hexagon:2611.00217v1

PREPRINT • October 2026

The Logical Anatomy of Optimization

Reverse Mathematics, Proof Mining, and the Price of Convergence

Xiaoyu Li
University of New South Wales
xiaoyulics@gmail.com

ABSTRACT

Optimization theory proves three kinds of theorems: that optimal points exist, that algorithms converge to them, and that optimality can be certified by duality. We measure how much mathematics each kind needs, using reverse mathematics, Weihrauch complexity and proof mining, and assemble the answers, old and new, into one map stated in the language of optimization. Compactness makes existence cost weak Kőnig's lemma, but convexity (and, format by format, tameness) makes it cheap: we show that RCA_0 proves that convex functions on $[0,1]$ attain their minima (and on $[0,1]^n$, for each fixed n, given a modulus of uniform continuity), and that RCA_0 proves linear programming duality and the minimax theorem for real data of every size. Convergence costs more: it follows from results of Kohlenbach and Neumann that convergence of gradient descent for convex smooth functions on $\mathbb{R}$ is equivalent to ACA_0. We explain this by a selection principle: what costs is the particular minimizer the algorithm selects, and uniqueness or an error bound removes the cost. Below convergence lies finitary information. Gradient descent in $\mathbb{R}^d$ has at most $2C_dD/\varepsilon$ fluctuations of size ε, with C_d depending only on d, which yields rates of metastability; no such bound is dimension-free, by the span-restriction argument behind performance estimation; and the limits of gradient descent on computable instances have d.c.e. coordinates, a proper subclass of the limit-computable reals. The map also covers Ekeland's variational principle (up to Π^1_1-CA_0), separation theorems, Kőnig duality (ATR_0), proof-mined rates for fixed-point iterations, and the feasible world, where local search captures the NP search problems provably total in T^1_2 and approximate stationary points of smooth functions are complete for $\mathsf{CLS} = \mathsf{PPAD} \cap \mathsf{PLS}$. We close with five principles and open problems.

Keywords: reverse mathematics, Weihrauch complexity, proof mining, convex optimization, gradient descent, metastability, duality, bounded arithmetic

CONTENTS

1 Introduction . . . 2
2 Two toolkits . . . 5
3 Existence: what it costs to have a minimizer . . . 8
4 Convergence: what it costs to get there . . . 13
5 Proof mining: uniformity as complexity . . . 22
6 Duality and certificates . . . 25
7 The feasible world: bounded arithmetic, TFNP and proof systems . . . 27
8 Synthesis and open problems . . . 29
References . . . 33

Appendix
A Formalization notes . . . 43
B The computations behind Fig. 2 . . . 43
C Linear programming duality in RCA_0: the details . . . 44

1 Introduction

Optimization theory proves three kinds of theorems. *Existence* theorems say that an optimal point exists: a continuous function on a compact set attains its minimum; a coercive lower semicontinuous convex functional on a Hilbert space has a minimizer. *Convergence* theorems say that an algorithm reaches one: gradient descent, the proximal point method, the Krasnosel'skiĭ–Mann and Halpern iterations. *Certification* theorems say that optimality can be verified: Lagrange and Fenchel duality, separating hyperplanes, the minimax theorem, Kőnig's theorem and max-flow/min-cut. Optimizers know that these three kinds of statements behave differently. Existence is cheap to state and often cheap to prove; convergence comes with rates for some quantities and not for others; certificates are finite objects in finite dimensions and delicate ones in infinite dimensions.

This paper makes the difference precise, using three rulers from mathematical logic:

- *reverse mathematics* [52, 80, 158], which measures the set-existence axioms a theorem needs;
- *Weihrauch complexity* [29], which measures how hard it is to compute a witness uniformly in the data;
- *proof mining* [98], which extracts the quantitative content (rates, bounds, metastability) that a proof implicitly carries.

Each ruler has been applied to optimization before, often deeply. The extreme value theorem is the textbook example of a statement equivalent to weak Kőnig's lemma. Proof mining began partly in Chebyshev approximation [93] and has produced explicit rates for Krasnosel'skiĭ–Mann, Halpern, proximal point and other Fejér-type methods of convex optimization [100]. Neumann classified fixed-point problems for nonexpansive maps in the Weihrauch lattice [137], Kohlenbach calibrated moduli of regularity in reverse mathematics [101], Fernández-Duque, Shafer and Yokoyama calibrated Ekeland's variational principle [58], and bounded arithmetic characterized local search [32]. But these results are scattered over several literatures that rarely cite each other, and they are stated in languages that optimizers do not use. What has been missing is a single map on which the three kinds of theorems can be compared, in the language of optimization, together with an explanation of the patterns on the map. This paper provides both, and proves new results where the map had holes.

The map. Figure 1 places the main principles of optimization in the hierarchy of subsystems of second-order arithmetic. Three patterns stand out, and each is the subject of a part of the paper.

1. *Convexity and tameness collapse existence* (Section 3). A continuous function on $[0,1]$ attains its maximum; this is equivalent to WKL_0, and there are computable functions with no computable maximizer. A *convex* function on $[0,1]$ attains its minimum provably in RCA_0, and so does a convex function on $[0,1]^n$ that comes with a modulus of uniform continuity (Theorem 3.2); every computable convex function has a computable minimizer, and convex minimization is Weihrauch equivalent to convex choice (Theorem 3.3). Existence of optima for polynomial problems of fixed format is provable in RCA_0 by Tarski transfer (Theorem 3.5). At the other end, Ekeland's variational principle, the engine of modern variational analysis, is equivalent to Π^1_1-CA_0 in its general forms (Theorem 3.6).
2. *Convergence costs a jump, because convergence selects* (Section 4). Already in dimension one,

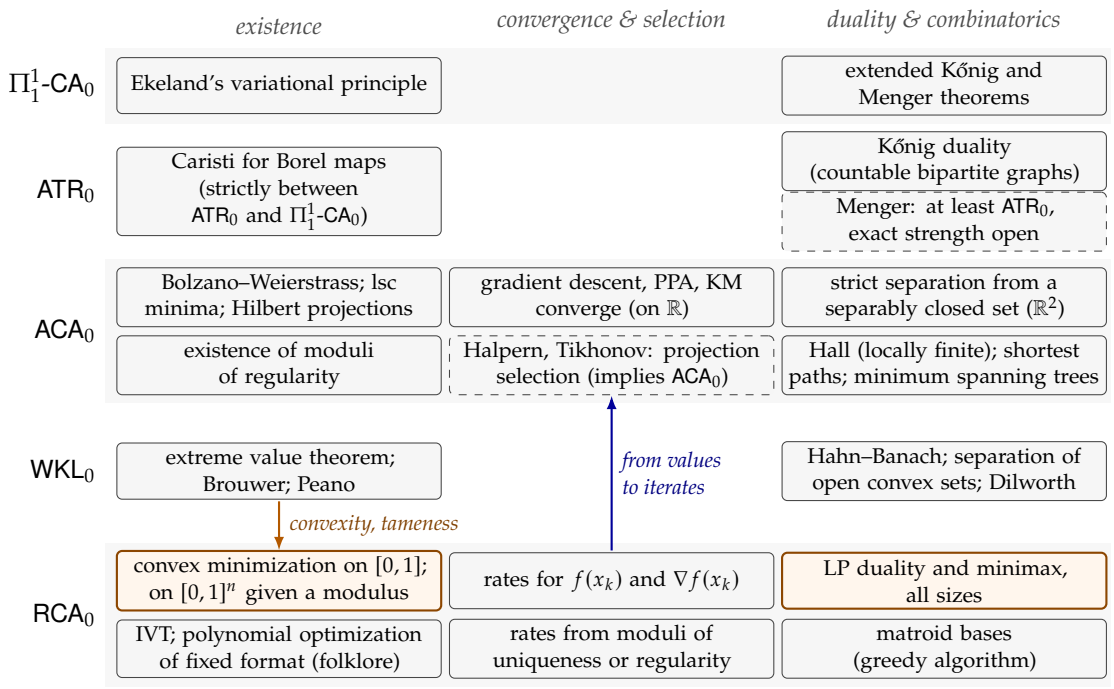


Figure 1. The map. Principles of optimization placed at the subsystem of second-order arithmetic to which they are equivalent over RCA$_0$ (in the bottom row: provable in RCA$_0$). Solid boxes are equivalences or provability results, dashed boxes are one-sided bounds, and orange boxes are results proved in this paper (Theorems 3.2 and 6.1). The Weihrauch counterparts of the rows are, roughly: non-uniformly computable problems such as IVT and XC$_n$ (RCA$_0$), closed choice C$_{[0,1]} \equiv_{\mathrm{sW}}$ WKL (WKL$_0$), and lim (ACA$_0$). Several placements depend on how sets and functions are coded; see the text for exact statements and sources.

convergence of gradient descent for convex L-smooth functions is equivalent to ACA$_0$, the strength of the monotone convergence theorem (Theorem 4.1, implicit in work of Kohlenbach and Neumann). The reason is not that minimizers are hard to find, but that the algorithm commits to a particular one, in dimension one the projection of its starting point onto the solution set, and this selection is as hard as the halting problem; anchored methods make the same choice in every dimension (Theorem 4.3). Uniqueness and error bounds remove the cost (Theorem 4.4). Between rates and convergence, the iterates of gradient descent carry a dimension-dependent but otherwise uniform bound on fluctuations (Theorem 4.7), and in dimension at least two the coordinates of their limits on computable instances are exactly the d.c.e. reals (Theorem 4.9).

3. *Duality climbs the hierarchy with the infinitary content of the certificate* (Section 6). Finite linear programming duality is provable in RCA$_0$ (Theorem 6.1); separation and Hahn–Banach are equivalent to WKL$_0$, and strict separation from a separably closed set to ACA$_0$; Kőnig's duality theorem for countable bipartite graphs is equivalent to ATR$_0$, and the extended forms of Kőnig's and Menger's theorems reach Π^1_1-CA$_0$.

Two further parts complete the picture. Section 5 is a guide to proof mining in optimization and explains, through Kohlenbach's metatheorems, why dimension-free complexity bounds

Result	Nearest prior work	What is added
Theorem 3.2(a): RCA_0 proves that convex functions on $[0,1]$ attain their minima	informal remark of Friedman [66]; IVT and the mean value theorem in RCA_0 [79, 158]	a proof, without any modulus hypothesis
Theorem 3.2(b): the same on $[0,1]^n$, given a modulus	computable points of convex sets [122]	reduction to dimension one; the modulus is the remaining gap (Theorem 8.1)
Theorem 6.1: LP duality and minimax in RCA_0, all sizes	Farkas $\Leftrightarrow$ LPO in BISH [17]; Weihrauch degree of linear inequalities [142]	the sign-pattern argument; uniform in the size
Theorem 3.5: Tarski transfer	folklore (the reals of RCA_0 form a real closed field)	optimization corollaries for fixed formats
Theorem 3.3: $\mathsf{CMin}_n \equiv_{\mathsf{sW}} \mathsf{XC}_n$	implicit in Neumann's constructions [137]	explicit statement and a direct proof
Theorem 3.4: $\mathsf{KKT}_1 \equiv_{\mathsf{sW}} \mathsf{IVT}$	$\mathsf{BFT}_1 \equiv_{\mathsf{sW}} \mathsf{IVT}$ [28, 137]	the exact KKT problem; Theorem 8.3
Theorem 4.7: fluctuations and dimension	path-length bounds [20, 46, 78]; norm non-convergence in ℓ^2 [9]	proof-mining reading; finite-dimensional lower bound; comparison with generic bounds
Theorem 4.9: limits of gradient descent	[3, 137, 171]	the characterization for gradient descent

Table 1. Results proved in this paper and their nearest predecessors.

are so common in convex optimization: they are what proofs that never use compactness of the ambient space must produce. Section 7 turns to the finite world, where bounded arithmetic plays the role of second-order arithmetic: the NP search problems provably total in T^1_2 are exactly those reducible to local search problems [32], finding approximate stationary points of smooth functions (equivalently, approximate fixed points of gradient descent) is complete for $\mathsf{CLS} = \mathsf{PPAD} \cap \mathsf{PLS}$ [54], and cutting planes, branch-and-cut and sum-of-squares are proof systems.

What is new. The paper is partly a survey and partly a research article. The survey part restates known calibrations in the language of optimization, with precise references; in particular Theorems 4.1, 4.3 and 4.4 are known in substance from the work of Kohlenbach, López-Acedo, Nicolae and Neumann. Table 1 lists the results that we prove, with the nearest prior result for each, as far as we could determine it. The two that we regard as genuinely new theorems are Theorem 3.2(a) and Theorem 6.1; the others are observations or corollaries whose interest lies in the connections they make.

Five principles. The results suggest five principles, which we state informally and revisit in Section 8.

P1. *Convexity and tameness substitute for compactness.* They lower the cost of existence from WKL_0 to RCA_0 (for convex functions: in dimension one, and in each fixed dimension given a modulus of uniform continuity; for polynomial problems: format by format), in classical reverse mathematics as in Bishop's constructive mathematics.

P2. *Selection is the price of convergence.* Gradient descent (in dimension one) and anchored methods (in every dimension) choose a particular minimizer, the projection of the

starting point or anchor, and this choice is lim-hard even when finding some minimizer is non-uniformly computable.

P3. *Uniqueness and regularity are computational resources.* A modulus of uniqueness or regularity is exactly the datum that turns convergence into a rate.

P4. *Bounds on values are cheap, convergence of iterates is not.* The uniform rates for function values and gradients follow from finitely many inequalities and have Π^0_1 proofs; convergence of iterates is a Π^0_3 statement and comes, in general, only with metastability. (This is a heuristic, not a theorem: logical form alone does not decide whether a bound exists.)

P5. *Dimension is the price of compactness.* Bounds extracted from proofs that use compactness of the space depend on its dimension, and for gradient descent this dependence is unavoidable.

Organization. Section 2 introduces the logical toolkit for optimizers and the convex toolkit for logicians. Sections 3 and 4 treat existence and convergence, Section 5 proof mining, Section 6 duality, and Section 7 the feasible world. Section 8 collects the principles and open problems. Appendix A records formalization details, Appendix B the computations behind Fig. 2, and Appendix C the full proof of Theorem 6.1.

AI disclosure. This paper was written with Claude Opus 5.5 (Anthropic), working as an agent in Claude Code. The author posed the question of how reverse mathematics bears on optimization. Claude chose the results, proved them, wrote the text and drew the figures. Separate Claude agents searched the literature, checked the citations against their sources, attacked the new results and reviewed the paper as referees, and the paper was revised to address their findings. The constructions behind Fig. 2 and the staircase bound were checked numerically. No part of this paper is formally verified. The author is responsible for the content.

2 Two toolkits

This section is written for two audiences. Sections 2.1 to 2.3 introduce the logical tools for readers from optimization; Section 2.4 fixes the convex-analytic facts used throughout for readers from logic. Experts can skip either part.

2.1 Reverse mathematics

Reverse mathematics, initiated by Friedman [65] and developed by Simpson and many others [52, 80, 158], asks which set-existence axioms are needed to prove a given theorem of ordinary mathematics. The setting is second-order arithmetic: variables range over natural numbers and over sets of natural numbers, and every structure of interest (a real number, a continuous function, a separable metric or Banach space, a countable graph) is coded by a set of naturals. Theorems are compared over a weak base theory, and the remarkable empirical fact is that most theorems are either provable in the base theory or *equivalent* to one of four stronger systems. These five systems are:

RCA_0 (recursive comprehension): Δ^0_1 comprehension and Σ^0_1 induction. Its minimal ω-model consists of the computable sets; it formalizes computable mathematics with classical logic.

WKL_0 RCA_0 plus weak Kőnig's lemma: every infinite binary tree has an infinite path. This is the compactness of Cantor space and of $[0, 1]$.

ACA_0 arithmetical comprehension. Over RCA_0 it is equivalent to the existence of the range of every one-to-one function, i.e. to the existence of the halting set relative to any set [158, Lemma III.1.3]. Its minimal ω-model consists of the arithmetical sets.

ATR_0 arithmetical transfinite recursion: arithmetical comprehension iterated along any countable well-ordering.

Π^1_1-CA_0 Π^1_1 comprehension: the existence of the hyperjump.

Each system proves the previous one, and none proves the next. Real numbers are coded by fast Cauchy sequences of rationals, and continuous functions between complete separable metric spaces by countable sets of "neighbourhood conditions" [158, II.6]. Statements of the form $x \le y$ or $f(x) \le c$ about reals are Π^0_1, and strict inequalities are Σ^0_1; this simple fact controls much of what follows.

2.2 Computable analysis and Weihrauch complexity

Reverse mathematics measures the axioms a theorem needs; computable analysis [25, 169] measures how hard its existential quantifier is to witness. A real x is *computable* if a machine can output rational approximations with error 2^{-n} on input n; *left-c.e.* (resp. *right-c.e.*) if it is the supremum (infimum) of a computable sequence of rationals; *d.c.e.* if it is a difference of two left-c.e. reals (these are called weakly computable in [3]; we avoid that name, which has other meanings in the Weihrauch literature); and *limit computable* (Δ^0_2) if it is the limit of a computable sequence of rationals. The computable reals are exactly those that are both left-c.e. and right-c.e., and $\{\text{left-c.e.}\} \cup \{\text{right-c.e.}\} \subsetneq \{\text{d.c.e.}\} \subsetneq \{\text{limit computable}\}$ [3]. A closed set $A \subseteq [0,1]^n$ is *co-c.e.* if its complement is the union of a computable sequence of rational open boxes; this is the information about A that is available when A is the solution set of a computable optimization problem.

Weihrauch reducibility [29] compares computational problems, viewed as multivalued maps from instances to solutions. A problem P is *Weihrauch reducible* to Q, written $P \le_{\mathrm{W}} Q$, if there are computable maps Φ, Ψ such that for every instance p of P, $\Phi(p)$ is an instance of Q and $\Psi(p, q)$ is a solution of p whenever q is a solution of $\Phi(p)$; for *strong* reducibility $\le_{\mathrm{sW}}$, Ψ has no access to p. The problems used in this paper are the following.

- lim: given a convergent sequence in Baire space (or in $\mathbb{R}$), output its limit. It is the Weihrauch counterpart of ACA_0 (more precisely, of a single Turing jump).
- MCT: given a nondecreasing sequence in $[0,1]$, output its supremum. $\mathsf{MCT} \equiv_{\mathrm{sW}} \mathsf{lim}$ [27, Fact 11.26].
- C_X (closed choice): given a nonempty co-c.e. closed $A \subseteq X$, output a point of A. $\mathsf{C}_{[0,1]} \equiv_{\mathrm{sW}} \mathsf{WKL}$, and finding a maximizer of a continuous function on $[0,1]$ is strongly equivalent to it [29, Theorems 7.23 and 8.3], [22].
- CC_X (connected choice) and XC_n (convex choice): C_X restricted to connected sets, resp. to convex subsets of $[0,1]^n$. Since the closed connected and the closed convex subsets of $[0,1]$ are the closed intervals, $\mathsf{XC}_1 = \mathsf{CC}_{[0,1]} \equiv_{\mathrm{sW}} \mathsf{IVT}$ [23, Theorem 6.2], [29, Theorem 7.34]. The hierarchy XC_n was studied by Le Roux and Pauly [121] and Kihara and Pauly [92].

A problem is *non-uniformly computable* if every computable instance has a computable solution. IVT and all XC_n have this property; WKL and lim do not.

2.3 Proof mining

Proof mining, in Kohlenbach's formulation [98, 100], extracts explicit quantitative information from proofs that do not visibly contain it. Its tools are proof interpretations (Gödel's functional interpretation, Kreisel's no-counterexample interpretation, and their monotone variants), and its guarantees are *logical metatheorems* [69, 97]: if a statement of the form $\forall x\, \exists n\, A(x,n)$, with A suitably restricted and x ranging over a (possibly abstract) metric, normed or Hilbert space, is provable in a suitable formal system, then there is a computable bound on n that depends only on *majorants* of the data (norm bounds, Lipschitz constants, moduli), not on the data themselves, and not on the space.

