m is continuous. Since $X \times X$ is compact, d_m, which maps from $(X \times X, d(a, a') + d(b, b'))$ to $(\mathbb{R}, |\cdot|)$, is uniformly continuous. Namely, for every $\eta > 0$, there is $\delta > 0$ such that

$$d(a, a') + d(b, b') < \delta \implies |d_m(a, b) - d_m(a', b')| < \eta \qquad \text{for all } a, a', b, b' \in X.$$

Also, $d_m(b, b) = 0$. Now if we let $a' = b' = b$, for every $\eta > 0$, there is $\delta > 0$ such that

$$d(a, b) < \delta \implies d_m(a, b) = |d_m(a, b) - d_m(b, b)| < \eta. \tag{D.14}$$

Let $A, B \in \mathcal{K}(X)$ with $H_d(A, B) < \delta$. For $a \in A$, compactness of B provides $b \in B$ with $d(a, b) = \operatorname{dist}_d(a, B) < \delta$, so $\operatorname{dist}_{d_m}(a, B) \leq d_m(a, b) < \eta$ by (D.14). Interchanging A and B gives

$$H_d(A, B) < \delta \implies H_{d_m}(A, B) \leq \eta. \tag{D.15}$$

Set $\Phi_m(x, y) := H_{d_m}(F(x), F(y))$. Since H_{d_m} is a pseudometric on $\mathcal{K}(X)$, the triangle inequality gives

$$|\Phi_m(x, y) - \Phi_m(x', y')| \leq H_{d_m}(F(x), F(x')) + H_{d_m}(F(y), F(y')).$$

If $(x', y') \to (x, y)$, then $H_d(F(x'), F(x)) \to 0$ and $H_d(F(y'), F(y)) \to 0$ by Hausdorff continuity of F, so both terms on the right tend to 0 by (D.15). Thus Φ_m is continuous, and so is $d_{m+1} = \max\{d, \Phi_m\}$. In particular, as $d \le d_m$, each d_m is topologically equivalent to d, so $\mathcal{K}(X)$ is the same family under d and d_m and H_{d_m} is a genuine metric on it.

By Lemma D.5, for all nonempty compact $A, B \subseteq X$,

$$H_{\overline{d}}(A, B) = \sup_{m \ge 0} H_{d_m}(A, B). \tag{D.16}$$

As a consequence, we can conclude that

$$H_{\overline{d}}(F(x), F(y)) \le \overline{d}(x, y) \qquad \text{for all } x, y \in X. \tag{D.17}$$

It remains to compare $\overline{d}$ with d. Fix $\gamma > 0$. By (D.12), choose q such that $\operatorname{diam}_d(K_q) \le \gamma$. Since K_q is forward invariant, an induction on m gives $\operatorname{diam}_{d_m}(K_q) \le \gamma$ for every m. We claim that for every $j \in \{0, \dots, q\}$, all $x, y \in K_j$, and all $p \ge 0$,

$$d_{q-j+p}(x, y) \le \max\{d_{q-j}(x, y), \gamma\}. \tag{D.18}$$

Since $K_0 = X$, the case $j = 0$ is

$$d_{q+p}(x, y) \le \max\{d_q(x, y), \gamma\} \qquad \text{for all } x, y \in X,\ p \ge 0. \tag{D.19}$$

We prove (D.18) by induction on j. For $j = q$, it is the diameter bound: $x, y \in K_q$ gives $d_p(x, y) \le \operatorname{diam}_{d_p}(K_q) \le \gamma$. Assume the claim for $j + 1$ and let $x, y \in K_j$ and $p \ge 1$ (for $p = 0$ there is nothing to prove). By definition $\widehat{F}(K_j) = K_{j+1}$, so $F(x), F(y) \subseteq K_{j+1}$, and the induction hypothesis (with the same p) gives, for all $u \in F(x)$ and $v \in F(y)$,

$$d_{q-j-1+p}(u, v) \le \max\{d_{q-j-1}(u, v), \gamma\}.$$

Therefore, we can also conclude that

$$H_{d_{q-j-1+p}}(F(x), F(y)) \le \max\big\{H_{d_{q-j-1}}(F(x), F(y)), \gamma\big\},$$

by taking the infimum over $v \in F(y)$, then the supremum over $u \in F(x)$, and repeating with x and y interchanged. Consequently, we have