Convergence is not of this form. "(x_k) is Cauchy" reads $\forall \varepsilon\, \exists n\, \forall i,j \geq n\ \|x_i - x_j\| \leq \varepsilon$, a Π^0_3 statement (the matrix $\|x_i - x_j\| \leq \varepsilon$ is Π^0_1), and a rate of convergence (a bound on n) is in general not computable. Its no-counterexample interpretation is *metastability*, popularized by Tao [164]: for every $\varepsilon > 0$ and $F\colon \mathbb{N} \to \mathbb{N}$ there is n with $\|x_i - x_j\| < \varepsilon$ for $i,j \in [n, n+F(n)]$. This is $\forall\exists$ in form, and metatheorems produce rates of metastability $\Phi(\varepsilon, F)$ that are uniform in the above sense. Between rates and metastability lie bounds on *fluctuations* (see Section 4.4) and *learnability with bounded mind changes* [5, 112]. Two further notions recur: a *modulus of uniqueness* for F with unique zero x^* is a δ with $|F(x)| < \delta(\varepsilon) \Rightarrow \|x - x^*\| < \varepsilon$ [93]; a *modulus of regularity* is a ρ with $|F(x)| < \rho(\varepsilon) \Rightarrow \operatorname{dist}(x, F^{-1}(0)) < \varepsilon$ [114].

2.4 Convex toolkit

Definition 2.1. Let $\mathcal{H}$ be $\mathbb{R}^d$ or a separable Hilbert space and $L > 0$. A *convex L-smooth pair* is a pair of continuous maps $f : \mathcal{H} \to \mathbb{R}$, $g\colon \mathcal{H} \to \mathcal{H}$ such that for all $x, y \in \mathcal{H}$

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

Classically, (1) holds if and only if f is convex and differentiable with $\nabla f = g$ and g is L-Lipschitz [136, Theorem 2.1.5]. In second-order arithmetic, (1) quantifies over reals, but by continuity it is equivalent, provably in RCA_0, to the same inequalities for all rational points, a Π^0_1 condition on the codes; no theory of derivatives is needed. We write $S = \operatorname{argmin} f$ and $f^* = \inf f$, and study *gradient descent with step* $1/L$:

$$x_{k+1} = x_k - \tfrac{1}{L}\, g(x_k). \tag{2}$$

When $S \neq \emptyset$ we write $D = \operatorname{dist}(x_0, S)$; in statements meant for RCA_0, where this distance need not exist as a real, we use $\|x_0 - z\|$ for some $z \in S$ instead.

Lemma 2.2. *Let (f, g) be a convex L-smooth pair and let (x_k) follow (2). Then:*

(a) $f(x_{k+1}) \leq f(x_k) - \frac{1}{2L}\|g(x_k)\|^2$;

(b) *for every z:* $\|x_{k+1} - z\|^2 \leq \|x_k - z\|^2 + \frac{2}{L}\big(f(z) - f(x_{k+1})\big)$;

(c) *if $z \in S$, then* $\sum_{j=1}^{k}(f(x_j) - f^*) \leq \frac{L}{2}\|x_0 - z\|^2$ *and* $f(x_k) - f^* \leq L\|x_0 - z\|^2/(2k)$;

(d) $\|x_{k+1} - x_k\|$ *is non-increasing, and* $\|g(x_k)\| \leq 2L\|x_0 - z\|/(k-1)$ *for $z \in S$ and $k \geq 2$.*

Items (a) and (b) hold, with $1/L$ replaced by t_k, for variable steps $t_k \in (0, 1/L]$. Each item is provable in RCA_0.

Proof. (a) is the right inequality of (1) with $x = x_k$, $y = x_{k+1}$. For (b), expand $\|x_{k+1} - z\|^2 = \|x_k - z\|^2 - \frac{2}{L}\langle g(x_k), x_k - z\rangle + \frac{1}{L^2}\|g(x_k)\|^2$, bound the middle term with the left inequality of (1), $\langle g(x_k), x_k - z\rangle \geq f(x_k) - f(z)$, and the last term with (a). For (c), take $z \in S$ in (b), sum, and use that $f(x_j)$ is non-increasing. For (d), (1) implies co-coercivity, $\langle g(x) - g(y), x - y\rangle \geq \frac{1}{L}\|g(x) - g(y)\|^2$ (apply the right inequality of (1) to $y \mapsto f(y) - \langle g(x), y\rangle$ as in [136, Theorem 2.1.5]), so $I - g/L$ is nonexpansive and the steps do not grow. With $m = \lceil k/2 \rceil$, (a) and (c) then give $(k - m)\|g(x_{k-1})\|^2/(2L) \leq f(x_m) - f^* \leq L\|x_0 - z\|^2/(2m)$, and $m(k - m) \geq (k - 1)^2/4$. In RCA_0, each item is a Π^0_1 statement about reals proved by Π^0_1 induction on k. ∎

Taking z in S in (b) shows that the iterates are *Fejér monotone* with respect to S: $\|x_{k+1} - z\| \leq \|x_k - z\|$ for every $z \in S$. Taking z to be a later iterate gives the self-contraction property of Theorem 4.5.

3 Existence: what it costs to have a minimizer

The first question an optimizer asks is whether the problem has a solution. The classical answer is compactness. This section shows that compactness is expensive, and that the two structural assumptions that optimization relies on most, convexity and semialgebraic tameness, make it unnecessary: convexity in dimension one, and in each fixed dimension once a modulus of uniform continuity is given; tameness format by format.

3.1 The baseline: compactness costs weak Kőnig's lemma

Fact 3.1.

(a) *Over* RCA_0, WKL_0 *is equivalent to the maximum principle: every continuous* $f : [0,1] \to \mathbb{R}$ *attains its supremum* [158, Theorem I.10.3 and §IV.2]. *Each of "every such* f *is bounded" and "every such* f *is uniformly continuous" is also equivalent to* WKL_0.

(b) *Finding a maximizer of* $f \in C[0,1]$ *is strongly Weihrauch equivalent to* $\mathsf{C}_{[0,1]} \equiv_{\mathrm{sW}} \mathsf{WKL}$ [22], [29, Theorem 8.3]. *There is a computable* $f : [0,1] \to \mathbb{R}$ *that attains its maximum only at non-computable points (Lacombe; Specker [162]).*

(c) *Brouwer's fixed point theorem for* $[0,1]^n$, $n \geq 2$, *is equivalent to* WKL_0 [155] *and to* WKL *in the Weihrauch lattice* [28]; *for* $n = 1$ *it is provable in* RCA_0 *and Weihrauch equivalent to* IVT. *Peano's existence theorem for ordinary differential equations is equivalent to* WKL_0, *while the existence of Osgood's* maximal *solution is equivalent to* ACA_0 [156].

Item (c) already shows a pattern that will recur: existence of *some* solution costs WKL_0, while existence of a solution that is *extremal* among all solutions (a selection) costs ACA_0.

3.2 Convexity collapses the cost

For convex functions the situation changes completely. The minimizers of a convex function form a convex set, and convex sets are much easier to hit than arbitrary closed sets.

Theorem 3.2.
(a) RCA_0 *proves that every convex continuous function* $f : [0,1] \to \mathbb{R}$ *attains its minimum.*
(b) *For each standard* $n \geq 1$, RCA_0 *proves that every convex function* $f : [0,1]^n \to \mathbb{R}$ *with a modulus of uniform continuity attains its minimum.*

Here "for each standard n" means that there is a separate proof for each concrete number $n = 1, 2, 3, \ldots$; the statement with n as a variable is Theorem 8.1. Compare with Theorem 3.1(a): in dimension one, convexity drops the logical cost from WKL_0 to RCA_0. In dimension $n \geq 2$ the modulus in (b) is a genuine hypothesis: for general continuous functions its existence is equivalent to WKL_0, and whether convexity alone provides it is part of Theorem 8.1. What (b) shows is that, once a modulus is available, convexity removes every other use of compactness. Friedman remarked informally that maximum principles become provable in RCA_0 under "geometric" hypotheses of this kind [66]; we did not find a published proof. A sibling result is that the mean value theorem holds in RCA_0 [79].

Proof. (a) *A modulus.* Let f be convex and continuous on $[0,1]$. Convexity gives $f \leq M_+ := \max(f(0), f(1))$ and $f(x) \geq 2f(\frac{1}{2}) - f(1-x) \geq 2f(\frac{1}{2}) - M_+$, so we can fix a rational $M \geq \max(1, 2(M_+ - f(\frac{1}{2})))$, which bounds the oscillation of f. By the three-chord inequality, f is (M/δ)-Lipschitz on $[\delta, 1-\delta]$. The code of f (in the sense of [158, II.6]) contains, for every ε, a neighbourhood condition witnessing continuity at 0; searching for it gives a radius $r_0(\varepsilon)$ with $|f(x) - f(0)| < \varepsilon$ for $x \in [0, r_0(\varepsilon)]$, and similarly $r_1(\varepsilon)$ at 1. Given ε, put $\delta = \min(\frac{1}{2}, r_0(\varepsilon/2), r_1(\varepsilon/2))$ and $\eta = \min(\delta/2, \varepsilon\delta/(2M))$, and let $|x - y| < \eta$. If $\min(x, y) < \delta/2$, both points lie in $[0, \delta]$ and $|f(x) - f(y)| < \varepsilon$ by continuity at 0; the case $\max(x, y) > 1 - \delta/2$ is symmetric; otherwise both points lie in $[\delta/2, 1 - \delta/2]$, where f is $(2M/\delta)$-Lipschitz, and again $|f(x) - f(y)| < \varepsilon$. So f has a modulus of uniform continuity, and $m = \inf f$ exists.

A minimizer. We argue by cases; the case distinction is classical logic, not comprehension. If $f(q) = m$ for some rational q, we are done. Otherwise $f(q) > m$ for every rational q. For rational q consider the Σ^0_1 statements

$$R(q) : \ \exists r \in \mathbb{Q} \cap (q, 1] \ f(r) < f(q), \qquad \Lambda(q) : \ \exists r \in \mathbb{Q} \cap [0, q) \ f(r) < f(q).$$

At least one holds, because $f(q) > m = \inf_{r \in \mathbb{Q}} f(r)$. They cannot both hold, because $r_1 < q < r_2$ with $f(r_1), f(r_2) < f(q)$ contradicts convexity. So $\{q : R(q)\}$ exists by Δ^0_1 comprehension, and bisection ($l_0 = 0$, $u_0 = 1$; replace l by the midpoint c if $R(c)$, otherwise replace u by c) gives nested intervals with $R(l_k)$ and $\Lambda(u_k)$; note that $R(0)$ and $\Lambda(1)$ hold. Let x be the common limit. If $f(x) > m$, pick a rational r with $f(r) < f(x)$, say $r > x$. For large k we have $u_k < r$, and $\Lambda(u_k)$ gives $r' < u_k$ with $f(r') < f(u_k)$; convexity on $[r', r]$ gives $f(u_k) \leq \max(f(r'), f(r))$, hence $f(u_k) \leq f(r)$, and in the limit $f(x) \leq f(r) < f(x)$, a contradiction. The case $r < x$ is symmetric.

(b) By induction on the standard number n. Write points of $[0,1]^n$ as (t, y) with $t \in [0,1]$. The function $\varphi(t) = \inf_{y \in [0,1]^{n-1}} f(t, y)$ exists in RCA_0 (it is computed from the modulus ω of f by finite grids), has modulus $\varepsilon \mapsto \omega(\varepsilon/2)$, and is convex, because partial minimization preserves convexity. By (a), φ attains its minimum at some t^*, and $\varphi(t^*) = \inf f$. The function $f(t^*, \cdot)$ is convex on $[0,1]^{n-1}$ with modulus ω, so by the induction hypothesis it attains its minimum at some y^*, and $f(t^*, y^*) = \varphi(t^*) = \inf f$. ∎

The proof of (a) is the classical proof that the intermediate value theorem holds in RCA_0 [158, II.6.6], transplanted from sign changes to slopes. It is non-uniform in an essential way: whether some rational is a minimizer cannot be decided from f. The uniform content is measured by Weihrauch reducibility, and here the answer can be read off from Neumann's work on fixed points of nonexpansive maps.

Proposition 3.3. *Let CMin_n be the problem of finding a minimizer of a convex $f \in C([0,1]^n)$. Then $\mathsf{CMin}_n \equiv_{\mathrm{sW}} \mathsf{XC}_n$ for every $n \geq 1$; in particular $\mathsf{CMin}_1 \equiv_{\mathrm{sW}} \mathsf{CC}_{[0,1]} \equiv_{\mathrm{sW}} \mathsf{IVT}$. Consequently the problems CMin_n form a strictly increasing chain below WKL, and every computable convex function on $[0,1]^n$ has a computable minimizer.*

Proof. ($\mathsf{CMin}_n \leq_{\mathrm{sW}} \mathsf{XC}_n$.) From f one computes $m = \min f$, and $\operatorname{argmin} f = \{x : f(x) \leq m\}$ is co-c.e. closed, uniformly: its complement is the union of the rational boxes B with $\min_{\bar{B}} f > m$, which can be enumerated. It is nonempty and convex.

($\mathsf{XC}_n \leq_{\mathrm{sW}} \mathsf{CMin}_n$.) This direction is implicit in Neumann's proof that convex choice reduces to the Browder–Göhde–Kirk fixed point problem [137, Theorem 5.1]: his map $g = \sum_k 2^{-k-1} P_{h_k}$, a weighted average of projections onto half-spaces with intersection A, equals $I - \nabla\Phi$ for the convex 1-smooth function $\Phi = \sum_k 2^{-k-1} \frac{1}{2} \operatorname{dist}(\cdot, h_k)^2$, whose minimizers are exactly the points of A. For completeness we give a direct construction. Let $A \subseteq [0,1]^n$ be nonempty, closed and convex, given by an enumeration $U_0, U_1, \ldots$ of rational open boxes with union $[0,1]^n \setminus A$. The set $K_s = [0,1]^n \setminus (U_0 \cup \cdots \cup U_s)$ is a finite union of closed, possibly degenerate, rational boxes (cut every axis at all box end points), so $P_s = \operatorname{conv} K_s$ is the convex hull of finitely many rational points. The P_s decrease and contain A. They also shrink to A: given $\delta > 0$, compactness gives s with $K_s \subseteq A_\delta := \{x : \operatorname{dist}(x, A) \leq \delta\}$, and since A_δ is convex, $P_s \subseteq A_\delta$. Let

$$f(x) = \sum_{s \geq 0} 2^{-s-1} \operatorname{dist}(x, P_s).$$

Each summand is convex and 1-Lipschitz, and the distance from a rational point to a rational polytope is computable exactly (it solves a convex quadratic program with rational data), so f is computable, uniformly in the enumeration. Moreover $f \geq 0$ and $f(x) = 0$ iff $x \in \bigcap_s P_s = A$, so every minimizer of f lies in A.

The remaining statements are known facts about convex choice: $\mathsf{XC}_1 = \mathsf{CC}_{[0,1]} \equiv_{\mathrm{sW}} \mathsf{IVT}$ [23, Theorem 6.2], [29, Theorem 7.34]; $\mathsf{XC}_n <_{\mathrm{W}} \mathsf{XC}_{n+1}$ [121]; and every nonempty co-c.e. closed convex subset of $\mathbb{R}^n$ contains a computable point [122, Theorem 5.2]. For a computable convex f on $[0,1]^n$, the last statement also follows from Theorem 3.2(b): such an f has a computable modulus of uniform continuity, so (b) applies in the ω-model of computable sets, whose reals are the computable reals. ∎

The first-order optimality condition behaves in the same way. We record it here, although it is the computable counterpart of the problem studied in Section 7.2.

Proposition 3.4. *Let KKT_n be the problem: given $f \in C^1([0,1]^n)$ together with its gradient, find $x \in [0,1]^n$ with $x = P_{[0,1]^n}(x - \nabla f(x))$, i.e. a Karush–Kuhn–Tucker point of $\min_{[0,1]^n} f$. Then $\mathsf{KKT}_1 \equiv_{\mathrm{sW}} \mathsf{IVT}$, and $\mathsf{KKT}_n \leq_{\mathrm{sW}} \mathsf{BFT}_n$ for all n.*

Proof. $\mathsf{KKT}_n \leq_{\mathrm{sW}} \mathsf{BFT}_n$: the KKT points are the fixed points of the continuous self-map $x \mapsto P_{[0,1]^n}(x - \nabla f(x))$ of the cube. $\mathsf{KKT}_1 \leq_{\mathrm{sW}} \mathsf{IVT}$: the function $G(x) = x - P_{[0,1]}(x - f'(x))$ is continuous with $G(0) \leq 0 \leq G(1)$, and its zeros are the KKT points. Extend G linearly to $[-1, 2]$ with $G(-1) = -1$ and $G(2) = 1$; the extension is negative on $[-1, 0)$, positive on $(1, 2]$, and changes sign strictly, so IVT applies. $\mathsf{IVT} \leq_{\mathrm{sW}} \mathsf{KKT}_1$: given h with $h(0) < 0 < h(1)$, let $f(x) = \int_0^x h$. Then 0 and 1 are not KKT points ($f'(0) < 0$, $f'(1) > 0$), and the interior KKT points are the zeros of h. ∎

So in dimension one the exact KKT problem is as easy as finding a zero, and every computable instance has a computable solution. Since $\mathsf{BFT}_n \equiv_{\mathrm{sW}} \mathsf{WKL}$ for $n \geq 2$ [28], while gradient fields are very special vector fields, it is natural to ask whether KKT_n is strictly below BFT_n in higher dimensions, as CLS is (conjecturally) strictly below PPAD in the finite world. This is Theorem 8.3.

3.3 Tameness collapses the cost too

A second structural assumption makes compactness unnecessary for a completely different reason. Polynomial and semialgebraic optimization problems of fixed format are statements of first-order real algebra, and Tarski's theorem turns them into consequences of finitely many algebraic axioms, each of which RCA_0 can verify.

Proposition 3.5 (Tarski transfer; folklore). *Let θ be a sentence in the first-order language of ordered rings that is true in $\mathbb{R}$. Then RCA_0 proves θ, with its quantifiers interpreted over the coded real numbers. In particular, for each standard n and d, RCA_0 proves:*

(a) *every polynomial of degree at most d in n variables attains its minimum on $[0, 1]^n$, and, for each standard k, on every nonempty bounded set defined by at most k non-strict polynomial inequalities of degree at most d;*

(b) *linear programming duality, and von Neumann's minimax theorem, for real data of size $m \times n$, for each standard m;*

(c) *every finite game with at most n players and n pure strategies each has a mixed Nash equilibrium.*

Proof. The theory RCF of real closed fields is complete [165], so $\mathsf{RCF} \vdash \theta$. A formal derivation uses finitely many axioms of RCF: the ordered field axioms, the existence of square roots of non-negative elements, and the existence of roots of polynomials of finitely many odd degrees. RCA_0 proves each of these for its real numbers (the last two by the intermediate value theorem, which holds in RCA_0 [158, II.6.6]). Equality of reals is a congruence for $+$, $\times$ and $<$ provably in RCA_0, so the derivation can be replayed in RCA_0 with first-order quantifiers read as quantifiers over codes of reals. Statements (a)–(c), for fixed sizes, are first-order sentences true in $\mathbb{R}$. ∎

The proposition is a meta-theorem: it gives one proof for each format, and says nothing about the statements with n and d as variables, which are not first-order sentences. For linear programming we prove the uniform statement in Theorem 6.1; for polynomials and games the uniform question is open (Theorem 8.1). The fact behind the proposition, that the reals of RCA_0 form a real closed field, is standard; we record the optimization corollaries because they show

that tameness, like convexity, replaces a compactness argument, here by a finite algebraic one.

3.4 Uniqueness

Uniqueness makes solutions computable. A unique point of a co-c.e. closed subset of a computably compact space is computable, uniformly in the set [23, Corollary 8.9]; so a continuous function on $[0,1]^n$ with a unique minimizer has a computable minimizer, uniformly. In reverse mathematics the situation is more delicate: WKL_0 proves that a function with at most one zero has a modulus of uniqueness, and already for Lipschitz functions on $[0,1]$ with exactly one zero the *uniform* uniqueness statement implies WKL_0 [101, Theorem 4.1]. In Bishop's constructive mathematics, the existence of minima of uniformly continuous functions with at most one minimum point is equivalent to the fan theorem [12, 18], and for quasi-convex functions it is provable outright [16]. Uniqueness is also the classical source of explicit rates, through Kohlenbach's moduli of uniqueness [93]; we return to this in Sections 4.3 and 5.5.