$$\begin{aligned} d_{q-j+p}(x, y) &= \max\big\{d(x, y), H_{d_{q-j-1+p}}(F(x), F(y))\big\} \\ &\le \max\big\{d(x, y), H_{d_{q-j-1}}(F(x), F(y)), \gamma\big\} \\ &= \max\{d_{q-j}(x, y), \gamma\}, \end{aligned}$$

which is (D.18) for j. Since d_q is continuous and $d_q(x, x) = 0$, (D.19) implies that $d(x, y) \to 0$ entails $\overline{d}(x, y) \to 0$. Together with $d \le \overline{d}$, this proves that d and $\overline{d}$ generate the same topology. Thus $(X, \overline{d})$ is compact and complete.

Step 3: Dynamical rank and geodesic distance. Choose N sufficiently large that $K_N \subseteq B_d(x^*, \varepsilon)$. For $x \neq x^*$, define its global-image depth by

$$\tau(x) := \max\{n \in \mathbb{N}_0 : x \in K_n\}, \qquad \tau(x^*) := +\infty, \qquad \ell(x) := \tau(x) - N. \tag{D.20}$$

The construction of τ is motivated by the construction of n_x in [Luc18, Theorem 3.3] by Luchian. The maximum is finite for $x \neq x^*$ because $\bigcap_n K_n = \{x^*\}$. Also, $\ell(x) \geq -N$ for every x. Most importantly, every admissible transition increases the rank: if $u \in F(x)$, then $x \in K_{\tau(x)}$ implies $u \in \widehat{F}(K_{\tau(x)}) = K_{\tau(x)+1}$, and hence

$$\ell(u) \geq \ell(x) + 1. \tag{D.21}$$

Define the symmetric rank-weighted distance

$$\rho(x, y) := c^{\min\{\ell(x), \ell(y)\}} \overline{d}(x, y), \tag{D.22}$$

with the convention $c^{+\infty} = 0$. The function ρ need not satisfy the triangle inequality. We therefore take its geodesic distance:

$$D(x, y) := \inf_{\substack{m \geq 1 \\ x = x_0, \ldots, x_m = y}} \sum_{i=1}^{m} \rho(x_{i-1}, x_i). \tag{D.23}$$

Symmetry and the triangle inequality are immediate for D. We next verify positivity of D.

Suppose first that $x \neq y$ and $\tau(x) \leq \tau(y) < +\infty$. Consider an arbitrary chain (or sequence) from x to y: $x = z_0, z_1, \ldots, z_m = y$. We call each consecutive pair (z_{i-1}, z_i) a link of the chain. Its ρ-length is

$$\rho(z_{i-1}, z_i) = c^{\min\{\ell(z_{i-1}), \ell(z_i)\}} \overline{d}(z_{i-1}, z_i).$$

The factor $c^{\min\{\ell(z_{i-1}), \ell(z_i)\}}$ will be called the rank multiplier of the link. Set $r := \tau(y) + 1$. If $z \notin K_r$, then $\tau(z) \leq \tau(y)$ and hence $\ell(z) \leq \ell(y)$. Since $0 < c < 1$, this implies $c^{\ell(z)} \geq c^{\ell(y)}$.

There are now two cases. First, suppose that the chain does not meet K_r. Then every point z_i lies outside K_r, so the rank multiplier of every link is at least $c^{\ell(y)}$. Therefore,

$$\sum_{i=1}^{m} \rho(z_{i-1}, z_i) \geq c^{\ell(y)} \sum_{i=1}^{m} \overline{d}(z_{i-1}, z_i) \geq c^{\ell(y)} \overline{d}(x, y).$$

Second, suppose that the chain meets K_r, and let j be the smallest index such that $z_j \in K_r$. Then $z_0, \ldots, z_{j-1} \notin K_r$. Hence every link in the initial segment $z_0, \ldots, z_j$ has rank multiplier at least $c^{\ell(y)}$, and

$$\sum_{i=1}^{m} \rho(z_{i-1}, z_i) \geq c^{\ell(y)} \sum_{i=1}^{j} \overline{d}(z_{i-1}, z_i) \geq c^{\ell(y)} \overline{d}(x, z_j) \geq c^{\ell(y)} \operatorname{dist}_{\overline{d}}\left(x, K_{\tau(y)+1}\right).$$