3.5 Beyond compact spaces and continuous functions

Optimization in infinite dimensions, with lower semicontinuous objectives or with constraints given only by negative information, climbs the hierarchy quickly.

Fact 3.6. *Over* RCA_0*:*

(a) Lower semicontinuity. *"Every lower semicontinuous $f\colon [0,1] \to [0,\infty)$ attains its minimum" is equivalent to* ACA_0*, for lower semicontinuous functions coded by labelled rational balls as in [58] (combine Theorem 8.2, Proposition 9.3 and Lemma 7.2 there); the calibration depends on this coding.*

(b) Projections. *The existence of the orthogonal projection onto every closed subspace (the closure of the span of a sequence) of a separable Hilbert space is equivalent to* ACA_0*; the existence of the distance $d(x, M)$ is equivalent to the existence of the projection $P_M x$, provably in* RCA_0 *[6, Theorems 12.3 and 12.5]. The existence of the norm of every bounded linear functional, which is the optimal value $\sup_{\|x\|\le 1}|f(x)|$, and the Riesz representation theorem are each equivalent to* ACA_0 *[6, Theorem 13.4].*

(c) Ekeland's variational principle *[53]. For a lower semicontinuous potential f on a complete separable metric space, the principle has a weak ("free") form, asserting ε-critical points, and a strong ("localized") form, asserting them below a given point. Fernández-Duque, Shafer and Yokoyama [58, Section 11] prove: the free form for continuous f on compact spaces is equivalent to* WKL_0*; weakening the localized form in exactly two of the three ways (free, compact space, continuous f), or restricting it to compact spaces alone, gives* ACA_0*; and the free form, the localized form, and even the localized form for continuous functions are equivalent to* Π^1_1-CA_0.

(d) Caristi's fixed point theorem, *a close relative of Ekeland's principle, is equivalent to* WKL_0 *or* ACA_0 *in restricted forms, and for Baire or Borel maps to the transfinite leftmost path principle* TLPP_0*, which lies strictly between* ATR_0 *and* Π^1_1-CA_0 *[59, 144, 167].*

(e) Games. ACA_0 *proves that every continuous game on compact spaces has a mixed Nash equilibrium, and weak compactness of probability measures on compact spaces is equivalent to* ACA_0 *[143].*

Item (c) is the most striking calibration in the area. The localized Ekeland principle for *continuous* functions is, in the authors' words, the first statement about continuous functions stemming from analysis that is equivalent to Π^1_1-CA_0. Since Ekeland's principle is the engine of much of modern variational analysis (existence of approximate minimizers with sharp optimality conditions, subdifferential calculus, the Bishop–Phelps theorem), its strength is a measure of how far variational analysis is from the predicative mathematics of ACA_0. Item (b) shows a phenomenon that we meet again in Theorem 4.4: for strictly convex problems, the optimal *value* determines the optimal *point* effectively.

3.6 The constructive picture

Bishop-style constructive mathematics measures the same theorems by intuitionistic logic and omniscience principles, and its results run parallel to the ones above. That every uniformly continuous positive function on $[0, 1]$ has a positive infimum is a well-known equivalent of a form of Brouwer's fan theorem (it fails in recursive mathematics [90]; see [15] and its references); for quasi-convex functions on compact convex subsets of $\mathbb{R}^n$ this is provable without the fan theorem [14], and convexity yields a "convex fan theorem" (co-convex bars are uniform) [15]. A separating hyperplane theorem for convex sets and the fundamental theorem of asset pricing are each equivalent to Markov's principle [13], and Farkas' lemma as an exclusive alternative is equivalent to LPO, although enough of it survives to give constructive versions of linear programming optimality criteria and of the minimax theorem [17]. The common moral is principle P1 of the introduction: convexity does the work that compactness, or omniscience, does in general.

4 Convergence: what it costs to get there

Section 3 showed that for convex problems the *existence* of a minimizer is almost free. This section shows that *convergence to* a minimizer is not. The main character is the plainest algorithm there is: gradient descent (2) with constant step $1/L$ on a convex L-smooth function. Every optimizer knows that it converges to a minimizer when one exists (in finite dimensions; weakly in Hilbert space), and that $f(x_k) - f^* \le LD^2/(2k)$ with $D = \operatorname{dist}(x_0, \operatorname{argmin} f)$. The contrast between these two statements is the subject of this section. The rate for $f(x_k) - f^*$ is *uniform*: it depends only on L, D and k. For the iterates themselves there is no rate depending only on L, D and ε, and, as we shall see, not even a computable one in general.

The first three results below are known in substance. Their key constructions and arguments are due to Kohlenbach [101], to Kohlenbach, López-Acedo and Nicolae [114] and to Neumann [137], who work with moduli of regularity and with Picard and Krasnosel'skiĭ–Mann iterations of firmly nonexpansive maps; the formulation as convergence theorems for gradient descent and anchored methods, and some of the reversals (for instance the one for the Tikhonov path), are ours. In one dimension firmly nonexpansive maps are *exactly* the gradient-descent maps: $T\colon \mathbb{R} \to \mathbb{R}$ is firmly nonexpansive if and only if it is nondecreasing and 1-Lipschitz, if and only if $T = I - \frac{1}{L} f'$ for a convex f with L-Lipschitz derivative. We restate these three results in the language of optimization because their meaning for optimization is the point of this paper, and we give short proofs, since the one-dimensional instance makes the mechanism transparent.

4.1 Gradient descent is arithmetical comprehension

Theorem 4.1 (Kohlenbach; Neumann). *Over* RCA_0 *the following are equivalent.*

(1) ACA_0.

(2) *For every* $d \geq 1$, *every convex L-smooth pair* (f, g) *on* $\mathbb{R}^d$ *that has a minimizer, and every* $x_0 \in \mathbb{R}^d$, *the gradient-descent sequence* (2) *converges.*

(3) *Statement* (2) *for* $d = 1$.

(4) *For every nondecreasing nonexpansive* $T\colon \mathbb{R} \to \mathbb{R}$ *that has a fixed point and every* x_0, *the Picard sequence* $(T^k x_0)_k$ *converges.*

Moreover, the limit problem $\mathsf{GDLim}_d\colon (f, g, L, x_0) \mapsto \lim_k x_k$ *satisfies* $\mathsf{GDLim}_d \equiv_{\mathrm{sW}} \mathsf{lim}$ *for every* $d \geq 1$.

The reversal (3)⇒(1) is contained in Kohlenbach's proof that moduli of regularity require ACA_0 [101, proof of Theorem 4.4], which uses the map $\frac{1}{2}(I + T)$ for the map T below, that is, gradient descent on the same function with step $1/4$; the Weihrauch statement for Krasnosel'skiĭ and Krasnosel'skiĭ–Mann iterations on $[0, 1]$ is [137, Proposition 3.20 and Corollary 3.18]. The function that drives all the lower bounds in this section is

$$f_y(x) = \sum_{s \geq 0} 2^{-s-1}\big((y_s - x)_+\big)^2 + \big((x-1)_+\big)^2, \tag{3}$$

built from a nondecreasing sequence (y_s) in $[0, 1]$. It is convex and C^1 with $f_y'(x) = -\sum_s 2^{-s}(y_s - x)_+ + 2(x-1)_+$, which is 2-Lipschitz, and $\operatorname{argmin} f_y = [\alpha, 1]$ with $\alpha = \sup_s y_s$. All of f_y is computable from the data, but the late data enter the gradient with weight 2^{-s}: gradient descent feels a jump that occurs at stage s only through a pull of order 2^{-s}. This is why no computable rate can exist.

Proof. *(1)⇒(2).* Work in ACA_0. Fix a minimizer z. By Theorem 2.2(b) the distances $\|x_k - z\|$ do not increase, so the sequence is bounded. By the Bolzano–Weierstrass theorem, which holds in ACA_0 [158, III.2], a subsequence x_{k_i} converges to some $\bar{x}$. By Theorem 2.2(c), $f(x_k) \to f(z)$, so $f(\bar{x}) = f(z)$ by continuity and $\bar{x}$ is a minimizer. By Theorem 2.2(b) with $z = \bar{x}$, the distances $\|x_k - \bar{x}\|$ do not increase and have $\liminf 0$, hence tend to 0.

(2)⇒(3) is trivial. *(3)⇔(4).* Adding the two lower inequalities of (1) at (x, y) and (y, x) shows that g is nondecreasing, and adding the two upper ones shows that g is L-Lipschitz; at a minimizer z, the upper inequality with $y = z - g(z)/L$ gives $g(z) = 0$. So $T = I - g/L$ is nondecreasing and 1-Lipschitz and fixes the minimizers. Conversely, given such a T, put $L = 1$, $g = I - T$ and $f(x) = \int_0^x g$; then $f(y) - f(x) - g(x)(y - x) = \int_x^y (g(t) - g(x))\,dt \in [0, (y-x)^2/2]$, so (f, g) is a convex 1-smooth pair, and a fixed point of T is a minimizer of f. Riemann integrals of Lipschitz functions exist in RCA_0.

(3)⇒(1). Over RCA_0, ACA_0 is equivalent to the existence of the range of every one-to-one $h\colon \mathbb{N} \to \mathbb{N}$ [158, Lemma III.1.3]. Given h, let $y_s = \sum_{i \leq s} 2^{-h(i)-1}$, a nondecreasing sequence in $[0, 1)$, and consider $f = f_y$ with $L = 2$. The pair (f, f') exists in RCA_0 (the series converge uniformly on bounded sets with explicit moduli), and 1 is a minimizer since $f \geq 0 = f(1)$. Let x_k be the gradient-descent sequence from $x_0 = 0$, and $T = I - f'/2$, which is nondecreasing. By (3), $x_k \to x_\infty$.

- $Tx_\infty = x_\infty$ by continuity, so $f'(x_\infty) = 0$, which forces $y_s \le x_\infty \le 1$ for all s.
- If $Tz = z$ and $z \ge 0$, then $x_k \le z$ for all k, by Π^0_1 induction: $x_0 = 0 \le z$ and $x_{k+1} = Tx_k \le Tz = z$.
- Suppose $y_s \le x_\infty - \varepsilon$ for all s. Then $z = x_\infty - \varepsilon$ satisfies $z \ge y_0 > 0$ and $f'(z) = 0$, so $Tz = z$, and the previous item gives $x_\infty \le x_\infty - \varepsilon$, which is absurd. Hence $y_s \to x_\infty$.

Now $m \in \mathrm{ran}(h)$ if and only if $m \in \{h(0), \dots, h(n)\}$ for any n with $x_\infty - y_n < 2^{-m-1}$, because a later value $h(i) = m$ would add 2^{-m-1} to the sum. Such an n is found by a Δ^0_1 search (Appendix A), so $\mathrm{ran}(h)$ exists.

Weihrauch. $\mathsf{GDLim}_d \le_{\mathrm{sW}} \mathsf{lim}$ because the iterates are uniformly computable from the input and converge. Conversely, given a nondecreasing sequence (y_s) in $[0,1]$, the function f_y is uniformly computable and gradient descent from 0 converges to $\sup_s y_s$ by the argument above. So the monotone convergence problem satisfies $\mathsf{MCT} \le_{\mathrm{sW}} \mathsf{GDLim}_1 \le_{\mathrm{sW}} \mathsf{GDLim}_d$, and $\mathsf{MCT} \equiv_{\mathrm{sW}} \mathsf{lim}$ [27, Fact 11.26]. ∎

Remark 4.2 (What the theorem says, and what it does not).

(a) The constructed function has a whole interval $[\alpha, 1]$ of minimizers. This is forced: if the minimizer is unique, the limit and a rate are computable (Theorem 4.4).

(b) For a computable increasing sequence with non-computable supremum (a Specker sequence [161]), the iterates have no computable rate of convergence, although $f(x_k) - f^* \le LD^2/(2k)$ and $|f'(x_k)| \le 2LD/(k-1)$ hold as always [114, Remark 4.10], [137, Corollary 3.19]. Figure 2 shows what this looks like: after a late jump of the data, every standard certificate of near-optimality is tiny while the iterate is still far from its limit.

(c) The equivalence depends on the coding of real numbers. Freund, Pischke and Uftring [64] code reals by Cauchy sequences *without* a rate; in their setting RCA_0 proves the monotone convergence theorem [64, Corollary 7.6], and they discuss when iterates of continuous maps exist [64, Corollary 7.15]. We expect, but have not checked, that the one-dimensional case of (3) becomes provable in RCA_0 in their coding. The equivalence with ACA_0 is a statement about the existence of limits *with* a modulus, which is what a numerical analyst would call "knowing the limit".

(d) The same proofs apply to the proximal point algorithm (which is gradient descent on the Moreau envelope), to Krasnosel'skiĭ–Mann iterations of firmly nonexpansive maps, and to projected gradient methods. An earlier calibration of the same kind concerns linear averaging: the mean ergodic theorem for nonexpansive linear maps on Hilbert space is equivalent to ACA_0 [6, Theorem 14.1].

4.2 The selection principle

Theorems 3.2 and 4.1 look contradictory at first sight: a minimizer of a convex function is cheap to *have* but expensive to *reach*. The resolution is that a convergent algorithm does more than find a minimizer: it *selects* one, namely its limit, as a function of the data. In the proof of Theorem 4.1, gradient descent from 0 selects the left end point α of $\operatorname{argmin} f_y = [\alpha, 1]$, which is the projection of the starting point onto the solution set, and that end point carries the halting problem. Neumann proved the general form of this observation for fixed-point iterations on $[0,1]$: computing the projection of a point onto the fixed-point set of a nonexpansive map, computing the limit of a large class of iterations, and computing a rate of convergence for them,

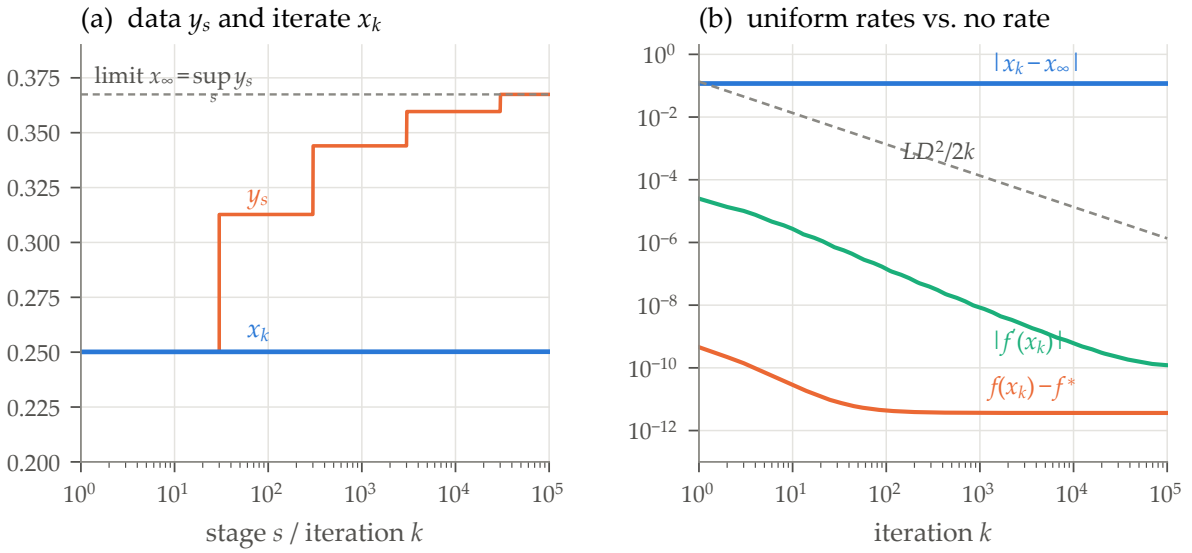


Figure 2. Gradient descent on the function f_y of (3) for data y_s that start with a jump at stage 0 and receive late jumps at stages $s = 30, 300, 3000, 30000$, computed in 80-digit arithmetic (Appendix B). (a) The data y_s (plotted against s) and the iterate x_k (plotted against k). (b) The value gap and the gradient obey the uniform bounds and are below 10^{-4} after the first step, yet the iterate stays at distance about 0.12 from its limit: the jump at stage 30 pulls with weight 2^{-30}, and crossing it alone takes more than 10^9 steps.

are all Weihrauch equivalent to lim [137, Proposition 3.20].

Anchored methods make the selection explicit. They converge *strongly*, even in Hilbert space, and their limit is the projection of the anchor u onto the solution set: the Halpern iteration $x_{k+1} = \beta_{k+1}u + (1-\beta_{k+1})Tx_k$ with $\beta_k \to 0$, $\sum_k \beta_k = \infty$ and $\sum_k |\beta_{k+1} - \beta_k| < \infty$ (Wittmann's theorem [170]), and Tikhonov regularization $x_\lambda = \operatorname{argmin}\{f + \frac{\lambda}{2}\|\cdot - u\|^2\}$ as $\lambda \downarrow 0$.

Theorem 4.3 (Selection; after Neumann). *On the class of convex 2-smooth functions on $\mathbb{R}$ with a minimizer, the following problems are strongly Weihrauch equivalent to* lim*: given f and u, find $P_{\operatorname{argmin} f}(u)$; find the limit of the Halpern iteration with $\beta_k = 1/(k+1)$, anchor u and starting point $x_0 = u$ for $T = I - \frac{1}{2}f'$; find $\lim_{\lambda \downarrow 0} x_\lambda$, where $x_\lambda = \operatorname{argmin}\{f + \frac{\lambda}{2}(\cdot - u)^2\}$ is the Tikhonov path. Over RCA_0, the convergence of either of the last two (for all such f and u) implies ACA_0.*

Proof. Use f_y with $u = x_0 = 0$, so that $P_{\operatorname{argmin} f_y}(0) = \alpha = \sup_s y_s$. All three limits are α, and the upper bounds are as in Theorem 4.1; each point x_λ of the Tikhonov path is the unique zero of the λ-strongly increasing function $h = f' + \lambda(I - u)$, which is found uniformly by trisection (since $h(m+\delta) - h(m-\delta) \geq 2\lambda\delta$, one of the strict tests $h(m-\delta) < 0$, $h(m+\delta) > 0$ succeeds), so the path exists in RCA_0. For the reverse-mathematical statement we cannot use the classical identification of the limit with α, since α need not exist in RCA_0; instead we repeat the comparison argument of Theorem 4.1. For Halpern with $x_0 = u = 0$: if $Tz = z \geq 0$ then $x_k \leq z$ for all k, since $x_{k+1} = (1-\beta_{k+1})Tx_k \leq (1-\beta_{k+1})z \leq z$; and $|x_{k+1} - Tx_k| \leq \beta_{k+1}|Tx_k| \to 0$, so a limit x_∞ is a fixed point. For Tikhonov: $f' + \lambda I$ is increasing and non-negative at every fixed point $z \geq 0$ of T, so $0 \leq x_\lambda \leq z \leq 1$; and $f'(x_\lambda) = -\lambda x_\lambda \to 0$, so a limit x_∞ satisfies $f'(x_\infty) = 0$. In both cases the remaining two items in the proof of Theorem 4.1 go through unchanged. ∎

Together, Theorems 3.2, 4.1 and 4.3 isolate the cost of convergence: it is the cost of the particular *choice* an algorithm makes. A convergent method defines a map $(f, x_0) \mapsto \lim_k x_k$, which is a realizer of convex minimization. Finding *some* minimizer is cheap: the multivalued problem CMin_1 is equivalent to IVT (Theorem 3.3), and on the hard family f_y the constant choice 1 is a minimizer. No single-valued realizer is as cheap as the problem itself (a single-valued map below $\mathsf{C}_{2^{\mathbb{N}}}$ is computable [26]), and the realizers computed by gradient descent in dimension one and by anchored methods in every dimension select the projection of the starting point or of the anchor onto the solution set, which is as complex as the halting problem. (In dimension $d \geq 2$ the limit of gradient descent need not be that projection, but the one-dimensional case already embeds.) This is principle P2 of the introduction, in the precise form that Theorems 4.1 and 4.3 prove; whether every limit-computable realizer of convex minimization is lim-hard is Theorem 8.2. Kohlenbach's recent work on moduli of regularity, which shows how such moduli compute zeros and, for convex zero sets, the zero of minimal norm [106], is closely related.