Since every chain falls into one of these two cases, taking the infimum over all chains gives

$$D(x, y) \geq c^{\ell(y)} \min\left\{\overline{d}(x, y), \operatorname{dist}_{\overline{d}}\left(x, K_{\tau(y)+1}\right)\right\}.$$

Finally, $x \neq y$ gives $\overline{d}(x, y) > 0$, while $\tau(x) \leq \tau(y)$ implies $x \notin K_{\tau(y)+1}$. Since $K_{\tau(y)+1}$ is closed, its $\overline{d}$-distance from x is strictly positive. Thus $D(x, y) > 0$. Therefore

$$D(x, y) \geq c^{\ell(y)} \min\left\{\overline{d}(x, y), \operatorname{dist}_{\overline{d}}\left(x, K_{\tau(y)+1}\right)\right\} > 0. \tag{D.24}$$

If $y = x^*$, the same argument with $K_{\tau(x)+1}$ gives

$$D(x, x^*) \geq c^{\ell(x)} \operatorname{dist}_{\overline{d}}\Big(x, K_{\tau(x)+1}\Big) > 0. \tag{D.25}$$

Thus D is a metric.

Now if we take a chain (x, y) with length one and apply the lower bound $\ell \geq -N$, we get $D(x, y) \leq c^{-N}\overline{d}(x, y)$. Hence the identity map from the compact space $(X, \overline{d})$ to the Hausdorff space (X, D) is continuous and bijective. It is therefore a homeomorphism. Thus D is topologically equivalent to both $\overline{d}$ and d, and (X, D) is compact and complete.

Step 4: Construct the contraction. Fix $x, y \in X$ and $u \in F(x)$. By Step 2, we know that $\overline{d}$ and d generate the same topology, so $F(y)$ is compact in $(X, \overline{d})$ and the continuous function $\overline{d}(u, \cdot)$ attains its infimum on $F(y)$ at some $v \in F(y)$. Since $u \in F(x)$, (D.17) gives

$$\overline{d}(u, v) = \operatorname{dist}_{\overline{d}}\big(u, F(y)\big) \leq \sup_{u' \in F(x)} \operatorname{dist}_{\overline{d}}\big(u', F(y)\big) \leq H_{\overline{d}}\big(F(x), F(y)\big) \leq \overline{d}(x, y).$$

Combining this with (D.21) yields

$$\rho(u, v) \leq c\rho(x, y). \tag{D.26}$$

Now fix a chain $x = x_0, \ldots, x_m = y$ and a point $u_0 \in F(x_0)$. Applying (D.26) successively, we can construct a chain $u_0, \ldots, u_m$ by choosing $u_i \in F(x_i)$ such that $\rho(u_{i-1}, u_i) \leq c\rho(x_{i-1}, x_i)$ at each $i \in [m]$. Then

$$D(u_0, u_m) \leq c \sum_{i=1}^{m} \rho(x_{i-1}, x_i). \tag{D.27}$$

Taking the infimum over all chains gives

$$\operatorname{dist}_D(u_0, F(y)) \leq cD(x, y).$$

Then taking the supremum over $u_0 \in F(x)$ gives

$$\sup_{u \in F(x)} \operatorname{dist}_D(u, F(y)) \leq cD(x, y).$$

Interchanging x and y gives the corresponding estimate from $F(y)$ to $F(x)$. Taking the maximum of the two directed estimates proves (4.3).

Step 5: Transfer to the original metric. Let $S := K_N$, so that $S \subseteq B_d(x^*, \varepsilon)$. If $x \notin S$, then $\ell(x) < 0$. Every chain $x = z_0, z_1, \ldots, z_m = x^*$ must enter S, and every link (z_{i-1}, z_i) up to and including the first entrance $z_j \in S$ has $z_{i-1} \notin S$ (so $\min\{\ell(z_{i-1}), \ell(z_i)\} \leq -1$) and hence has ρ-length at least its $\overline{d}$-length. Consequently,

$$D(x, x^*) \geq \operatorname{dist}_{\overline{d}}(x, S). \tag{D.28}$$

If $D(x, x^*) \leq \varepsilon$, then either $x \in S$, in which case $d(x, x^*) \leq \varepsilon$, or $\operatorname{dist}_d(x, S) \leq \operatorname{dist}_{\overline{d}}(x, S) \leq \varepsilon$. In the latter case,

$$d(x, x^*) \leq \operatorname{dist}_d(x, S) + \sup_{z \in S} d(z, x^*) \leq 2\varepsilon. \tag{D.29}$$

This proves (4.4).

For the pairwise statement, suppose first that $x, y \notin S$. A chain from x to y either avoids S, in which case its length is at least $\overline{d}(x,y)$, or enters S, in which case its initial segment has length at least $\mathrm{dist}_{\overline{d}}(x,S)$. Hence

$$D(x,y) \geq \min\left\{\overline{d}(x,y), \mathrm{dist}_{\overline{d}}(x,S)\right\}. \tag{D.30}$$

Thus $D(x,y) \leq \varepsilon$ implies either $d(x,y) \leq \varepsilon$ or $d(x,x^*) \leq 2\varepsilon$. If one of x and y already belongs to S, then its d-distance from x^* is at most ε. This proves (4.5) and completes the proof.

Remark D.6. Theorem 4.1 is a mathematical universality result: every robustly globally convergent compact-valued iteration admits a complete equivalent metric under which all branches contract geometrically. Its construction is not automatically efficient. In particular, it uses the globally reachable sets K_t, the infinite supremum defining $\overline{d}$, and the shortest-chain closure defining D. These objects need not have polynomial-size circuit representations. Consequently, a computational converse suitable for a CLS reduction requires additional succinctness assumptions, such as an efficiently evaluable potential that decreases for every $u \in F(x)$. The compact, Hausdorff-continuous formulation above is nevertheless compatible with the representation of correspondences used in computational Kakutani problems [PVZ23].

D.6 Proof of Corollary 4.2

Since $F(x^*) = \{x^*\}$, the contraction inequality implies that every $u \in F(x)$ satisfies $D(u,x^*) \leq cD(x,x^*)$. This gives (4.6). It also yields, for every $u \in F(x)$,

$$\begin{aligned} D(x,u) &\geq D(x,x^*) - D(u,x^*) \geq (1-c)D(x,x^*), \\ D(x,u) &\leq D(x,x^*) + D(u,x^*) \leq (1+c)D(x,x^*). \end{aligned} \tag{D.31}$$

Taking the infimum over $u \in F(x)$ proves (4.7). In particular, a fixed point has zero distance from x^*, proving uniqueness. Finally, (4.7) and (4.6) give

$$D(x_t,x^*) \leq \frac{c^t}{1-c} r_D(x_0).$$

If $r_D(x_0) > 0$, the choice of t in (4.8) makes the right-hand side at most ε, so (4.4) gives $d(x_t,x^*) \leq 2\varepsilon$. If $r_D(x_0) = 0$, the residual lower bound gives $x_0 = x^*$, and every admissible trajectory remains at the endpoint.

E Omitted Proofs in Section 5

E.1 Technical lemmas

We first isolate the quantitative facts used by the reduction.

Lemma E.1 (Conversion between approximate FOSP and KKT point)**.** *Let $K = [0,1]^q$, and suppose that $\mathcal{L}$ is continuously differentiable on K and $\nabla\mathcal{L}$ is Λ-Lipschitz on K for some $\Lambda > 0$. Let $\delta > 0$, and assume that $z \in K$ is a δ-FOSP for $\mathcal{L}$. Then*

$$y := \Pi_K\left(z - \frac{1}{\Lambda}\nabla\mathcal{L}(z)\right) \tag{E.1}$$

is a $2\sqrt{\Lambda\delta}$-KKT point of $\mathcal{L}$ over K.

Proof. Let $g := \nabla \mathcal{L}(z)$. By the first-order characterization of Euclidean projection, (E.1) implies

$$z - \frac{1}{\Lambda} g - y \in N_K(y).$$

Since $z \in K$, we may test $z - \frac{1}{\Lambda} g - y$ against $z - y$. This gives

$$0 \geq \left\langle z - \frac{1}{\Lambda} g - y, z - y \right\rangle = \|z - y\|_2^2 - \frac{1}{\Lambda} \langle g, z - y \rangle. \tag{E.2}$$