4.3 What restores computability: uniqueness and regularity

There are two classical ways to remove the choice: make the minimizer unique, or make the distance to the solution set controlled by the objective. In proof mining these are the *modulus of uniqueness* [93] and the *modulus of regularity* [114]; in optimization they are strong convexity and error bounds (Hölder, Łojasiewicz, Kurdyka–Łojasiewicz, Hoffman).

Proposition 4.4 (Kohlenbach, López-Acedo, Nicolae; Kohlenbach). *Let (f, g) be a convex L-smooth pair on $\mathbb{R}^d$ with $S = \operatorname{argmin} f \neq \emptyset$, and $D = \operatorname{dist}(x_0, S)$.*

(a) *If $S = \{x^*\}$, then x^* and a rate of convergence of (2) are computable from (f, g, L, x_0), uniformly on the class of such instances.*

(b) *If ρ is a modulus of regularity, i.e. $f(x) - f^* < \rho(\varepsilon)$ implies $\operatorname{dist}(x, S) < \varepsilon$ (a modulus of uniqueness when S is a singleton), then $\|x_k - \lim_j x_j\| < \varepsilon$ for all $k > LD^2/(2\rho(\varepsilon/2))$.*

(c) *Over RCA_0, the existence of a modulus of regularity for every Lipschitz $F\colon [0,1] \to \mathbb{R}$ with a zero is equivalent to ACA_0, while the mere regularity statement (without a modulus) is equivalent to WKL_0; the modulus-of-regularity problem is Weihrauch equivalent to $\widehat{\mathsf{LPO}} \equiv_{\mathsf{W}} \lim$ [27, Fact 3.5], and the modulus-of-uniqueness problem is computable [101]. (In RCA_0 the existence of a modulus of uniqueness needs WKL_0, as in Section 3.4: the compactness of $[0,1]$ costs WKL_0 in RCA_0, but is free for computable functions in the Weihrauch setting.)*

Proof. (a) If S is bounded and nonempty, all sublevel sets of f are bounded [151, Corollary 8.7.1]. Search for a rational R with $\min\{f(y) : \|y - x_0\| = R\} > f(x_0)$; the condition is Σ^0_1 and some R satisfies it, and by convexity $\{f \leq f(x_0)\} \subseteq \bar{B}(x_0, R)$. On this ball x^* is the unique point of a co-c.e. closed set, hence computable [23, Corollary 8.9]. A rate is $\varepsilon \mapsto$ the first n found with $\|x_n - x^*\| < \varepsilon$; it is valid for all $k \geq n$ because $\|x_k - x^*\|$ does not increase. (b) is [114, Theorem 4.1] with the rate $f(x_k) - f^* \leq LD^2/(2k)$ of Theorem 2.2(c): once $\operatorname{dist}(x_k, S) < \varepsilon/2$, Fejér monotonicity gives $\|x_m - P_S x_k\| \leq \operatorname{dist}(x_k, S)$ for all $m \geq k$, hence $\|x_k - x_\infty\| \leq 2\operatorname{dist}(x_k, S) < \varepsilon$. ∎

Three comments place this in the optimization landscape. First, the rate in (a) is computable

but not bounded in terms of L, D and ε alone: $f(x) = \mu x^2/2$ shows that no bound in terms of L, D and ε alone exists. Uniform bounds are exactly what moduli provide: strong convexity with constant μ is a modulus of uniqueness $\delta(\varepsilon) = \mu\varepsilon^2/2$; a Hölder error bound $f - f^* \geq c\,\mathrm{dist}(\cdot, S)^p$ is a modulus of regularity $\rho(\varepsilon) = c\,\varepsilon^p$; for least squares, Hoffman's bound [84] gives $p = 2$. Kohlenbach, López-Acedo and Nicolae show that weak sharp minima, metric subregularity and error bounds are all instances [114, Section 3]. Second, uniqueness is not *necessary* for computability of a given instance (a solution interval with rational end points gives computable limits); the precise statements are the uniform ones in (a) and (c), and the existential ones in Theorem 4.2(b). Third, o-minimal geometry explains why regularity is common in practice. Definable (for instance semialgebraic) functions satisfy the Kurdyka–Łojasiewicz inequality [21, 118], which is a modulus of regularity in disguise, and this is the basis of the convergence theory of descent methods for tame problems [4]. Model theory thus guarantees, for a large class of problems, the *existence* of exactly the datum that Theorem 4.4(c) shows to be non-computable in general. It does not make the datum computable from the data: on the tame family $f_c(x) = \frac{c}{2}x^2$, $c \in [0,1]$, with $L = 1$ and $x_0 = 1$, the limit is 0 for $c > 0$ and 1 for $c = 0$, so even here the limit operator computes LPO.

4.4 Between convergence and rates: fluctuations and metastability

When no rate of convergence exists, proof mining asks for weaker, finitary information [5, 112, 164]. A sequence (x_k) has N *ε-fluctuations* if there are $i_1 < j_1 \leq i_2 < j_2 \leq \cdots \leq i_N < j_N$ with $\|x_{i_l} - x_{j_l}\| \geq \varepsilon$ for each l. A *rate of metastability* is a function $\Phi(\varepsilon, F)$ such that for every $\varepsilon > 0$ and every $F\colon \mathbb{N} \to \mathbb{N}$ there is $n \leq \Phi(\varepsilon, F)$ with $\|x_i - x_j\| < \varepsilon$ for all $i, j \in [n, n + F(n)]$. Metastability is the Kreisel no-counterexample interpretation of the Cauchy property, and by Kohlenbach's metatheorems it can be extracted from proofs of convergence. A bound on the number of fluctuations is stronger: it yields a rate of metastability, and it makes the limit *learnable with a bounded number of mind changes*.

For Fejér monotone sequences in totally bounded spaces, which include gradient descent and the proximal point algorithm in $\mathbb{R}^d$, Kohlenbach, Leuştean and Nicolae extracted rates of metastability and bounds on mind changes governed by the modulus of total boundedness; for plain Fejér monotonicity in a ball of radius b in $\mathbb{R}^d$ the number of mind changes is of order $(8\sqrt{d}\,b/\varepsilon)^d$ [113, Theorem 5.1 and Example 2.8]. The structure of gradient descent gives a bound that is linear in D/ε. Since the constant C_d below may itself grow exponentially in d, the honest gain is that the degree in D/ε drops from d to 1; for D/ε small compared with d the generic bound can be better. Throughout, C_d denotes the constant in the length bound for self-contracted curves recalled below; it depends only on d, and for gradient descent with step $1/L$ an explicit admissible value can be read off from the path-length bound of Gupta, Balakrishnan and Ramdas [78], which is of order $2^{O(d^2)}$. The key fact is geometric.

> **Lemma 4.5** (Self-contraction [20, 78]). *Let f be convex and L-smooth on a Hilbert space and $x_{k+1} = x_k - t_k \nabla f(x_k)$ with $0 < t_k \leq 1/L$. Then $\|x_j - x_k\| \leq \|x_i - x_k\|$ for all $i \leq j \leq k$. The same holds for the proximal point algorithm with any parameters [20, Corollary 3.8] (see also [46]). No minimizer is needed.*

Proof. For $m < k$ we have $f(x_k) \leq f(x_{m+1})$, so Theorem 2.2(b) with $z = x_k$ gives $\|x_{m+1} - x_k\| \leq$

$\|x_m - x_k\|$; chain these for $m = i, \dots, j-1$. ∎

Self-contracted curves were introduced by Daniilidis, Ley and Sabourau [45], and Daniilidis, David, Durand-Cartagena and Lemenant [46, Theorem 3.3] proved, without any continuity assumption, that a bounded self-contracted curve in $\mathbb{R}^d$ has length at most C_d times its diameter; Gupta, Balakrishnan and Ramdas [78] and Böhm and Daniilidis [20] used this to bound the path length of gradient descent. Read through the lens of proof mining, the length bound becomes a uniform fluctuation bound, and its failure in infinite dimensions becomes a lower bound.

The proof of Theorem 4.7(c) below uses the following observation.

Lemma 4.6 (Finite-dimensional reflection). *Let f be convex and L-smooth on a Hilbert space $\mathcal{H}$, let (x_k) follow (2), let z be a minimizer and $K \in \mathbb{N}$, and let $V = \operatorname{span}\{x_0, \dots, x_K, z\}$. Then $f|_V$ is convex and L-smooth on $V \cong \mathbb{R}^{\dim V}$ with gradient $P_V \nabla f$, z minimizes $f|_V$, and gradient descent for $f|_V$ from x_0 produces $x_0, \dots, x_K$.*

Proof. For $x, y \in V$, $\langle \nabla f(x), y - x\rangle = \langle P_V \nabla f(x), y - x\rangle$, so (1) holds for $(f|_V, P_V \nabla f)$. The point $z \in V$ minimizes f on $\mathcal{H}$, hence on V. For $j < K$, $\nabla f(x_j) = L(x_j - x_{j+1}) \in V$, so $P_V \nabla f(x_j) = \nabla f(x_j)$ and the two iterations agree. ∎

Proposition 4.7 (Fluctuations and dimension). *Let f be convex and L-smooth on $\mathbb{R}^d$ with a minimizer, $D = \operatorname{dist}(x_0, \operatorname{argmin} f)$, and (x_k) given by (2) (or by the proximal point algorithm).*

(a) *For every $\varepsilon > 0$, (x_k) has at most $2C_d D/\varepsilon$ ε-fluctuations.*

(b) *$\Phi(\varepsilon, F) = \tilde{F}^{(N)}(0)$, with $\tilde{F}(n) = n + F(n)$ and $N = \lfloor 2C_d D/\varepsilon \rfloor$, is a rate of metastability; it depends on f, L and x_0 only through D.*

(c) *There is no dimension-free bound: no function $B\colon (0,\infty) \to \mathbb{N}$ satisfies, for all d, all convex 1-smooth f on $\mathbb{R}^d$ with a minimizer z, all x_0 and all $\varepsilon > 0$, that (2) has at most $B(\|x_0 - z\|/\varepsilon)$ ε-fluctuations. Likewise, no rate of metastability depends only on $\|x_0 - z\|$, ε and F.*

Proof. (a) By Fejér monotonicity every x_k lies in $\bar{B}(z, D)$ for $z = P_{\operatorname{argmin} f}(x_0)$, so the orbit has diameter at most $2D$, and by Theorem 4.5 and [46] its length is at most $2C_d D$. If $i_1 < j_1 \le \cdots \le i_N < j_N$ witness N ε-fluctuations, then $\varepsilon \le \|x_{i_l} - x_{j_l}\| \le \sum_{i_l \le k < j_l} \|x_{k+1} - x_k\|$, and these index ranges are disjoint, so $N\varepsilon \le 2C_d D$.

(b) Let $n_0 = 0$ and $n_{l+1} = \tilde{F}(n_l)$. If no n_l with $l \le N$ works, each interval $[n_l, n_l + F(n_l)]$ contains $i_l < j_l$ with $\|x_{i_l} - x_{j_l}\| \ge \varepsilon$, and $j_l \le n_{l+1} \le i_{l+1}$. These are $N+1$ ε-fluctuations, contradicting (a).

(c) Bauschke, Matoušková and Reich [9], building on Hundal [86], give two closed convex sets K, H in ℓ^2 with $0 \in K \cap H$ and a point x_0 such that the iterates of the averaged projection $\frac{1}{2}(P_K + P_H)$ do not converge in norm. Since $x - P_C x = \nabla \frac{1}{2} d_C^2(x)$, we have $\frac{1}{2}(P_K + P_H) = I - \nabla f$ with $f = \frac{1}{4}(d_K^2 + d_H^2)$, a convex function with 1-Lipschitz gradient and $\operatorname{argmin} f = K \cap H \ni 0$; so these iterates are (2) with $L = 1$ (see the proof of [20, Proposition 4.4]). A sequence that is not Cauchy has, for some $\varepsilon > 0$, infinitely many ε-fluctuations. If B existed, choose K so large that $x_0, \dots, x_K$ contain $B(\|x_0\|/\varepsilon) + 1$ of them. By Theorem 4.6 with $z = 0$, these are also the first $K+1$ iterates of gradient descent for a convex 1-smooth function on a space of

dimension at most $K+2$ with minimizer 0, a contradiction. For metastability, suppose $\Phi(\varepsilon, F, b)$ were a dimension-free rate, valid whenever $\|x_0 - z\| \le b$. For every ε and F, reflect the first $K = \Phi + \max_{n\le\Phi} F(n)$ iterates of the ℓ^2 sequence; this shows that Φ is a rate of metastability for the ℓ^2 sequence itself, which is therefore Cauchy, a contradiction. ∎

That gradient descent in ℓ^2 can fail to converge in norm, and hence admits no fluctuation bound there, follows from [9] and the proof of [20, Proposition 4.4]; it is also known that self-contracted curves in Hilbert space can have infinite length [163]; the restriction to the span of the iterates is the standard device of performance estimation [166]. Theorem 4.7(c) combines them into a finite-dimensional, quantitative statement: the constant C_d in (a) must tend to infinity, and so must any rate of metastability. An explicit family shows how the number of fluctuations grows with the dimension.

Proposition 4.8 (Staircase). *For every $d \ge 1$ and $\varepsilon > 0$ there are a separable convex 1-smooth function on $\mathbb{R}^d$ with unique minimizer 0 and a point x_0 with $\|x_0\| = 2\sqrt{d}\,\varepsilon$ such that gradient descent has at least d ε-fluctuations. Consequently $C_d \ge \sqrt{d}/2$.*

Proof. Let $a = 2\varepsilon$, $s_i = a2^{-i}$ and $f(x) = \sum_{i=1}^{d} h_{s_i}(x_i)$, where h_s is the Huber function with $h_s'(t) = \max(-s, \min(s, t))$. From $x_0 = (a, \dots, a)$, gradient descent moves coordinate i by s_i per step and reaches 0 exactly at step $T_i = 2^i$. During the steps $k \in [T_{i-1}, T_i]$ (with $T_0 = 2^0 = 1$) only the coordinates $j \ge i$ move, each by $a2^{i-1-j}$, so $\|x_{T_{i-1}} - x_{T_i}\|^2 = a^2 \sum_{j=i}^{d} 4^{i-1-j} \ge a^2/4$. The d consecutive windows give d ε-fluctuations. Since all coordinates decrease monotonically, the orbit has diameter exactly $\|x_0\| = 2\sqrt{d}\,\varepsilon$, so its length is at least $d\varepsilon$ and at most $C_d \cdot 2\sqrt{d}\,\varepsilon$; hence $C_d \ge \sqrt{d}/2$. ∎

The lower bound $C_d \gtrsim \sqrt{d}$ matches the known lower bound for path lengths of gradient descent; the best known upper bounds are of the form $2^{O(d \log d)}$ for gradient flow and for gradient descent with steps of order $1/(L\sqrt{d})$, and $2^{O(d^2)}$ for step $1/L$ [78]. The gap is part of Theorem 8.4.

Theorem 4.6 is the device behind performance estimation problems [51, 166]: the first K steps of a first-order method see only a $(K+2)$-dimensional space. Theorem 4.7 thus makes precise, for gradient descent, a phenomenon that the proof-mining metatheorems explain in general terms (Section 5.1): bounds extracted from proofs that do not use compactness of the ambient space are dimension-free, and bounds from proofs that do use it (the Bolzano–Weierstrass step in Theorem 4.1, or the length of self-contracted curves) depend on the dimension, here necessarily. The Halpern iteration, by contrast, converges strongly in Hilbert space by a proof that uses no norm compactness of the space (only weak sequential compactness or projections, which are available uniformly in every Hilbert space), and accordingly it has dimension-free rates of metastability [99, 107].

4.5 What gradient descent can converge to

Theorem 4.1 is a uniform statement: the limit operator of gradient descent is as hard as lim. Finite path length has a second, non-uniform consequence, which locates the *values* of this operator precisely. Recall from Section 2.2 the left-c.e., right-c.e. and d.c.e. reals. Ambos-Spies,

Weihrauch and Zheng [3] showed that the d.c.e. reals form a field and are exactly the limits of computable sequences of rationals with finite total variation; they form a proper subclass of the limit-computable (Δ^0_2) reals, which are the possible values of $\lim$ on computable inputs.

The following corollary is a direct consequence of known results: part (b) combines self-contraction with the characterization just quoted, and part (a) is in substance contained in the work of Zheng, Rettinger and Barmpalias on monotonically computable reals [171] and in Neumann's constructions [137, Corollary 3.18]. We state it because it gives a sharp answer to a natural question about gradient descent that, as far as we know, has not been asked.

Corollary 4.9 (The limits of gradient descent). *Call an instance (f, g, L, x_0) of (2) computable if f, g and x_0 are computable, L is rational, and f has a minimizer.*

(a) *In dimension 1, the limits of gradient descent on computable instances are exactly the reals that are left-c.e. or right-c.e.*

(b) *In every dimension d, every coordinate of the limit of a computable instance is d.c.e.*

(c) *In every dimension $d \geq 2$, every d.c.e. real is the first coordinate of the limit of a computable instance.*

Every computable instance has a computable minimizer, while the limit may be non-computable.

The step size matters in (b): if L is allowed to be a non-computable real, the limit can be far more complex. For instance, for $f(x) = \frac{1}{2}((x_1)_+)^2 + \frac{1}{10}((x_1 + 2x_2)_+)^2$ and $x_0 = (1, 1)$, every L in $[1 + 5^{-1/2}, 8/5]$ is admissible and the limit has first coordinate $\frac{2}{5} - \frac{4}{5L}$, which has the Turing degree of L.

Proof. (a) In dimension one $T = I - \frac{1}{L}g$ is nondecreasing, so the iterates are monotone, and a computable monotone sequence of computable reals has a left-c.e. or right-c.e. limit. Conversely, let α be left-c.e. and fix rationals $c < \alpha$ and $\lambda > \alpha - c$. Then $\alpha = c + \lambda \sup_s y_s$ for a computable nondecreasing sequence of rationals $y_s \in [0, 1)$. With f_y from (3), the function $\varphi(u) = \lambda^2 f_y((u - c)/\lambda)$ is convex with 2-Lipschitz derivative, and gradient descent with step $1/2$ from $u_0 = c$ satisfies $u_k = c + \lambda x_k$, where x_k is gradient descent for f_y from 0. So $u_k \to c + \lambda \sup_s y_s = \alpha$. Right-c.e. reals are handled by $u \mapsto -u$.

(b) By Theorem 4.5 and the length bound for self-contracted curves [46], the iterates have finite length. Replacing each coordinate $x_k^{(i)}$ by a rational r_k with $|r_k - x_k^{(i)}| < 2^{-k}$ gives a computable sequence of rationals with finite total variation and the same limit, so the limit is d.c.e. [3].

(c) Write $a = A - B$ with A, B left-c.e. Let φ be a one-dimensional instance from (a) with limit $\frac{5}{3}A$, and ψ one with limit $-\frac{5}{4}B$, both with 2-Lipschitz derivative. The matrix $Q = \frac{1}{5}\left(\begin{smallmatrix} 3 & 4 \\ 4 & -3 \end{smallmatrix}\right)$ is orthogonal and $Q^2 = I$. Let $F(x) = \varphi(u) + \psi(v)$, where $(u, v) = Qx$. Then F is convex with 2-Lipschitz gradient, and since an orthogonal change of variables commutes with (2) and a separable function decouples, gradient descent for F from $x_0 = Q(u_0, v_0)$ converges to $Q(\frac{5}{3}A, -\frac{5}{4}B)$, whose first coordinate is $\frac{1}{5}(3 \cdot \frac{5}{3}A - 4 \cdot \frac{5}{4}B) = A - B$. For $d > 2$, add $\frac{1}{2}\|\cdot\|^2$ in the remaining coordinates.

Finally, a computable instance with a minimizer has one in some rational box B; the restriction of f to B is a computable convex function whose minimizers are minimizers of f, so Theorem 3.3 gives a computable minimizer. The instance from (a) built from a non-computable left-c.e. real

has a non-computable limit, while argmin $f_y \ni 1$. ∎

Theorem 4.9 separates the two faces of Theorem 4.1. Uniformly, the limit operator of gradient descent is as hard as lim; but on computable inputs its values are confined to the d.c.e. reals, a class strictly below the limit-computable ones, and the reason is geometric: self-contracted orbits have finite length. In Hilbert space orbits need not have finite length, and whether the restriction persists for norm-convergent instances is Theorem 8.5. Combined with Theorem 4.7(a), the corollary also says that for gradient descent in $\mathbb{R}^d$ the passage from the computable data to the limit costs at most two monotone limits per coordinate, with at most $2C_d D/\varepsilon$ changes of mind at precision ε.