Consequently, $\|z - y\|_2^2 \leq \frac{1}{\Lambda} \langle g, z - y \rangle$. Applying the δ-FOSP condition with $u = y$ yields $\langle g, z - y \rangle \leq \delta$. Combining this inequality with (E.2), we obtain $\|z - y\|_2 \leq \sqrt{\frac{\delta}{\Lambda}}$. Now define

$$n := \Lambda(z - y) - g \in N_K(y).$$

The Λ-Lipschitz continuity of the gradient implies

$$\begin{aligned} \|\nabla \mathcal{L}(y) + n\|_2 &= \|\nabla \mathcal{L}(y) - \nabla \mathcal{L}(z) + \Lambda(z - y)\|_2 \\ &\leq \|\nabla \mathcal{L}(y) - \nabla \mathcal{L}(z)\|_2 + \Lambda \|z - y\|_2 \\ &\leq 2\Lambda \|z - y\|_2 \leq 2\sqrt{\Lambda \delta}. \end{aligned} \tag{E.3}$$

It follows in particular that $\|\nabla \mathcal{L}(y) + n\|_\infty \leq 2\sqrt{\Lambda \delta}$. It remains to translate (E.3) into the coordinatewise KKT conditions. For the box $K = [0,1]^q$, every $n \in N_K(y)$ satisfies

$$n_i \begin{cases} \leq 0, & y_i = 0, \\ = 0, & 0 < y_i < 1, \\ \geq 0, & y_i = 1. \end{cases}$$

Suppose that $y_i > 0$. If $y_i < 1$, then $n_i = 0$; if $y_i = 1$, then $n_i \geq 0$. In either case,

$$\partial_i \mathcal{L}(y) \leq \partial_i \mathcal{L}(y) + n_i \leq |\partial_i \mathcal{L}(y) + n_i| \leq 2\sqrt{\Lambda \delta}.$$

Similarly, suppose that $y_i < 1$. If $y_i > 0$, then $n_i = 0$; if $y_i = 0$, then $n_i \leq 0$. Hence

$$\partial_i \mathcal{L}(y) \geq \partial_i \mathcal{L}(y) + n_i \geq -|\partial_i \mathcal{L}(y) + n_i| \geq -2\sqrt{\Lambda \delta}.$$

Finally, substituting $\delta = \frac{\varepsilon^2}{4\Lambda}$ gives $2\sqrt{\Lambda \delta} = \varepsilon$, so an $\varepsilon^2/(4\Lambda)$-FOSP is converted into an ε-KKT point. This is an analytic conclusion, since continuous differentiability alone gives no algorithm for evaluating the gradient. If $\nabla \mathcal{L}$ can be evaluated exactly in polynomial time on rational inputs, with polynomially bounded output length, and $\Lambda > 0$ is a rational of polynomial encoding length, then y is computable in polynomial time from rational z. Specifically, we can evaluate the gradient, take the rational step, and clip each coordinate to $[0,1]$. The rational quadratic objectives constructed below have this property. The running time here is measured in both the instance length and the encoding length of z. □

Lemma E.2 (Large margin leads to non-negativity)**.** *Let $W := \sum_{t \in \mathcal{C}} w_t$, and assume that $q \geq 1$, $W > 0$, and $\alpha \geq d_{\text{emb}} \geq 1$. Then the following properties hold.*

(i) *For every $z \in K$ and every triplet $t \in \mathcal{C}$, we have*

$$\alpha - d_{\text{emb}} \le \|z_{i_t} - z_{j_t}\|_2^2 - \|z_{i_t} - z_{k_t}\|_2^2 + \alpha \le \alpha + d_{\text{emb}}.$$

Consequently, $\mathcal{L}$ agrees on K with the quadratic polynomial

$$\mathcal{L}(z) = \sum_{t \in \mathcal{C}} w_t \left(\|z_{i_t} - z_{j_t}\|_2^2 - \|z_{i_t} - z_{k_t}\|_2^2 + \alpha \right). \tag{E.4}$$

(ii) *For every $z \in K$, $W(\alpha - d_{\text{emb}}) \le \mathcal{L}(z) \le W(\alpha + d_{\text{emb}})$. Define $R := 2d_{\text{emb}}W$ and*

$$p(z) := \frac{\mathcal{L}(z) - W(\alpha - d_{\text{emb}})}{R}. \tag{E.5}$$

Then $p(z) \in [0,1]$ for every $z \in K$.