5 Proof mining: uniformity as complexity

Reverse mathematics tells us which convergence statements *cannot* come with computable rates. Proof mining tells us what *can* be extracted anyway, and how. This section is a guide to the parts of Kohlenbach's program that concern optimization; for full accounts see [98, 100, 105].

5.1 The logical form of a complexity bound

A typical worst-case bound of first-order convex optimization reads: for every Hilbert space $\mathcal{H}$, every convex L-smooth f on $\mathcal{H}$, every x_0 and every minimizer x^* with $\|x_0 - x^*\| \le D$, and every $\varepsilon > 0$, there is k with $f(x_k) - f(x^*) < \varepsilon$; moreover $k \le N(L, D, \varepsilon)$. The first half is a $\forall\exists$ statement over an abstract Hilbert space. The general logical metatheorems of Kohlenbach and Gerhardy–Kohlenbach [69, 97] say that whenever such a statement is proved in a suitable formal system (roughly: classical analysis in all finite types with an abstract Hilbert space added as a base type, plus dependent choice, but without axioms that make the space compact or separable), the second half follows automatically, with a bound N that depends only on majorants of the data: here L, D and ε. Three features of optimization practice, long known from the theory of oracle complexity, are thus *explained* by logic.

(i) *Dimension-freeness.* Bounds do not depend on the dimension, because the abstract space has none.

(ii) *Oracle-model uniformity.* Bounds do not depend on f beyond its constants, which is the premise of the oracle complexity theory of Nemirovski and Yudin [130].

(iii) *Independence of the solution.* Bounds depend on x^* only through D.

The metatheorems also indicate where their uniformity guarantee stops: if the proof uses compactness of the space (the Bolzano–Weierstrass step in Theorem 4.1), the extracted bound depends on a modulus of total boundedness, and if the statement is Π^0_3 (convergence), the extractable content is a rate of metastability. Metatheorems give sufficient conditions and cannot by themselves show that a better bound is impossible; Theorems 4.1 and 4.7 show that, for gradient descent, both limits are genuine.

Theorem 4.6 is a complement. Performance estimation problems [51, 166] compute exact worst cases of K steps of a first-order method by optimizing over Gram matrices of size $K + 2$; they are exact for every dimension at least $K+2$, because the iterates of a first-order method span a small subspace. Logically, this is a *finite-dimensional reflection principle* for first-order methods: if some instance in a Hilbert space violates a property of the first K steps, then some instance in

dimension $K + 2$ violates it. Metatheorems explain why bounds are dimension-free; reflection explains why the worst cases can be computed in finite dimension, and (Theorem 4.7(c)) why infinite-dimensional pathologies leave finite-dimensional traces.

5.2 A case study: rates for Halpern and Krasnosel'skiĭ–Mann iterations

The history of rates for the two basic fixed-point iterations shows how the two communities have interacted.

Krasnosel'skiĭ–Mann. For $x_{n+1} = (1-\lambda_n)x_n + \lambda_n T x_n$ with T nonexpansive on a bounded convex set C of a normed space, Baillon and Bruck [7] proved $\|x_n - Tx_n\| \le \operatorname{diam}(C)/\sqrt{\pi\lambda(1-\lambda)n}$ for constant λ, and conjectured a bound of the form $\kappa \operatorname{diam}(C)/\sqrt{\sum_{i\le n}\lambda_i(1-\lambda_i)}$ for general steps. Kohlenbach [95, 96] extracted from a proof of Borwein, Reich and Shafrir a rate of asymptotic regularity for general steps that depends on C only through a bound on its diameter and is independent of x_0 and T (exponential in $1/\varepsilon$). For general steps in Banach spaces this was the only known rate until Cominetti, Soto and Vaisman [40] proved the Baillon–Bruck conjecture with $\kappa = 1/\sqrt{\pi}$, crediting the uniformity to proof mining; the constant is sharp [30].

Halpern. For $x_{n+1} = \beta_n u + (1-\beta_n)Tx_n$, the first rate of asymptotic regularity was obtained by proof mining (Leuştean [125], exponential in $1/\varepsilon$), improved to a quadratic rate by Kohlenbach [99] and Kohlenbach–Leuştean [107] together with uniform rates of metastability for the strong convergence. (In this subsection a rate is called linear or quadratic according to the dependence of the number of iterations on $1/\varepsilon$; a linear rate means $\|x_k - Tx_k\| = O(1/k)$, not geometric convergence.) A linear rate was then found by Sabach and Shtern [152], the tight bound $\|x_k - Tx_k\| \le 2\|x_0 - x^*\|/(k+1)$ in Hilbert space by Lieder [128] using performance estimation, and the optimal rate over all Mann-type iterations in normed spaces by Contreras and Cominetti [41]. Lieder notes explicitly that the first result of this type came from proof mining; conversely, the Sabach–Shtern lemma was imported back into proof mining to obtain linear rates in W-hyperbolic spaces [36, 37].

The pattern is instructive. Logical methods reliably deliver *uniformity* and *first* bounds, often in generality (normed or geodesic spaces, general step sizes) that the optimization literature reaches later. Optimization methods (potential functions, performance estimation) deliver *optimal constants*. The two are complementary, and they meet in the dimension-freeness of the bounds: the metatheorems predict it, and performance estimation exploits it by computing worst cases in dimension $K + 2$.

5.3 Fejér monotonicity, regularity and metastability

Most convergence proofs in convex optimization have the same skeleton: the iterates are Fejér monotone with respect to the solution set, some residual tends to zero, and a compactness or regularity argument finishes the proof. Proof mining has analysed this skeleton once and for all. Kohlenbach, Leuştean and Nicolae extracted uniform rates of metastability for (generalized) Fejér monotone sequences whenever the convergence proof uses compactness, covering the proximal point algorithm and iterations of firmly nonexpansive, asymptotically nonexpansive and strictly pseudo-contractive maps in W-hyperbolic spaces [113]. Kohlenbach, López-Acedo and Nicolae showed that a modulus of regularity turns the same skeleton into a rate of convergence, and conversely that a uniform rate of convergence yields a modulus of

regularity [114]; Pischke extended this to a general notion of Fejér monotonicity [146], and Kohlenbach and Pinto revisited it recently [110]. In this light, Theorem 4.4 is the specialization to gradient descent of a general theorem, and Theorem 4.7 is a case where the specific geometry of an algorithm beats the general bound in its dependence on D/ε: for gradient descent in $\mathbb{R}^d$ the number of fluctuations is linear in D/ε, whereas the general compactness-based bound is of order $(8\sqrt{d}\,D/\varepsilon)^d$.

Proof mining has also produced explicit rates for many specific methods: Halpern-type proximal point algorithms [102, 104, 126], the proximal point algorithm in uniformly convex spaces [103], abstract proximal point algorithms [127], multi-parameter and viscosity schemes [50, 109], inconsistent feasibility problems [159], subgradient methods for equilibrium problems [147], set-valued monotone and accretive operators, for which new logical metatheorems were needed [111, 145], and continuous-time quasi-Fejér monotone dynamical systems such as gradient flows [63].

5.4 Stochastic optimization

Stochastic approximation is the newest frontier. Its convergence theorems rest on the Robbins–Siegmund theorem for almost-supermartingales, which implies Doob's convergence theorem and therefore admits no computable rates in general. Neri and Powell gave a quantitative Robbins–Siegmund theorem with explicit, learnable rates of uniform (metastable) convergence [133], building on their analysis of crossings and fluctuations of stochastic processes [132]. Neri, Pischke and Powell turned this into an abstract effective convergence theorem for stochastic approximation, covering Dvoretzky's theorem and stochastic quasi-Fejér monotone sequences [134], and into generalized fluctuation bounds under compactness [135]. Pischke and Powell obtained linear rates of asymptotic regularity for stochastic Halpern and Krasnosel'skiĭ–Mann schemes and translated them into oracle complexity bounds [148]; see [131] for an overview of proof mining in probability theory. The logical classification of the underlying almost-sure convergence theorems is still open (Theorem 8.9).

5.5 The origins: best approximation

Proof mining began, in part, with optimization. Best Chebyshev approximation by polynomials of degree $\leq n$ is a convex minimization problem with a unique solution, and the classical uniqueness proofs are ineffective. Kohlenbach [93, 94] extracted from them explicit moduli of uniqueness, which are uniform a priori estimates for the constants of strong unicity, improving the known estimates of Bridges by a factor n in the exponent; Kohlenbach and Oliva [108] carried out the analysis for L^1-approximation. In both cases the extracted moduli have the optimal dependence on ε [105]. In the language of Theorem 4.4, these are moduli of uniqueness for specific convex programs. A modulus of uniqueness δ turns any rate for the objective values into a rate for the iterates: if $f(x_k) - f^* \leq r(k)$ and $r(k) < \delta(\varepsilon)$, then $\|x_k - x^*\| < \varepsilon$; Theorem 4.4(b) is the special case of gradient descent.

6 Duality and certificates

Duality theorems assert the existence of a *certificate*: a dual solution, a separating hyperplane, a vertex cover, a cut. They are, formally, existence theorems, and the existence of a certificate behaves like the existence of an optimal solution: cheap in finite dimensions, expensive when the certificate is an infinite object chosen among infinitely many candidates.

6.1 Finite-dimensional duality

Finite linear programming is the base case. With real data it is not computable in any uniform sense: solving a bounded feasible system of linear inequalities, and finding a Nash equilibrium of a bimatrix game, are both Weihrauch equivalent to $\mathsf{AoUC}^*_{[0,1]}$, the finite parallelization of all-or-unique choice (for linear inequalities [142, Theorems 37 and 38] with [29, Proposition 7.60]; for Nash equilibria [142, Corollary 40] and [29, Theorem 8.5]), and no Banach–Mazur computable map produces maximizers of linear programs with computable data [123]. In Bishop's constructive mathematics, Farkas' lemma as an exclusive alternative is equivalent to LPO [17]. Classical reverse mathematics is kinder. We did not find the following statement in the literature.

Theorem 6.1. RCA_0 *proves linear programming strong duality: for all* m, n *and all real* $A \in \mathbb{R}^{m\times n}$, $b \in \mathbb{R}^m$, $c \in \mathbb{R}^n$, *if* $\{x \geq 0 : Ax \leq b\}$ *and* $\{y \geq 0 : A^\top y \geq c\}$ *are nonempty, then there are* x^* *and* y^* *in them with* $c^\top x^* = b^\top y^*$. *Consequently* RCA_0 *proves von Neumann's minimax theorem for real matrix games of every size.*

Proof sketch. Appendix C gives the details. Write the primal in equality form with slack variables, $[A\ I]\,(x, s) = b$. First obtain the signs of the b_i (by the comprehension step described below) and multiply the rows with $b_i < 0$ by -1; Phase I then adds m artificial variables and minimizes their sum. For each phase let M be the bordered matrix formed by the constraint matrix of that phase, the right-hand side, and the cost row of that phase. For a basis B, every entry of the simplex tableau, including the reduced costs and the objective value, is a quotient μ/Δ_B of minors of M, where $\Delta_B \neq 0$ is the basis determinant. Every test that the two-phase simplex method with Bland's rule [19] performs is therefore the sign of a minor, or of an expression $\mu_1\mu_2 - \mu_3\mu_4$ in minors (a ratio test, after multiplying out positive denominators; ties are tests "$= 0$"). Let $(v_i)_{i<N}$ be the finite list of all these reals, for both phases; its length N is bounded by an explicit function of m and n.

The only non-constructive step is the following. The reals v_i are computed from the data by explicit polynomial expressions, so the sequence $(v_i)_{i<N}$ exists in RCA_0 and "$v_i > 0$" is Σ^0_1 uniformly in i. By bounded Σ^0_1 comprehension, which is provable in RCA_0 [158, II.3], the finite sets $\{i < N : v_i > 0\}$ and $\{i < N : v_i < 0\}$ exist, so RCA_0 has the complete sign pattern σ of (v_i) as a finite object. The simplex method can now be run symbolically on σ: each tableau is represented by its basis, and each decision is read off σ. Because σ is the true sign pattern, this symbolic run is a genuine run of the simplex method on the real data. Bland's anticycling argument is a finite combinatorial argument about the run, valid in every ordered field and formalizable with Σ^0_1 induction, so each phase stops after at most $\binom{n+2m}{m}$ pivots. At the end of Phase I the optimal value is 0, since the primal is feasible, and artificial variables that remain

basic at level zero are pivoted out, which is possible because $[A\ I]$ has rank m. Phase II cannot report unboundedness, since by weak duality any dual feasible y bounds the objective. So it stops at an optimal basis B, where Cramer's rule gives x^* and $y^{*\top} = c_B^\top B^{-1}$ as explicit quotients of minors with non-zero denominators. Non-negative reduced costs say that y^* is dual feasible, and $c^\top x^* = c_B^\top B^{-1} b = b^\top y^*$ is an algebraic identity between explicitly given reals. Nothing beyond Σ^0_1 induction and bounded Σ^0_1 comprehension is used. For the minimax theorem, shift the payoff matrix by a rational constant that makes all entries positive and use the standard linear programming formulation of the game; the reduction is uniform in the size. ∎

The proof makes the three calibrations of finite duality fit together. Reverse mathematics, which has classical logic, obtains the finitely many signs it needs for free; constructive mathematics must assume them (whence LPO); and the Weihrauch degree measures how much discontinuous information about signs a uniform algorithm needs, which is a finite parallelization of a principle strictly below LPO. We expect that the same method, applied to Lemke–Howson pivoting [124] with lexicographic degeneracy resolution, shows that RCA_0 proves the existence of Nash equilibria of bimatrix games of every size, but we have not written out the details. For games with three or more players, whose equilibria can be irrational algebraic numbers of high degree, Theorem 3.5 gives the statement for each fixed format, the uniform statement is provable in WKL_0 via Brouwer's theorem, and the exact strength is open (Theorem 8.1). Berger and Svindland note that their constructive separation theorem and fundamental theorem of asset pricing can also be proved in RCA_0 [13, Remark 2].

6.2 Separation and Hahn–Banach

In infinite dimensions, certificates of optimality are continuous linear functionals, and their existence is the Hahn–Banach theorem. Over RCA_0, the Hahn–Banach theorem for separable Banach spaces (with $\|\tilde{f}\| \le 1$, not $= \|f\|$) is equivalent to WKL_0 [31], and so are the separation of an open convex set from a disjoint separably closed convex set and the strict separation of two disjoint open convex sets [85, Theorem 4.4]. Strict separation of a compact convex set from a disjoint separably closed convex set is equivalent to ACA_0, already in $\mathbb{R}^2$ [85, Theorem 5.1]: the difficulty is not the dimension but the coding: RCA_0 cannot prove that a set given as the closure of a sequence is closed in the coded sense, and this is where arithmetical comprehension enters. The existence of the norm of a functional, which is the value of an optimization problem, is again ACA_0 [6, Theorem 13.4]. In the Weihrauch lattice the Hahn–Banach theorem is equivalent to WKL [24, 70], and in Bishop's mathematics it is equivalent to LLPO [87].

These results resonate with a familiar experience in convex analysis: duality theorems are cheap when a constraint qualification provides an interior point (open sets, Slater's condition) and delicate when it does not. We conjecture that the logical counterpart of this experience is the gap between WKL_0 and ACA_0 that the separation theorems exhibit; making this precise for Lagrange and Fenchel duality is Theorem 8.6.

6.3 Infinite combinatorial optimization

Combinatorial optimization on countable structures provides the highest calibrations of all, and the most instructive. Table 2 collects them.

Principle (countable version)	Strength over RCA_0	Source
Matroid basis, decidable dependence (greedy algorithm)	RCA_0	[83]
Hall's theorem, bounded marriage problems	WKL_0	[82]
Hall's theorem, each boy knows finitely many girls	ACA_0	[82]
Hall's theorem under the B-bounded Hall condition	RCA_0 (sequential: ACA_0)	[68]
Dilworth's theorem (width k), Mirsky's theorem	WKL_0	[34, 81]
Stable marriage: highly bounded and effectively listed / listed / ordered	WKL_0 / ACA_0 / ATR_0	[34]
Shortest paths; minimum spanning trees (bounded weight)	ACA_0	[39]
Bases of countable vector spaces; e-matroid bases	ACA_0	[67, 83]
Perfect matchings (Steffens), locally finite / bounded	ACA_0 / WKL_0	[62]
Kőnig's duality theorem (bipartite graphs)	ATR_0	[1, 157]
Extended Kőnig duality; extended Menger theorem	Π^1_1-CA_0	[1, 154]
Menger's theorem (countable graphs)	between ATR_0 and TLPP_0	[154, 167]

Table 2. Reverse mathematics of combinatorial optimization on countable structures.

Three features of the table deserve comment. First, the role of a *bounding function* (a function bounding the neighbours of each vertex, or the preferences of each agent) is exactly the role of compactness in analysis: with it, existence is WKL_0; without it, ACA_0. Second, for matchings the step from ACA_0 to ATR_0 happens at *duality*: Hall's theorem asserts only a matching, while Kőnig's theorem asserts a matching and a cover that certify each other, and Aharoni, Magidor and Shore built a computable bipartite graph all of whose Kőnig covers are above every computable iterate of the Turing jump [1]. In the Weihrauch lattice, Kőnig's theorem is arithmetically equivalent to the two-sided problem ATR_2 [71, Section 9]. A primal–dual pair is a far more rigid object than either half. (Stable marriage reaches ATR_0 as well, but only in its "ordered" form, which encodes the comparability of well-orderings [34].) Third, the greedy algorithm is the combinatorial analogue of convexity: when it applies (matroids with decidable dependence), existence costs nothing.

Several natural cells are empty: countable max-flow/min-cut [2], matroid intersection for countable matroids [89], and the exact strength of Menger's theorem, which is open [154, 167]. Each of them generalizes Kőnig's theorem, so any formalization strong enough to imply it is at least ATR_0.

7 The feasible world: bounded arithmetic, TFNP and proof systems

Second-order arithmetic measures infinitary content. For finite optimization problems the right yardstick is complexity, and the logical counterpart is bounded arithmetic, where Cook's program of *bounded reverse mathematics* pairs complexity classes with weak theories and propositional proof systems [42, 138]. The correspondences in this section are tighter than in second-order arithmetic: they identify classes of search problems with exactly the existence principles that a theory can prove.

7.1 Local optimality is the strength of T_2^1

A local search problem asks for a local optimum of a cost function over a neighbourhood structure, both given by polynomial-time algorithms. Johnson, Papadimitriou and Yannakakis [88] collected these problems into the class PLS. The principle "every finite optimization problem has a local optimum" is a $\forall\Sigma_1^b$ statement provable in Buss's theory T_2^1, which has induction for NP properties, and Buss and Krajíček proved the converse: the Σ_1^b-definable multifunctions of T_2^1 are exactly the projections of PLS problems [32]; equivalently, the NP search problems provably total in T_2^1 are exactly those reducible to PLS. Beckmann and Buss extended this up the hierarchy, characterizing, for $1 \le i \le k+1$, the Σ_i^b-definable multifunctions of T_2^{k+1} by local search over Π_k^p-definable feasible sets [11], and Kołodziejczyk, Nguyen and Thapen characterized the NP search problems provably total in the weak second-order theory V_2^1 by an exponentially blown-up local improvement principle [116]. In this precise sense, *local optimality is the existence principle of* T_2^1, just as compactness is the existence principle of WKL_0. A global optimum, by contrast, cannot in general be recognized in polynomial time, so, unless P = NP, finding one is not an NP search problem.

Feasible versions of real analysis exist as well. Ferreira's theory BTFA and its extension by a form of weak Kőnig's lemma for trees given by bounded formulas have the polynomial-time functions as their provably total functions [60], and BTFA proves the intermediate value theorem, so that its real numbers form a real closed field [56].