Proof. For any $a, b \in [0,1]^{d_{\text{emb}}}$, one has $0 \le \|a - b\|_2^2 \le d_{\text{emb}}$. Hence

$$\|z_{i_t} - z_{j_t}\|_2^2 - \|z_{i_t} - z_{k_t}\|_2^2 + \alpha$$

lies in $[\alpha - d_{\text{emb}}, \alpha + d_{\text{emb}}]$ and is non-negative. This proves (E.4). The same bound implies $W(\alpha - d_{\text{emb}}) \le \mathcal{L}(z) \le W(\alpha + d_{\text{emb}})$, so (E.5) takes values in $[0,1]$. □

Expanding the squared distances in (E.4) gives a rational quadratic in the embeddings, which gives the following bounds.

Lemma E.3 (Lipschitz bound for quadratic formula)**.** *Let*

$$\mathcal{L}(z) = \frac{1}{2} z^\top H z + h^\top z + c,$$

where $H \in \mathbb{Q}^{q \times q}$ is symmetric and c is constant. Define

$$B_H = \max_{\ell \in [q]} \sum_{j=1}^{q} |H_{\ell j}|, \quad \Lambda = \max\{1, B_H\}, \quad G = \max\{1, \|h\|_\infty + B_H\}, \quad L_p = \frac{G}{R}.$$

Then, for all $x, y, z \in K$, $\nabla \mathcal{L}$ is Λ-Lipschitz with respect to ℓ_2, with $\|\nabla \mathcal{L}(z)\|_\infty \le G$, and

$$|p(x) - p(y)| \le L_p \|x - y\|_1.$$

The matrix H, the vector h, and all constants introduced above can be computed from the triplet instance in polynomial time and have polynomial encoding length.

Proof. For $x, y \in K$, we can show that $\nabla \mathcal{L}(z) = Hz + h$ and

$$\nabla \mathcal{L}(x) - \nabla \mathcal{L}(y) = H(x - y).$$

Because H is symmetric, its maximum absolute row sum equals its maximum absolute column sum. Therefore,

$$\|H\|_2 \le \sqrt{\|H\|_1 \|H\|_\infty} = B_H \le \Lambda. \tag{E.6}$$

Thus $\nabla\mathcal{L}$ is Λ-Lipschitz in ℓ_2. Since $\|z\|_\infty \le 1$ on K, we also have $\|Hz+h\|_\infty \le B_H + \|h\|_\infty \le G$. The mean-value theorem then gives, for all $x, y \in K$,

$$|p(x) - p(y)| \le \frac{1}{R} \sup_{z\in K} \|\nabla\mathcal{L}(z)\|_\infty \|x-y\|_1 \le L_p \|x-y\|_1. \tag{E.7}$$

Expanding the squared distances in (E.4) into quadratic form will produce H, h, and c using polynomially many rational arithmetic operations, which proves the final claim. □

Lemma E.4 (Projected-gradient certificate)**.** *Under the assumptions of Lemma E.2 and E.3, define*

$$T(z) := \Pi_K\left(z - \frac{1}{\Lambda}\nabla\mathcal{L}(z)\right), \tag{E.8}$$

where Π_K is Euclidean projection onto K, and set

$$\delta := \min\left\{\frac{1}{2}, \frac{\varepsilon^2}{8qR\Lambda}\right\}. \tag{E.9}$$

The map T is represented by a polynomial-size arithmetic circuit and is 2-Lipschitz with respect to ℓ_1. Furthermore, if $x \in K$ satisfies $p(T(x)) \ge p(x) - \delta$, then $T(x)$ is an ε-FOSP of $\mathcal{L}$.

Proof. Projection onto the box K is coordinatewise clipping to $[0,1]$, so it is computable by a polynomial-size circuit and is nonexpansive in ℓ_1. Using $\nabla\mathcal{L}(x) - \nabla\mathcal{L}(y) = H(x-y)$ and

$$\|H\|_{1,1} = \max_{\ell\in[q]} \sum_{j=1}^{q} |H_{\ell j}| = B_H$$

for symmetric matrix H, we obtain

$$\begin{aligned} \|T(x) - T(y)\|_1 &\le \left\| (x-y) - \frac{1}{\Lambda}\left(\nabla\mathcal{L}(x) - \nabla\mathcal{L}(y)\right)\right\|_1 \\ &\le \left(1 + \frac{B_H}{\Lambda}\right)\|x-y\|_1 \le 2\|x-y\|_1. \end{aligned} \tag{E.10}$$