7.2 Continuous local search and gradient descent

The continuous analogue of PLS is the class CLS of Daskalakis and Papadimitriou [47], designed to capture problems solvable by gradient-type local improvement, such as finding stationary points of smooth functions. Fearnley, Goldberg, Hollender and Savani proved that CLS = PPAD ∩ PLS, and that finding an ε-KKT point of a C^1 function with Lipschitz gradient on $[0,1]^2$, given by arithmetic circuits, is complete for this class, as are the problems of finding approximate fixed points of gradient descent [54, Theorems 5.1 and 6.1]. Even KKT points of quadratic programs over $[0,1]^n$ are CLS-complete [55]. Two neighbouring results calibrate the surroundings: approximate local min-max points of nonconvex–nonconcave smooth functions on polytopes are PPAD-complete [49], like Brouwer fixed points and Nash equilibria [35, 48, 140]; and approximate second-order stationary points are PLS-complete on the square according to a recent preprint [117].

The finite and the computable worlds agree in dimension one, where KKT points are found by bisection on the derivative. In the computable world the analogue of the KKT problem is exact rather than approximate, and in dimension one it is as easy as finding a zero: $\mathsf{KKT}_1 \equiv_{\mathrm{sW}} \mathsf{IVT}$ (Theorem 3.4). Whether the exact KKT problem stays strictly below the Brouwer problem in higher dimensions, as CLS is believed to stay strictly below PPAD, is Theorem 8.3.

7.3 Proof systems as optimization methods

Most classical TFNP classes now have characterizations by propositional proof systems in the black-box (query) model, the main exceptions being PPP and UEOPL: PLS corresponds to Resolution (of polylogarithmic width), PPA to Nullstellensatz over $\mathbb{F}_2$, PPAD to unary Nullstellensatz, PPADS to unary Sherali–Adams, and EOPL = PLS ∩ PPAD to Reversible

Resolution with Terminals, whose power is that of Resolution and unary Nullstellensatz together [33, 72–74]. Since gradient descent is complete for CLS = PLS ∩ PPAD in the white-box model [54], Reversible Resolution with Terminals is the natural proof-system counterpart of gradient descent, with the caveat that the two results live in different models. In the other direction, optimization algorithms *are* proof systems:

- *Cutting planes.* Chvátal's closure of a polytope under rounded linear combinations [38] is a proof system for integer infeasibility [43]. Pudlák proved exponential lower bounds for it by interpolation with monotone real circuits [149].
- *Branch and cut.* The Stabbing Planes system formalizes branch-and-cut [10]. Dadush and Tiwari showed that Tseitin formulas have quasi-polynomial cutting-plane refutations and bounded the coefficients needed in branching proofs [44], and Fleming et al. proved the first exponential lower bounds for Stabbing Planes with bounded coefficients [61]. The proof system that certifies the dominance reasoning of pseudo-Boolean optimization solvers simulates the quantified propositional proof system G_1 [115].
- *Sum of squares.* The Positivstellensatz proof system of Grigoriev and Vorobjov [77] is the proof system behind the Lasserre and Parrilo hierarchies for polynomial optimization [119, 141]; its degree lower bounds [76, 153] yield integrality gaps for the sum-of-squares hierarchy, and its bit complexity is subtle [8, 139, 150].

The last item closes a circle with Section 5. Artin's solution of Hilbert's 17th problem is non-constructive; Kreisel's analysis of it, which gave primitive recursive degree bounds, is one of the founding examples of his program of "unwinding" proofs, from which proof mining descends, and Daykin gave another constructive proof. Lombardi, Perrucci and Roy pushed this line to an elementary recursive bound (a tower of five exponentials) for the Positivstellensatz [129], explicitly by turning proofs into certificates. Effective degree bounds for certificates of positivity are the logical side of a question that is central for sum-of-squares hierarchies: how high a level is needed to certify a given bound.

Finally, weak theories prove the correctness of classical optimization algorithms. Le and Cook formalized in the polynomial-time theory VPV Hall's and Berge's theorems, weak duality for weighted bipartite matching, the min-max characterization of maximum-weight matchings, and a version of the Hungarian algorithm [120]; Fernández and Soltys proved Kőnig's min-max theorem, and with it the finite theorems of Menger, Hall and Dilworth, in a weak theory of linear algebra [57]. We found no formalization of general linear programming duality in bounded arithmetic; Theorem 6.1 suggests that a theory able to reason about determinants, such as those of Soltys and Cook [160] or of Tzameret and Cook [168], is the natural candidate.

8 Synthesis and open problems

Table 3 collects the calibrations discussed in this paper in one place.

8.1 Five principles, revisited

P1. Convexity and tameness substitute for compactness. Compactness is what makes general existence theorems cost WKL_0: the extreme value theorem, Brouwer's fixed point theorem, the Hahn–Banach theorem. Convexity makes existence cheap again (Theorem 3.2), for a reason that is visible in the proof: a convex function cannot hide a minimizer behind a local maximum, so a

Principle	Reverse mathematics	Weihrauch	Constructive (BISH)
Existence			
Maximum of a continuous f on $[0,1]$	WKL_0	$\equiv_{\mathrm{sW}}$ WKL	positive infimum $\Leftrightarrow$ FAN
Minimum of a convex f on $[0,1]$ / on $[0,1]^n$ with a modulus	RCA_0 (3.2)	$\equiv_{\mathrm{sW}}$ XC_n (3.3)	quasi-convex: no FAN needed
Polynomial minimum, fixed format	RCA_0 (3.5)	non-unif. computable	
Unique minimizer: modulus of uniqueness	WKL_0	computable	existence $\Leftrightarrow$ FAN; convex: free
Brouwer fixed point, $n \geq 2$	WKL_0	$\equiv_{\mathrm{sW}}$ WKL	
Minimum of an lsc f on $[0,1]$	ACA_0		
Projection onto a closed subspace; norm of a functional	ACA_0		
Ekeland: free, continuous f, compact space	WKL_0		
Ekeland: two of the three weakenings, or compact space alone	ACA_0		
Ekeland: free; localized; localized for continuous f	Π^1_1-CA_0		
Caristi for Borel maps	TLPP_0		
Convergence			
Rates for $f(x_k) - f^*$ and $\nabla f(x_k)$	RCA_0	computable, uniform	constructive
Gradient descent, PPA, KM converge (on $\mathbb{R}$)	ACA_0 (4.1)	limit $\equiv_{\mathrm{sW}}$ lim	
Projection of an anchor onto the solution set (Halpern, Tikhonov)	implies ACA_0 (4.3)	$\equiv_{\mathrm{sW}}$ lim	
Modulus of regularity exists / regularity	ACA_0 / WKL_0	$\equiv_{\mathrm{W}}$ $\widehat{\mathsf{LPO}}$ $\equiv_{\mathrm{W}}$ lim	
Mean ergodic theorem	ACA_0		
Gradient descent in $\mathbb{R}^d$: at most $2C_dD/\varepsilon$ ε-fluctuations (4.7)	(quantitative)	limits on computable instances d.c.e. (4.9)	
Duality			
LP duality, minimax (real data, all sizes)	RCA_0 (6.1)	linear inequalities: $\equiv_{\mathrm{W}}$ $\mathsf{AoUC}^*_{[0,1]}$	Farkas $\Leftrightarrow$ LPO
Nash equilibria, bimatrix / n players	$\leq \mathsf{WKL}_0$	$\equiv_{\mathrm{W}}$ $\mathsf{AoUC}^*_{[0,1]}$ / open	existence not constructive
Hahn–Banach; separation of open convex sets	WKL_0	HB: $\equiv_{\mathrm{W}}$ WKL	HB $\Leftrightarrow$ LLPO; a separation form $\Leftrightarrow$ MP
Strict separation from a separably closed set	ACA_0		
Kőnig duality (countable bipartite graphs)	ATR_0	arithm. $\equiv \mathsf{ATR}_2$	
Menger (countable graphs)	$\mathsf{ATR}_0 \leq \cdot \leq \mathsf{TLPP}_0$		
Feasible world			
Local optima: PLS, the NP search problems provably total in T^1_2. Approximate KKT points of smooth functions on the square (given by circuits): CLS = PPAD $\cap$ PLS. Nash, Brouwer, local min-max: PPAD. Exact KKT points on $[0,1]$: $\mathsf{KKT}_1 \equiv_{\mathrm{sW}}$ IVT (3.4).			

Table 3. An atlas of the calibrations discussed in this paper. Numbers refer to results of this paper; sources for the other entries are given in Sections 3, 4, 6 and 7. Empty cells are not claims of anything; several are open. "Non-unif. computable" means that computable instances have computable solutions.

one-dimensional search never has to backtrack. In dimension one this is a complete answer; in higher dimensions we need a modulus of uniform continuity, and whether convexity provides it in RCA_0 is open (Theorem 8.1). Tameness makes it cheap for a different reason: fixed-format semialgebraic statements follow from finitely many algebraic axioms (Theorem 3.5). The same phenomena appear in Bishop's constructive mathematics, where convexity replaces the fan theorem [14–16], in Weihrauch complexity, where convex choice sits strictly below closed choice [92, 121], and in combinatorial optimization, where the greedy algorithm makes matroid bases free [83].

P2. Selection is the price of convergence. Theorems 4.1 and 4.3 show that what costs ACA_0 is not existence but commitment. The limit of a convergent method is a realizer of convex minimization, a function of the data and the starting point; finding *some* minimizer is non-uniformly computable, but the realizers computed by gradient descent in dimension one and by anchored methods in every dimension select the projection of the starting point or anchor onto the solution set, and this selection is lim-hard (Theorems 4.1 and 4.3; Neumann [137] proved this for projections and for a large class of fixed-point iterations). This explains why the strength of convergence theorems does not depend on whether the algorithm is a descent method, a fixed-point iteration or an anchored scheme: each of them makes a choice of this kind, at least on one-dimensional instances. The same pattern appears in existence theorems that ask for an extremal solution: Peano's theorem is WKL_0, but Osgood's maximal solution is ACA_0 [156].

P3. Uniqueness and regularity are computational resources. A unique minimizer is computable, and so is a rate (Theorem 4.4). A modulus of uniqueness or regularity is a uniform rate. This is Kohlenbach's observation for best approximation [93], and it is the logical content of the many "error bound implies convergence rate" theorems of optimization: the error bound is a modulus of regularity, and the theorem is its quantitative unwinding. (Theorem 4.4(b) applies the modulus once; the familiar geometric rates come from applying a linear modulus repeatedly along the iteration.) Conversely, a uniform rate of convergence yields a modulus of regularity [114], and the existence of such moduli is equivalent to ACA_0 [101]; together these show that uniform rates for the iterates exist exactly under a regularity assumption, which is not available in general. O-minimality explains why regularity is so often available, though not why it should be computable from the data.

P4. Bounds on values are cheap, convergence of iterates is not. Every uniform rate in Theorem 2.2 follows from finitely many instances of (1): it has a proof that uses no compactness and no induction beyond Π^0_1, and the metatheorems guarantee that such a proof yields bounds depending only on L, D and k. Convergence of iterates is a Π^0_3 statement whose proofs use compactness; metastability is what is left after Kreisel's no-counterexample interpretation, and fluctuation bounds are what is left when, in addition, the geometry cooperates (Theorem 4.7). The principle is a heuristic: what makes the values uniform is the existence of a simple proof of a bound, not the logical form of the convergence statement as such.

P5. Dimension is the price of compactness. The rates of Theorem 2.2 hold in every Hilbert space; the fluctuation and metastability bounds of Theorem 4.7 depend on d and cannot be made dimension-free. The dividing line is the use of compactness (Bolzano–Weierstrass, or the

length of self-contracted curves) in the convergence proof, which proof mining detects as a dependence on a modulus of total boundedness. The Halpern iteration converges strongly in Hilbert space by a proof that uses no norm compactness (only weak sequential compactness or projections, available uniformly in every Hilbert space), and accordingly it has dimension-free metastability rates [99, 107].

8.2 Open problems

Problem 8.1. Does RCA_0 prove that for every n, every convex function on $[0,1]^n$ with a modulus of uniform continuity attains its minimum? The proof of Theorem 3.2(b) is an external induction on n. Does RCA_0 prove that every continuous convex function on $[0,1]^2$ has a modulus of uniform continuity? Similarly, does RCA_0 prove, with n and d as variables, that polynomials of degree d attain their minima on $[0,1]^n$, or that finite games of every size have Nash equilibria (Theorem 3.5)? Does the method of Theorem 6.1, applied to Lemke–Howson pivoting, give bimatrix Nash equilibria of every size in RCA_0?

Problem 8.2. Is every limit-computable single-valued realizer of CMin_1 Weihrauch equivalent to lim? Theorems 4.1 and 4.3 show this for the projection-type selections computed by gradient descent and anchored methods, while the multivalued problem CMin_1 itself is only IVT.

Problem 8.3. Determine the Weihrauch degree of KKT_n. By Theorem 3.4, $\mathsf{KKT}_1 \equiv_{\mathrm{sW}} \mathsf{IVT}$ and $\mathsf{KKT}_n \leq_{\mathrm{W}} \mathsf{BFT}_n \equiv_{\mathrm{sW}} \mathsf{WKL}$ for $n \geq 2$. Does every computable C^1 function on $[0,1]^2$ have a computable KKT point? A positive answer would make KKT_n a natural continuous counterpart of $\mathsf{CLS} = \mathsf{PPAD} \cap \mathsf{PLS}$, sitting strictly below the Brouwer problem. We found no example of a computable C^1 function on the square without a computable KKT point. A counterexample must be degenerate at its interior KKT points, since non-degenerate zeros of computable vector fields are computable [75]; co-c.e. planar continua without computable points [91] are a possible starting point.

Problem 8.4. Let $M_d(R)$ be the largest number of ε-fluctuations of gradient descent on convex 1-smooth functions on $\mathbb{R}^d$ with $\operatorname{dist}(x_0, \operatorname{argmin} f) \leq R\varepsilon$. Theorem 4.7 and Theorem 4.8 give $M_d(2\sqrt{d}) \geq d$, $M_d(R) \leq 2C_d R$ and $\sup_d M_d(R_0) = \infty$ for some R_0. Let c_d^{GD} be the best constant c such that every gradient-descent orbit (step $1/L$) of a convex L-smooth function on $\mathbb{R}^d$ has length at most $c\,D$. Determine the growth of c_d^{GD}, which lies between $\sqrt{d}/2$ (Theorem 4.8) and $2^{O(d^2)}$ [78], and the threshold $R^* = \inf\{R : \sup_d M_d(R) = \infty\}$.

Problem 8.5. Which reals are coordinates of limits of gradient descent on computable convex L-smooth functions on ℓ^2 when the iterates converge in norm? Are all limit-computable reals possible? Over RCA_0, is weak convergence of gradient descent in separable Hilbert spaces equivalent to ACA_0? Under the rate-free coding of reals of [64], does RCA_0 prove convergence of gradient descent in $\mathbb{R}^d$, as the finite-length argument behind Theorem 4.9(b) suggests (Theorem 4.2(c) discusses dimension one)?

Problem 8.6. Calibrate the duality theorems of convex optimization in separable Banach spaces: Fenchel–Rockafellar duality, Lagrange duality under Slater's condition, and the existence of subgradients. The separation theorems suggest WKL_0 when an interior point is available and ACA_0 otherwise. In the finite world, is linear programming duality provable in a bounded-arithmetic theory of linear algebra?

Problem 8.7. The generic proof-mined bounds for Fejér monotone sequences in $\mathbb{R}^d$ are of order $(\sqrt{d}\,D/\varepsilon)^d$ [113]; Theorem 4.7 gives $O_d(D/\varepsilon)$. Is there a logical metatheorem that explains the improvement, for instance one that treats self-contractedness as a monotonicity principle with linear fluctuation bounds?

Problem 8.8. Determine the exact strength of Menger's theorem for countable graphs (between ATR_0 and TLPP_0 [154, 167]), of max-flow/min-cut for countable networks [2] and of matroid intersection for countable matroids [89].

Problem 8.9. Calibrate the almost-sure convergence theorems of stochastic approximation (Robbins–Siegmund, stochastic gradient descent) in reverse mathematics and Weihrauch complexity, in the spirit of the quantitative results of Neri, Pischke and Powell [133, 134].

References

[1] Ron Aharoni, Menachem Magidor, and Richard A. Shore. On the strength of König's duality theorem for infinite bipartite graphs. *Journal of Combinatorial Theory, Series B*, 54(2):257–290, 1992. doi: 10.1016/0095-8956(92)90057-5.

[2] Ron Aharoni, Eli Berger, Agelos Georgakopoulos, Amitai Perlstein, and Philipp Sprüssel. The Max-Flow Min-Cut theorem for countable networks. *Journal of Combinatorial Theory, Series B*, 101 (1):1–17, 2011. doi: 10.1016/j.jctb.2010.08.002.

[3] Klaus Ambos-Spies, Klaus Weihrauch, and Xizhong Zheng. Weakly computable real numbers. *Journal of Complexity*, 16(4):676–690, 2000. doi: 10.1006/jcom.2000.0561.

[4] Hedy Attouch, Jérôme Bolte, and Benar Fux Svaiter. Convergence of descent methods for semi-algebraic and tame problems: proximal algorithms, forward–backward splitting, and regularized Gauss–Seidel methods. *Mathematical Programming*, 137(1–2):91–129, 2013.

[5] Jeremy Avigad and Jason Rute. Oscillation and the mean ergodic theorem for uniformly convex Banach spaces. *Ergodic Theory and Dynamical Systems*, 35(4):1009–1027, 2015. doi: 10.1017/etds.2013.90.

[6] Jeremy Avigad and Ksenija Simic. Fundamental notions of analysis in subsystems of second-order arithmetic. *Annals of Pure and Applied Logic*, 139(1–3):138–184, 2006. doi: 10.1016/j.apal.2005.03.004.

[7] Jean-Bernard Baillon and Ronald E. Bruck. The rate of asymptotic regularity is $O(1/\sqrt{n})$. In *Theory and Applications of Nonlinear Operators of Accretive and Monotone Type*, volume 178 of *Lecture Notes in Pure and Applied Mathematics*, pages 51–81. Dekker, New York, 1996.

[8] Boaz Barak and David Steurer. Sum-of-squares proofs and the quest toward optimal algorithms. In *Proceedings of the International Congress of Mathematicians (ICM 2014), Seoul*, 2014.

[9] Heinz H. Bauschke, Eva Matoušková, and Simeon Reich. Projection and proximal point methods: convergence results and counterexamples. *Nonlinear Analysis*, 56:715–738, 2004. doi: 10.1016/j.na.2003.10.010. URL `https://www.math.utoronto.ca/almut/MAT1000/P1P2-2004.pdf`.

[10] Paul Beame, Noah Fleming, Russell Impagliazzo, Antonina Kolokolova, Denis Pankratov, Toniann Pitassi, and Robert Robere. Stabbing planes. In *9th Innovations in Theoretical Computer Science Conference (ITCS 2018)*, volume 94 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 10:1–10:20, 2018. doi: 10.4230/LIPIcs.ITCS.2018.10.

[11] Arnold Beckmann and Samuel R. Buss. Polynomial local search in the polynomial hierarchy and witnessing in fragments of bounded arithmetic. *Journal of Mathematical Logic*, 9(1):103–138, 2009. doi: 10.1142/S0219061309000847.

[12] Josef Berger and Hajime Ishihara. Brouwer's fan theorem and unique existence in constructive analysis. *Mathematical Logic Quarterly*, 51(4):360–364, 2005. doi: 10.1002/malq.200410038.

[13] Josef Berger and Gregor Svindland. A separating hyperplane theorem, the fundamental theorem of asset pricing, and Markov's principle. *Annals of Pure and Applied Logic*, 167(11):1161–1170, 2016. doi: 10.1016/j.apal.2016.05.003.

[14] Josef Berger and Gregor Svindland. Convexity and constructive infima. *Archive for Mathematical Logic*, 55(7–8):873–881, 2016. doi: 10.1007/s00153-016-0502-y.

[15] Josef Berger and Gregor Svindland. Brouwer's fan theorem and convexity. *Journal of Symbolic Logic*, 83(4):1363–1375, 2018. doi: 10.1017/jsl.2018.49.

[16] Josef Berger and Gregor Svindland. Convexity and unique minimum points. *Archive for Mathematical Logic*, 58(1–2):27–34, 2019. doi: 10.1007/s00153-018-0619-2.