Here $\|\cdot\|_{1,1}$ is the operator norm when both input and output are measured by ℓ_1 norm. Now fix $x \in K$ with $p(T(x)) \ge p(x) - \delta$ and write $y := T(x)$. Projection optimality with the comparison point x gives $\langle \nabla\mathcal{L}(x), y - x\rangle \le -\Lambda\|y-x\|_2^2$. By Λ-smoothness,

$$\mathcal{L}(y) \le \mathcal{L}(x) + \langle\nabla\mathcal{L}(x), y-x\rangle + \frac{\Lambda}{2}\|y-x\|_2^2 \le \mathcal{L}(x) - \frac{\Lambda}{2}\|y-x\|_2^2. \tag{E.11}$$

Since $\mathcal{L} = Rp + W(\alpha - d_{\text{emb}})$, the assumed potential inequality and (E.11) imply $\|y-x\|_2 \le \sqrt{2R\delta/\Lambda}$. For any $u \in K$, the variational inequality for the projection in (E.8) yields

$$\langle \nabla\mathcal{L}(x), u - y\rangle \ge \Lambda\langle x - y, u - y\rangle. \tag{E.12}$$

Combining (E.12) with the Λ-Lipschitz continuity of the gradient and using $\|u-y\|_2 \le \sqrt{q}$, we obtain

$$\langle u - y, \nabla\mathcal{L}(y)\rangle \ge -2\Lambda\|x-y\|_2\|u-y\|_2 \ge -\sqrt{8qR\Lambda\delta} \ge -\varepsilon. \tag{E.13}$$

Thus $y = T(x)$ satisfies Definition 3. □

Lemma E.5 (Potential ultrametric). *Let $p : K \to [0,1]$ be the potential from (E.5), and define*

$$d_p(x,y) := \begin{cases} 0, & x = y, \\ 1 + \max\{p(x), p(y)\}, & x \neq y. \end{cases} \tag{E.14}$$

Then d_p is a complete ultrametric on K with diameter at most two. In addition, for any off-diagonal pairs (a,b) and (a',b'),

$$|d_p(a,b) - d_p(a',b')| \le L_p\big(\|a - a'\|_1 + \|b - b'\|_1\big). \tag{E.15}$$

Proof. If x, y, z are pairwise distinct, then $d_p(x,z) \ge 1 + p(x)$ and $d_p(z,y) \ge 1 + p(y)$, so $d_p(x,y) \le \max\{d_p(x,z), d_p(z,y)\}$. The remaining cases are immediate. Thus d_p is an ultrametric. Every nonzero distance belongs to $[1,2]$; consequently, every d_p-Cauchy sequence is eventually constant. Hence (K, d_p) is complete and has diameter at most two.

For off-diagonal pairs, the constant term in (E.14) cancels. The elementary Lipschitz property of the maximum and Lemma E.3 give

$$\begin{aligned} |d_p(a,b) - d_p(a',b')| &\le \max\big\{|p(a) - p(a')|, |p(b) - p(b')|\big\} \\ &\le L_p\big(\|a - a'\|_1 + \|b - b'\|_1\big). \end{aligned}$$

This proves (E.15). □

E.2 Proof of Theorem 5.1 and Corollary 5.2

Proof of Theorem 5.1. If $V = \emptyset$, the reduction maps the input to any fixed valid β-PROJECTED-NADLER instance (for $\beta = 1$) and returns the empty tuple as the exact FOSP. If $V \neq \emptyset$ and $W = \sum_{t \in \mathcal{C}} w_t = 0$, the objective is constant; in this case, we also map the input to any fixed valid instance and decode every answer as $z = 0$, which is an exact FOSP. Assume that $W > 0$ and $V \neq \emptyset$, and construct p from Lemma E.2, T, δ from Lemma E.4, and d_p from Lemma E.5. Define the singleton correspondence $F_{\mathcal{L}}(x) := \{T(x)\}$ and its projection circuit by $\Pi_{F_{\mathcal{L}}}(x,z) := T(x)$. Since every value is a singleton, this is an exact projection circuit under d_p. The target instance uses

$$D = 2, \quad k = 1 - \delta/2, \quad \varepsilon_1 = 1/2, \quad \varepsilon_2 = \delta/8, \quad L_P = 2, \quad L_d = L_p.$$

Lemma E.5 proves completeness of the metric space and the diameter bound, and $\delta \le 1/2$ ensures $k \in (0,1)$. Also, $(1-k)\varepsilon_1 = \delta/4 > \delta/8 = \varepsilon_2 > 0$, which satisfies the promise on accuracy.

Suppose first that the solver returns a point x of type (N1). Then $d_p(x, T(x)) \le 1/2$. Because every nonzero d_p-distance is at least one, $x = T(x)$. In particular, $p(T(x)) = p(x)$, so Lemma E.4 shows that x is an ε-FOSP.

Now suppose that the solver returns a certificate of type (N2). Since the projection circuit ignores its second argument, either branch of the certificate gives

$$d_p(T(x), T(y)) > k\, d_p(x,y) + \varepsilon_2 > k\, d_p(x,y).$$

In particular, $T(x) \neq T(y)$, which also implies $x \neq y$. We show that at least one $v \in \{x, y\}$ satisfies $p(T(v)) \geq p(v) - \delta$. Suppose otherwise, so that $p(T(x)) < p(x) - \delta$ and $p(T(y)) < p(y) - \delta$. Since both pairs (x, y) and $(T(x), T(y))$ consist of distinct points, the off-diagonal formula for d_p gives

$$\begin{aligned} d_p(T(x), T(y)) &= 1 + \max\{p(T(x)), p(T(y))\} \\ &< 1 + \max\{p(x), p(y)\} - \delta \\ &= d_p(x, y) - \delta \\ &\leq \left(1 - \frac{\delta}{2}\right) d_p(x, y) \\ &= k\, d_p(x, y), \end{aligned} \tag{E.16}$$

where the last inequality follows from $d_p(x, y) \leq 2$. This contradicts the returned certificate. Thus, the decoder checks the potential-decrease condition at x and y, selects a point v satisfying $p(T(v)) \geq p(v) - \delta$, and returns $T(v)$. By Lemma E.4, this point is an ε-FOSP.

Finally, equation (E.10) rules out type (N3) for $\beta = 1$, and equation (E.15) rules out type (N4). The matrices, constants, and arithmetic circuits used above have polynomial encoding length; the metric circuit uses two evaluations of p, a maximum gate, and comparison gates for equality, and the decoder evaluates at most two projected-gradient updates. Therefore both the instance map and the solution decoder run in polynomial time. □

Proof of Corollary 5.2. The CLS-membership follows from Theorems 5.1 and Theorem 3.2. For CLS-hardness, we reduce from QUADRATICPROGRAM-KKT (Definition 4), which remains CLS-complete when the accuracy $\varepsilon > 0$ is given in binary [FGHS25]. Let (Q, b, ε) be an instance on $[0, 1]^n$, and write $p(x) = x^\top Q x + b^\top x$.

We use the construction in the proof of [YLCP+25, Theorem 4.1] with embedding dimension $d_{\text{emb}} = 1$, pivots $a_A = 0$ and $a_B = 1/2$, margin $\alpha = 1 = d_{\text{emb}}$, and one movable point per variable. Thus, we have $K = [0, 1]^n$ and $z_i = x_i$. Each coefficient of p is realized by a triplet and its dual, with nonnegative rational weights whose difference is a prescribed rational linear combination of the entries of Q and b. This gives $\mathcal{L}(z) = p(z) + C$ on K for some constant C, and consequently $\nabla\mathcal{L} = \nabla p$ on K. Since $\alpha \geq d_{\text{emb}}$, this is an instance of Definition 3. Let Λ be the constant of Lemma E.3 for $\mathcal{L}$, and set the target accuracy to $\varepsilon_T := \varepsilon^2/(4\Lambda)$. All weights, Λ and ε_T have polynomial encoding length.

Given an ε_T-FOSP $z \in K$, output $y := \Pi_K\big(z - \nabla\mathcal{L}(z)/\Lambda\big)$, which is computable in polynomial time by coordinatewise clipping. By Lemma E.1, y is an ε-KKT point of $\mathcal{L}$ over K. Since $\nabla\mathcal{L} = \nabla p$ on K, y is an ε-KKT point of (5.3). □