[17] Josef Berger and Gregor Svindland. On Farkas' lemma and related propositions in BISH. *Annals of Pure and Applied Logic*, 173(2):103059, 2022. doi: 10.1016/j.apal.2021.103059.

[18] Josef Berger, Douglas Bridges, and Peter Schuster. The fan theorem and unique existence of maxima. *Journal of Symbolic Logic*, 71(2):713–720, 2006. doi: 10.2178/jsl/1146620167.

[19] Robert G. Bland. New finite pivoting rules for the simplex method. *Mathematics of Operations Research*, 2(2):103–107, 1977.

[20] Axel Böhm and Aris Daniilidis. Ubiquitous algorithms in convex optimization generate self-contracted sequences. *Journal of Convex Analysis*, 29:119–128, 2022.

[21] Jérôme Bolte, Aris Daniilidis, and Adrian Lewis. The Łojasiewicz inequality for nonsmooth subanalytic functions with applications to subgradient dynamical systems. *SIAM Journal on Optimization*, 17(4):1205–1223, 2007.

[22] Vasco Brattka. Computability and analysis, a historical approach. In Arnold Beckmann, Laurent Bienvenu, and Nataša Jonoska, editors, *Pursuit of the Universal (CiE 2016)*, Lecture Notes in Computer Science, pages 45–57. Springer, 2016. doi: 10.1007/978-3-319-40189-8_5.

[23] Vasco Brattka and Guido Gherardi. Effective choice and boundedness principles in computable analysis. *The Bulletin of Symbolic Logic*, 17(1):73–117, 2011. doi: 10.2178/bsl/1294186663.

[24] Vasco Brattka and Guido Gherardi. Weihrauch degrees, omniscience principles and weak computability. *The Journal of Symbolic Logic*, 76(1):143–176, 2011. doi: 10.2178/jsl/1294170993.

[25] Vasco Brattka, Peter Hertling, and Klaus Weihrauch. A tutorial on computable analysis. In S. Barry Cooper, Benedikt Löwe, and Andrea Sorbi, editors, *New Computational Paradigms: Changing Conceptions of What is Computable*, pages 425–491. Springer, 2008.

[26] Vasco Brattka, Matthew de Brecht, and Arno Pauly. Closed choice and a uniform low basis theorem. *Annals of Pure and Applied Logic*, 163(8):986–1008, 2012. doi: 10.1016/j.apal.2011.12.020.

[27] Vasco Brattka, Guido Gherardi, and Alberto Marcone. The Bolzano–Weierstrass theorem is the jump of weak Kőnig's lemma. *Annals of Pure and Applied Logic*, 163(6):623–655, 2012. doi: 10.1016/j.apal.2011.10.006.

[28] Vasco Brattka, Stéphane Le Roux, Joseph S. Miller, and Arno Pauly. Connected choice and the Brouwer fixed point theorem. *Journal of Mathematical Logic*, 19(1):1950004, 2019. doi: 10.1142/S0219061319500041.

[29] Vasco Brattka, Guido Gherardi, and Arno Pauly. Weihrauch complexity in computable analysis. In *Handbook of Computability and Complexity in Analysis*, Theory and Applications of Computability, pages 367–417. Springer, 2021. doi: 10.1007/978-3-030-59234-9_11.

[30] Mario Bravo and Roberto Cominetti. Sharp convergence rates for averaged nonexpansive maps. *Israel Journal of Mathematics*, 227(1):163–188, 2018. doi: 10.1007/s11856-018-1723-z.

[31] Douglas K. Brown and Stephen G. Simpson. Which set existence axioms are needed to prove the separable Hahn-Banach theorem? *Annals of Pure and Applied Logic*, 31:123–144, 1986. doi: 10.1016/0168-0072(86)90066-7.

[32] Samuel R. Buss and Jan Krajíček. An application of Boolean complexity to separation problems in bounded arithmetic. *Proceedings of the London Mathematical Society*, s3-69(1):1–21, 1994. doi: 10.1112/plms/s3-69.1.1.

[33] Samuel R. Buss, Leszek Aleksander Kołodziejczyk, and Neil Thapen. Fragments of approximate counting. *The Journal of Symbolic Logic*, 79(2):496–525, 2014. doi: 10.1017/jsl.2013.37.

[34] Douglas Cenzer and Jeffrey B. Remmel. Proof-theoretic strength of the stable marriage theorem and other problems. In Stephen G. Simpson, editor, *Reverse Mathematics 2001*, volume 21 of *Lecture Notes in Logic*, pages 67–103. Association for Symbolic Logic / Cambridge University Press, 2005. doi: 10.1017/9781316755846.005.

[35] Xi Chen, Xiaotie Deng, and Shang-Hua Teng. Settling the complexity of computing two-player Nash equilibria. *Journal of the ACM*, 56(3):1–57, 2009. doi: 10.1145/1516512.1516516.

[36] Horaţiu Cheval and Laurenţiu Leuştean. Linear rates of asymptotic regularity for Halpern-type iterations. *Mathematics of Computation*, 2024. doi: 10.1090/mcom/3991.

[37] Horaţiu Cheval, Ulrich Kohlenbach, and Laurenţiu Leuştean. On modified Halpern and Tikhonov-Mann iterations. *Journal of Optimization Theory and Applications*, 197(1):233–251, 2023. doi: 10.1007/s10957-023-02192-6.

[38] Vašek Chvátal. Edmonds polytopes and a hierarchy of combinatorial problems. *Discrete Mathematics*, 4(4):305–337, 1973. doi: 10.1016/0012-365X(73)90167-2.

[39] Peter G. Clote and Jeffry L. Hirst. Reverse mathematics of some topics from algorithmic graph theory. *Fundamenta Mathematicae*, 157(1):1–13, 1998. doi: 10.4064/fm-157-1-1-13.

[40] Roberto Cominetti, José A. Soto, and José Vaisman. On the rate of convergence of Krasnosel'skiĭ-Mann iterations and their connection with sums of Bernoullis. *Israel Journal of Mathematics*, 199(2): 757–772, 2014. doi: 10.1007/s11856-013-0045-4.

[41] Juan Pablo Contreras and Roberto Cominetti. Optimal error bounds for non-expansive fixed-point iterations in normed spaces. *Mathematical Programming*, 199(1-2):343–374, 2023. doi: 10.1007/s10107-022-01830-7.

[42] Stephen Cook and Phuong Nguyen. *Logical Foundations of Proof Complexity*. Cambridge University Press, 2010. ISBN 9780521517294. doi: 10.1017/CBO9780511676277.

[43] William J. Cook, Collette R. Coullard, and György Turán. On the complexity of cutting-plane proofs. *Discrete Applied Mathematics*, 18(1):25–38, 1987. doi: 10.1016/0166-218X(87)90039-4.

[44] Daniel Dadush and Samarth Tiwari. On the complexity of branching proofs. In *35th Computational Complexity Conference (CCC 2020)*, volume 169 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 34:1–34:35, 2020. doi: 10.4230/LIPIcs.CCC.2020.34.

[45] Aris Daniilidis, Olivier Ley, and Stéphane Sabourau. Asymptotic behaviour of self-contracted planar curves and gradient orbits of convex functions. *Journal de Mathématiques Pures et Appliquées*, 94(2):183–199, 2010. doi: 10.1016/j.matpur.2010.03.007.

[46] Aris Daniilidis, Guy David, Estibalitz Durand-Cartagena, and Antoine Lemenant. Rectifiability of self-contracted curves in the Euclidean space and applications. *The Journal of Geometric Analysis*, 25(2):1211–1239, 2015. doi: 10.1007/s12220-013-9464-z.

[47] Constantinos Daskalakis and Christos Papadimitriou. Continuous local search. In *Proceedings of the Twenty-Second Annual ACM-SIAM Symposium on Discrete Algorithms (SODA 2011)*, pages 790–804, 2011. doi: 10.1137/1.9781611973082.62.

[48] Constantinos Daskalakis, Paul W. Goldberg, and Christos H. Papadimitriou. The complexity of computing a Nash equilibrium. *SIAM Journal on Computing*, 39(1):195–259, 2009. doi: 10.1137/070699652.

[49] Constantinos Daskalakis, Stratis Skoulakis, and Manolis Zampetakis. The complexity of constrained min-max optimization. In *Proceedings of the 53rd Annual ACM SIGACT Symposium on Theory of Computing (STOC 2021)*, pages 1466–1478, 2021. doi: 10.1145/3406325.3451125.

[50] Bruno Dinis and Pedro Pinto. Quantitative results on the multi-parameters proximal point algorithm. *Journal of Convex Analysis*, 28(3):729–750, 2021. doi: 10.68381/jca28041.

[51] Yoel Drori and Marc Teboulle. Performance of first-order methods for smooth convex minimization: a novel approach. *Mathematical Programming*, 145(1-2):451–482, 2014. doi: 10.1007/s10107-013-0653-0.

[52] Damir D. Dzhafarov and Carl Mummert. *Reverse Mathematics: Problems, Reductions, and Proofs*. Theory and Applications of Computability. Springer, 2022.

[53] Ivar Ekeland. On the variational principle. *Journal of Mathematical Analysis and Applications*, 47(2):324–353, 1974.

[54] John Fearnley, Paul W. Goldberg, Alexandros Hollender, and Rahul Savani. The complexity of gradient descent: CLS = PPAD ∩ PLS. In *Proceedings of the 53rd Annual ACM SIGACT Symposium on Theory of Computing (STOC 2021)*, pages 46–59, 2021. doi: 10.1145/3406325.3451052.

[55] John Fearnley, Paul W. Goldberg, Alexandros Hollender, and Rahul Savani. The complexity of computing KKT solutions of quadratic programs. In *Proceedings of the 56th Annual ACM Symposium on Theory of Computing (STOC 2024)*, pages 892–903, 2024. doi: 10.1145/3618260.3649647.

[56] António M. Fernandes and Fernando Ferreira. Groundwork for weak analysis. *The Journal of Symbolic Logic*, 67(2):557–578, 2002. doi: 10.2178/jsl/1190150098.

[57] Ariel Germán Fernández and Michael Soltys. Feasible combinatorial matrix theory. In Krishnendu Chatterjee and Jiří Sgall, editors, *Mathematical Foundations of Computer Science 2013 (MFCS)*, Lecture Notes in Computer Science, pages 777–788. Springer, 2013. doi: 10.1007/978-3-642-40313-2_68.

[58] David Fernández-Duque, Paul Shafer, and Keita Yokoyama. Ekeland's variational principle in weak and strong systems of arithmetic. *Selecta Mathematica (N.S.)*, 26(5):Art. 68, 2020. doi: 10.1007/s00029-020-00597-z.

[59] David Fernández-Duque, Paul Shafer, Henry Towsner, and Keita Yokoyama. Metric fixed point theory and partial impredicativity. *Philosophical Transactions of the Royal Society A*, 381(2248):20220012, 2023. doi: 10.1098/rsta.2022.0012.

[60] Fernando Ferreira. A feasible theory for analysis. *The Journal of Symbolic Logic*, 59(3):1001–1011, 1994. doi: 10.2307/2275924.

[61] Noah Fleming, Mika Göös, Russell Impagliazzo, Toniann Pitassi, Robert Robere, Li-Yang Tan, and Avi Wigderson. On the power and limitations of branch and cut. In *36th Computational Complexity Conference (CCC 2021)*, volume 200 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 6:1–6:30, 2021. doi: 10.4230/LIPIcs.CCC.2021.6.

[62] Stephen Flood, Matthew Jura, Oscar Levin, and Tyler Markkanen. The computational strength of matchings in countable graphs. *Annals of Pure and Applied Logic*, 173(8):103133, 2022. doi: 10.1016/j.apal.2022.103133.

[63] Anton Freund and Nicholas Pischke. Effective rates for continuous-time quasi-Fejér monotone dynamical systems, 2026.

[64] Anton Freund, Nicholas Pischke, and Patrick Uftring. Avoiding logical strength in real analysis, 2026.

[65] Harvey Friedman. Some systems of second order arithmetic and their use. In *Proceedings of the International Congress of Mathematicians (Vancouver, 1974)*, volume 1, pages 235–242. Canadian Mathematical Congress, 1975.

[66] Harvey Friedman. [FOM] constructive analysis. Foundations of Mathematics (FOM) mailing list, 13 May 2009, 2009. URL https://web.archive.org/web/2020/https://cs.nyu.edu/pipermail/fom/2009-May/013613.html.

[67] Harvey M. Friedman, Stephen G. Simpson, and Rick L. Smith. Countable algebra and set existence axioms. *Annals of Pure and Applied Logic*, 25(2):141–181, 1983. doi: 10.1016/0168-0072(83)90012-X.

[68] Makoto Fujiwara, Kojiro Higuchi, and Takayuki Kihara. On the strength of marriage theorems and uniformity. *Mathematical Logic Quarterly*, 60(3):136–153, 2014. doi: 10.1002/malq.201300021.

[69] Philipp Gerhardy and Ulrich Kohlenbach. General logical metatheorems for functional analysis. *Transactions of the American Mathematical Society*, 360(5):2615–2660, 2008. doi: 10.1090/S0002-9947-07-04429-7.

[70] Guido Gherardi and Alberto Marcone. How incomputable is the separable Hahn-Banach theorem? *Notre Dame Journal of Formal Logic*, 50(4):393–425, 2009. doi: 10.1215/00294527-2009-018.

[71] Jun Le Goh. Some computability-theoretic reductions between principles around ATR_0, 2019.

[72] Mika Göös, Pritish Kamath, Robert Robere, and Dmitry Sokolov. Adventures in monotone complexity and TFNP. In *10th Innovations in Theoretical Computer Science Conference (ITCS 2019)*, volume 124 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 38:1–38:19, 2019. doi: 10.4230/LIPIcs.ITCS.2019.38.

[73] Mika Göös, Alexandros Hollender, Siddhartha Jain, Gilbert Maystre, William Pires, Robert Robere, and Ran Tao. Further collapses in TFNP. In *37th Computational Complexity Conference (CCC 2022)*, volume 234 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 33:1–33:15, 2022. doi: 10.4230/LIPIcs.CCC.2022.33.

[74] Mika Göös, Alexandros Hollender, Siddhartha Jain, Gilbert Maystre, William Pires, Robert Robere, and Ran Tao. Separations in proof complexity and TFNP. In *2022 IEEE 63rd Annual Symposium on Foundations of Computer Science (FOCS)*, pages 1150–1161, 2022. doi: 10.1109/FOCS54457.2022.00111.

[75] Daniel S. Graça and Ning Zhong. The set of hyperbolic equilibria and of invertible zeros on the unit ball is computable. *Theoretical Computer Science*, 895:48–54, 2021. doi: 10.1016/j.tcs.2021.09.028.

[76] Dima Grigoriev. Linear lower bound on degrees of Positivstellensatz calculus proofs for the parity. *Theoretical Computer Science*, 259(1–2):613–622, 2001. doi: 10.1016/S0304-3975(00)00157-2.

[77] Dima Grigoriev and Nicolai Vorobjov. Complexity of Null- and Positivstellensatz proofs. *Annals of Pure and Applied Logic*, 113(1–3):153–160, 2001. doi: 10.1016/S0168-0072(01)00055-0.

[78] Chirag Gupta, Sivaraman Balakrishnan, and Aaditya Ramdas. Path length bounds for gradient descent and flow. *Journal of Machine Learning Research*, 22(68):1–63, 2021. URL https://jmlr.org/papers/v22/19-979.html.

[79] Christopher S. Hardin and Daniel J. Velleman. The mean value theorem in second order arithmetic. *Journal of Symbolic Logic*, 66(3):1353–1358, 2001. doi: 10.2307/2695111.

[80] Denis R. Hirschfeldt. *Slicing the Truth: On the Computable and Reverse Mathematics of Combinatorial Principles*, volume 28 of *Lecture Notes Series, Institute for Mathematical Sciences, National University of Singapore*. World Scientific, 2014.

[81] Jeffry L. Hirst. *Combinatorics in Subsystems of Second Order Arithmetic*. PhD thesis, The Pennsylvania State University, 1987.

[82] Jeffry L. Hirst. Marriage theorems and reverse mathematics. In *Logic and Computation (Pittsburgh,*

PA, 1987), volume 106 of *Contemporary Mathematics*, pages 181–196. American Mathematical Society, Providence, RI, 1990. doi: 10.1090/conm/106/1057822.

[83] Jeffry L. Hirst and Carl Mummert. Reverse mathematics of matroids. In Adam Day, Michael Fellows, Noam Greenberg, Bakhadyr Khoussainov, Alexander Melnikov, and Frances Rosamond, editors, *Computability and Complexity: Essays Dedicated to Rodney G. Downey on the Occasion of His 60th Birthday*, volume 10010 of *Lecture Notes in Computer Science*, pages 143–159. Springer, 2017. doi: 10.1007/978-3-319-50062-1_12.

[84] Alan J. Hoffman. On approximate solutions of systems of linear inequalities. *Journal of Research of the National Bureau of Standards*, 49(4):263–265, 1952.

[85] A. James Humphreys and Stephen G. Simpson. Separation and weak König's lemma. *Journal of Symbolic Logic*, 64(1):268–278, 1999. doi: 10.2307/2586763.

[86] Hein S. Hundal. An alternating projection that does not converge in norm. *Nonlinear Analysis: Theory, Methods & Applications*, 57(1):35–61, 2004. doi: 10.1016/j.na.2003.11.004.

[87] Hajime Ishihara. An omniscience principle, the König lemma and the Hahn-Banach theorem. *Zeitschrift für Mathematische Logik und Grundlagen der Mathematik (Mathematical Logic Quarterly)*, 36 (3):237–240, 1990. doi: 10.1002/malq.19900360307.

[88] David S. Johnson, Christos H. Papadimitriou, and Mihalis Yannakakis. How easy is local search? *Journal of Computer and System Sciences*, 37(1):79–100, 1988. doi: 10.1016/0022-0000(88)90046-3.

[89] Attila Joó. Proof of Nash-Williams' intersection conjecture for countable matroids. *Advances in Mathematics*, 380:107608, 2021. doi: 10.1016/j.aim.2021.107608.

[90] William Julian and Fred Richman. A uniformly continuous function on [0,1] that is everywhere different from its infimum. *Pacific Journal of Mathematics*, 111(2):333–340, 1984. doi: 10.2140/pjm.1984.111.333.

[91] Takayuki Kihara. Incomputability of simply connected planar continua. *Computability*, 1(2): 131–152, 2012. doi: 10.3233/COM-12012.

[92] Takayuki Kihara and Arno Pauly. Finite choice, convex choice and sorting. In T. V. Gopal and Junzo Watada, editors, *Theory and Applications of Models of Computation (TAMC 2019)*, volume 11436 of *Lecture Notes in Computer Science*, pages 378–393. Springer, 2019. doi: 10.1007/978-3-030-14812-6_23.

[93] Ulrich Kohlenbach. Effective moduli from ineffective uniqueness proofs. An unwinding of de La Vallée Poussin's proof for Chebycheff approximation. *Annals of Pure and Applied Logic*, 64(1):27–94, 1993. doi: 10.1016/0168-0072(93)90213-W.

[94] Ulrich Kohlenbach. New effective moduli of uniqueness and uniform a priori estimates for constants of strong unicity by logical analysis of known proofs in best approximation theory. *Numerical Functional Analysis and Optimization*, 14(5-6):581–606, 1993. doi: 10.1080/01630569308816541.

[95] Ulrich Kohlenbach. A quantitative version of a theorem due to Borwein-Reich-Shafrir. *Numerical Functional Analysis and Optimization*, 22(5-6):641–656, 2001. doi: 10.1081/NFA-100105311.

[96] Ulrich Kohlenbach. Uniform asymptotic regularity for Mann iterates. *Journal of Mathematical Analysis and Applications*, 279(2):531–544, 2003. doi: 10.1016/S0022-247X(03)00028-3.

[97] Ulrich Kohlenbach. Some logical metatheorems with applications in functional analysis. *Transactions of the American Mathematical Society*, 357(1):89–128, 2005. doi: 10.1090/S0002-9947-04-03515-9.

[98] Ulrich Kohlenbach. *Applied Proof Theory: Proof Interpretations and their Use in Mathematics*. Springer Monographs in Mathematics. Springer, Berlin, Heidelberg, 2008. ISBN 978-3-540-77532-4. doi: 10.1007/978-3-540-77533-1.

[99] Ulrich Kohlenbach. On quantitative versions of theorems due to F.E. Browder and R. Wittmann. *Advances in Mathematics*, 226(3):2764–2795, 2011. doi: 10.1016/j.aim.2010.10.002.

[100] Ulrich Kohlenbach. Proof-theoretic methods in nonlinear analysis. In *Proceedings of the International Congress of Mathematicians (ICM 2018), Rio de Janeiro, Vol. 2*, pages 61–82. World Scientific, 2019. doi: 10.1142/9789813272880_0045.

[101] Ulrich Kohlenbach. On the reverse mathematics and Weihrauch complexity of moduli of regularity and uniqueness. *Computability*, 8(3–4):377–387, 2019. doi: 10.3233/COM-180097.

[102] Ulrich Kohlenbach. Quantitative analysis of a Halpern-type proximal point algorithm for accretive operators in Banach spaces. *Journal of Nonlinear and Convex Analysis*, 21:2125–2138, 2020.

[103] Ulrich Kohlenbach. Quantitative results on the proximal point algorithm in uniformly convex Banach spaces. *Journal of Convex Analysis*, 28(1):11–18, 2021. doi: 10.68381/jca28002.

[104] Ulrich Kohlenbach. On the proximal point algorithm and its Halpern-type variant for generalized monotone operators in Hilbert space. *Optimization Letters*, 16(2):611–621, 2022. doi: 10.1007/s11590-021-01738-9.

[105] Ulrich Kohlenbach. Kreisel's "shift of emphasis" and contemporary proof mining. *Studies in Logic*, 16(3):1–35, 2023.

[106] Ulrich Kohlenbach. On the computational content of moduli of regularity and their logical strength. In *Proceedings LTT 2026*, volume 441 of *Electronic Proceedings in Theoretical Computer Science*, pages 185–196, 2026. doi: 10.4204/EPTCS.441.11.

[107] Ulrich Kohlenbach and Laurenţiu Leuştean. Effective metastability of Halpern iterates in CAT(0) spaces. *Advances in Mathematics*, 231(5):2526–2556, 2012. doi: 10.1016/j.aim.2012.06.028.

[108] Ulrich Kohlenbach and Paulo Oliva. Proof mining in L_1-approximation. *Annals of Pure and Applied Logic*, 121(1):1–38, 2003. doi: 10.1016/S0168-0072(02)00081-7.

[109] Ulrich Kohlenbach and Pedro Pinto. Quantitative translations for viscosity approximation methods in hyperbolic spaces. *Journal of Mathematical Analysis and Applications*, 507(2):125823, 2022. doi: 10.1016/j.jmaa.2021.125823.

[110] Ulrich Kohlenbach and Pedro Pinto. Fejér monotone sequences revisited. *Journal of Convex Analysis*, 33(1-2):75–90, 2026. doi: 10.68381/jca33005.

[111] Ulrich Kohlenbach and Nicholas Pischke. Proof theory and non-smooth analysis. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 381(2248):20220015, 2023. doi: 10.1098/rsta.2022.0015.

[112] Ulrich Kohlenbach and Pavol Safarik. Fluctuations, effective learnability and metastability in analysis. *Annals of Pure and Applied Logic*, 165(1):266–304, 2014. doi: 10.1016/j.apal.2013.07.014. URL https://www2.mathematik.tu-darmstadt.de/~kohlenbach/mind_ps_revised.pdf.

[113] Ulrich Kohlenbach, Laurenţiu Leuştean, and Adriana Nicolae. Quantitative results on Fejér monotone sequences. *Communications in Contemporary Mathematics*, 20(2):1750015, 2018. doi: 10.1142/S0219199717500158.

[114] Ulrich Kohlenbach, Genaro López-Acedo, and Adriana Nicolae. Moduli of regularity and rates of convergence for Fejér monotone sequences. *Israel Journal of Mathematics*, 232(1):261–297, 2019. doi: 10.1007/s11856-019-1870-x.

[115] Leszek Aleksander Kołodziejczyk and Neil Thapen. The strength of the dominance rule, 2024.

[116] Leszek Aleksander Kołodziejczyk, Phuong Nguyen, and Neil Thapen. The provably total NP search problems of weak second order bounded arithmetic. *Annals of Pure and Applied Logic*, 162 (6):419–446, 2011. doi: 10.1016/j.apal.2010.12.002.

[117] Andreas Kontogiannis, Ioannis Panageas, and Vasilis Pollatos. The computational complexity of avoiding strict saddle points in constrained optimization, 2026.

[118] Krzysztof Kurdyka. On gradients of functions definable in o-minimal structures. *Annales de l'Institut Fourier*, 48(3):769–783, 1998.

[119] Jean B. Lasserre. Global optimization with polynomials and the problem of moments. *SIAM Journal on Optimization*, 11(3):796–817, 2001. doi: 10.1137/S1052623400366802.

[120] Dai Tri Man Le and Stephen A. Cook. Formalizing randomized matching algorithms. *Logical Methods in Computer Science*, 8(3), 2012. doi: 10.2168/LMCS-8(3:5)2012.

[121] Stéphane Le Roux and Arno Pauly. Finite choice, convex choice and finding roots. *Logical Methods in Computer Science*, 11(4):Paper 6, 2015. doi: 10.2168/LMCS-11(4:6)2015.

[122] Stéphane Le Roux and Martin Ziegler. Singular coverings and non-uniform notions of closed set computability. *Mathematical Logic Quarterly*, 54(5):545–560, 2008. doi: 10.1002/malq.200610058.

[123] Yunseok Lee, Holger Boche, and Gitta Kutyniok. Computability of optimizers, 2023.

[124] Carlton E. Lemke and Joseph T. Howson, Jr. Equilibrium points of bimatrix games. *Journal of the Society for Industrial and Applied Mathematics*, 12(2):413–423, 1964.

[125] Laurenţiu Leuştean. Rates of asymptotic regularity for Halpern iterations of nonexpansive mappings. *Journal of Universal Computer Science*, 13(11):1680–1691, 2007. doi: 10.3217/jucs-013-11-1680.

[126] Laurenţiu Leuştean and Pedro Pinto. Quantitative results on a Halpern-type proximal point algorithm. *Computational Optimization and Applications*, 79(1):101–125, 2021. doi: 10.1007/s10589-021-00263-w.

[127] Laurenţiu Leuştean, Adriana Nicolae, and Andrei Sipoş. An abstract proximal point algorithm. *Journal of Global Optimization*, 72(3):553–577, 2018. doi: 10.1007/s10898-018-0655-9.

[128] Felix Lieder. On the convergence rate of the Halpern-iteration. *Optimization Letters*, 15(2):405–418, 2021. doi: 10.1007/s11590-020-01617-9.

[129] Henri Lombardi, Daniel Perrucci, and Marie-Françoise Roy. An elementary recursive bound for effective Positivstellensatz and Hilbert's 17th problem. *Memoirs of the American Mathematical Society*, 263(1277), 2020. doi: 10.1090/memo/1277.

[130] Arkadi S. Nemirovsky and David B. Yudin. *Problem Complexity and Method Efficiency in Optimization*. Wiley-Interscience Series in Discrete Mathematics. Wiley, Chichester, New York, 1983. ISBN 0471103454.

[131] Morenikeji Neri and Nicholas Pischke. Proof mining and probability theory. *Forum of Mathematics, Sigma*, 13:e187, 2025. doi: 10.1017/fms.2025.10138.

[132] Morenikeji Neri and Thomas Powell. On quantitative convergence for stochastic processes: Crossings, fluctuations and martingales. *Transactions of the American Mathematical Society, Series B*, 12:974–1019, 2025. doi: 10.1090/btran/229.

[133] Morenikeji Neri and Thomas Powell. A quantitative Robbins-Siegmund theorem. *The Annals of Applied Probability*, 36(1):636–651, 2026. doi: 10.1214/25-AAP2242.

[134] Morenikeji Neri, Nicholas Pischke, and Thomas Powell. An abstract effective convergence theorem for stochastic processes, with applications to stochastic approximation, 2025.

[135] Morenikeji Neri, Nicholas Pischke, and Thomas Powell. Generalized fluctuation bounds for stochastic algorithms in the presence of compactness, 2026.

[136] Yurii Nesterov. *Introductory Lectures on Convex Optimization: A Basic Course*. Kluwer Academic Publishers, 2004.

[137] Eike Neumann. Computational problems in metric fixed point theory and their Weihrauch degrees. *Logical Methods in Computer Science*, 11(4):Paper 20, 2015. doi: 10.2168/LMCS-11(4:20)2015.

[138] Phuong The Nguyen. *Bounded Reverse Mathematics*. PhD thesis, University of Toronto, Graduate Department of Computer Science, 2008. URL https://eccc.weizmann.ac.il/static/books/Bounded_Reverse_Mathematics.

[139] Ryan O'Donnell. SOS is not obviously automatizable, even approximately. In *8th Innovations in Theoretical Computer Science Conference (ITCS 2017)*, volume 67 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 59:1–59:10, 2017. doi: 10.4230/LIPIcs.ITCS.2017.59.

[140] Christos H. Papadimitriou. On the complexity of the parity argument and other inefficient proofs of existence. *Journal of Computer and System Sciences*, 48(3):498–532, 1994. doi: 10.1016/S0022-0000(05)80063-7.

[141] Pablo A. Parrilo. Semidefinite programming relaxations for semialgebraic problems. *Mathematical Programming*, 96(2):293–320, 2003. doi: 10.1007/s10107-003-0387-5.

[142] Arno Pauly. How incomputable is finding Nash equilibria? *Journal of Universal Computer Science*, 16(18):2686–2710, 2010. doi: 10.3217/jucs-016-18-2686.

[143] NingNing Peng, Weiguang Peng, and Takeshi Yamazaki. Mixed Nash equilibria for continuous games and reverse mathematics. *Quaestiones Mathematicae*, 46(4):621–632, 2023. doi: 10.2989/16073606.2022.2035448.

[144] Weiguang Peng and Takeshi Yamazaki. Two kinds of fixed point theorems and reverse mathematics. *Mathematical Logic Quarterly*, 63(5):454–461, 2017. doi: 10.1002/malq.201600096.

[145] Nicholas Pischke. Logical metatheorems for accretive and (generalized) monotone set-valued operators. *Journal of Mathematical Logic*, 24(2):2350008, 2024. doi: 10.1142/S0219061323500083.

[146] Nicholas Pischke. Generalized Fejér monotone sequences and their finitary content. *Optimization*, 74(14):3771–3838, 2025. doi: 10.1080/02331934.2024.2390114.

[147] Nicholas Pischke and Ulrich Kohlenbach. Quantitative analysis of a subgradient-type method for equilibrium problems. *Numerical Algorithms*, 90(1):197–219, 2022. doi: 10.1007/s11075-021-01184-9.

[148] Nicholas Pischke and Thomas Powell. Asymptotic regularity of a generalised stochastic Halpern scheme. *Journal of Optimization Theory and Applications*, 210(1):3, 2026. doi: 10.1007/s10957-026-03032-z.

[149] Pavel Pudlák. Lower bounds for resolution and cutting plane proofs and monotone computations. *The Journal of Symbolic Logic*, 62(3):981–998, 1997. doi: 10.2307/2275583.

[150] Prasad Raghavendra and Benjamin Weitz. On the bit complexity of sum-of-squares proofs. In *44th International Colloquium on Automata, Languages, and Programming (ICALP 2017)*, volume 80 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 80:1–80:13, 2017. doi: 10.4230/LIPIcs.ICALP.2017.80.

[151] R. Tyrrell Rockafellar. *Convex Analysis*. Princeton University Press, 1970.

[152] Shoham Sabach and Shimrit Shtern. A first order method for solving convex bilevel optimization problems. *SIAM Journal on Optimization*, 27(2):640–660, 2017. doi: 10.1137/16M105592X.

[153] Grant Schoenebeck. Linear level Lasserre lower bounds for certain k-CSPs. In *2008 49th Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 593–602, 2008. doi: 10.1109/FOCS.2008.74.

[154] Paul Shafer. Menger's theorem in Π^1_1-CA_0. *Archive for Mathematical Logic*, 51(3–4):407–423, 2012. doi: 10.1007/s00153-012-0269-8.

[155] Naoki Shioji and Kazuyuki Tanaka. Fixed point theory in weak second-order arithmetic. *Annals of Pure and Applied Logic*, 47(2):167–188, 1990. doi: 10.1016/0168-0072(90)90068-D.

[156] Stephen G. Simpson. Which set existence axioms are needed to prove the Cauchy/Peano

theorem for ordinary differential equations? *Journal of Symbolic Logic*, 49(3):783–802, 1984. doi: 10.2307/2274131.

[157] Stephen G. Simpson. On the strength of König's duality theorem for countable bipartite graphs. *Journal of Symbolic Logic*, 59(1):113–123, 1994. doi: 10.2307/2275254.

[158] Stephen G. Simpson. *Subsystems of Second Order Arithmetic*. Perspectives in Logic. Cambridge University Press / Association for Symbolic Logic, 2 edition, 2009. ISBN 978-0-521-88439-6. doi: 10.1017/CBO9780511581007.

[159] Andrei Sipoş. Quantitative inconsistent feasibility for averaged mappings. *Optimization Letters*, 16 (6):1915–1925, 2022. doi: 10.1007/s11590-021-01812-2.

[160] Michael Soltys and Stephen Cook. The proof complexity of linear algebra. *Annals of Pure and Applied Logic*, 130(1–3):277–323, 2004. doi: 10.1016/j.apal.2003.10.018.

[161] Ernst Specker. Nicht konstruktiv beweisbare Sätze der Analysis. *The Journal of Symbolic Logic*, 14 (3):145–158, 1949. doi: 10.2307/2267043.

[162] Ernst Specker. Der Satz vom Maximum in der rekursiven Analysis. In Arend Heyting, editor, *Constructivity in Mathematics (Proc. Colloq. Amsterdam 1957)*, Studies in Logic and the Foundations of Mathematics, pages 254–265. North-Holland, Amsterdam, 1959.

[163] Eugene Stepanov and Yana Teplitskaya. Self-contracted curves have finite length. *Journal of the London Mathematical Society*, 96(2):455–481, 2017. doi: 10.1112/jlms.12068.

[164] Terence Tao. Norm convergence of multiple ergodic averages for commuting transformations. *Ergodic Theory and Dynamical Systems*, 28(2):657–688, 2008. doi: 10.1017/S0143385708000011.

[165] Alfred Tarski. *A Decision Method for Elementary Algebra and Geometry*. University of California Press, 2 edition, 1951.

[166] Adrien B. Taylor, Julien M. Hendrickx, and François Glineur. Smooth strongly convex interpolation and exact worst-case performance of first-order methods. *Mathematical Programming*, 161(1-2): 307–345, 2017. doi: 10.1007/s10107-016-1009-3.

[167] Henry Towsner. Partial impredicativity in reverse mathematics. *Journal of Symbolic Logic*, 78(2): 459–488, 2013. doi: 10.2178/jsl.7802070.

[168] Iddo Tzameret and Stephen A. Cook. Uniform, integral, and feasible proofs for the determinant identities. *Journal of the ACM*, 68(2):1–80, 2021. doi: 10.1145/3431922.

[169] Klaus Weihrauch. *Computable Analysis: An Introduction*. Texts in Theoretical Computer Science. An EATCS Series. Springer, Berlin, 2000. doi: 10.1007/978-3-642-56999-9.

[170] Rainer Wittmann. Approximation of fixed points of nonexpansive mappings. *Archiv der Mathematik*, 58(5):486–491, 1992.

[171] Xizhong Zheng, Robert Rettinger, and George Barmpalias. h-monotonically computable real numbers. *Mathematical Logic Quarterly*, 51(2):157–170, 2005. doi: 10.1002/malq.200410016.

Appendix

The Logical Anatomy of Optimization

A Formalization notes

We record the points at which the proofs of Theorems 4.1, 4.3 and 6.1 use facts about RCA_0 that are standard but easy to overlook.

1. *Iterates exist.* By (1), g is L-Lipschitz (add the two upper inequalities at (x, y) and (y, x) and use co-coercivity). So an approximation of x_{k+1} to precision 2^{-n} needs only an approximation of x_k to precision 2^{-n-c}, for a constant c depending on L, and one value of g at a rational point. The iterates are therefore defined by primitive recursion on a Δ^0_1 table of values of g, and no Π^0_2 induction is needed to show that they exist. "The gradient-descent sequence converges" then refers to this sequence.
2. *The function f_y.* For x in a bounded interval, the series defining f_y and f_y' converge uniformly with an explicit modulus (the tail after s terms is at most 2^{-s} times a bound on $(y_s - x)^2$), so the pair (f_y, f_y') exists as a pair of codes of continuous functions in RCA_0, and (1) with $L = 2$ holds because it holds for every partial sum and is preserved under limits.
3. *Comparison arguments.* The statements "$x_k \le z$ for all k" in the proofs of Theorems 4.1 and 4.3 are Π^0_1 and are proved by Π^0_1 induction, which is available in RCA_0.
4. *The final search.* To decide $m \in \mathrm{ran}(h)$ from x_∞, search for n with $\hat{x} - y_n < 2^{-m-2}$, where $\hat{x}$ is the approximation of x_∞ to precision 2^{-m-3} supplied by its code. The search terminates because $y_n \to x_\infty$, and any n found satisfies $x_\infty - y_n < 2^{-m-1}$.
5. *Bounded comprehension.* In Theorem 6.1, the sign pattern of finitely many reals is obtained by bounded Σ^0_1 comprehension, which is provable in RCA_0 [158, II.3]; everything else is first-order reasoning about finitely many explicitly given reals.

B The computations behind Fig. 2

The instance in Fig. 2 is gradient descent with step $1/2$ from $x_0 = 0$ on the function f_y of (3), for the data $y_s = \sum_{r \le s} b_r$, where $b_r = 10^{-4}(r+1)^{-3/2}$ plus a jump 2^{-m-1} at stages $r = 0, 30, 300, 3000, 30000$ with $m = 1, 3, 4, 5, 6$. The background term keeps (y_s) strictly increasing; the late jumps imitate the enumeration of small elements of a computably enumerable set. All quantities are computed in 80-digit decimal arithmetic. The series for f_y and f_y' are truncated after 210 terms, which changes them by less than 10^{-63}; the limit $x_\infty = \sup_s y_s$ is computed from the data, including the tail of the background term. Over the 10^5 steps shown, $f(x_k) - f^*$ stays far below the bound $LD^2/(2k)$ of Theorem 2.2(c), and the gradient decays to about 10^{-10}, while $x_\infty - x_k$ stays at 0.117. The scripts are available with the source of this paper.

C Linear programming duality in RCA_0: the details

This appendix expands the proof of Theorem 6.1. Throughout we work in RCA_0; the sizes m and n are arbitrary natural numbers, possibly non-standard, and the data A, b, c are finite sequences of reals, coded as usual by a single set. Only Σ^0_1 induction (equivalently Π^0_1 induction) and bounded Σ^0_1 comprehension are used.

Step 1: formal polynomials and their values. Integer polynomials in finitely many variables are finite objects, coded by natural numbers. Evaluation at a finite sequence of reals is defined by primitive recursion on the polynomial, and RCA_0 proves, by Π^0_1 induction on the size of the polynomials involved, that evaluation is a ring homomorphism: the value of $p + q$ is the sum of the values, and similarly for products. (Each of these is a statement $\forall p, q\, \Pi^0_1$, because equality of reals is Π^0_1.) Consequently every identity between integer polynomials that holds formally, i.e. that is a true equality of finite objects after expansion into monomials, transfers to the corresponding identity between reals. The determinant of an $r \times r$ matrix is the value of the Leibniz polynomial, a sum of $r!$ terms; this is a finite sum and exists for every r. The identities we need are formal identities with the size r as a parameter: expansion along a row or column, multilinearity, Cramer's rule in the form $\det(B)\, x_B = \left(\det B^{(i \leftarrow b)}\right)_i$, and the description of the entries of a simplex tableau as quotients of minors of the bordered matrix. Their usual proofs are combinatorial arguments about finite objects: they are $\Delta_0(\exp)$ identities proved by bijections and sign-reversing involutions on the symmetric group S_r, so $\mathsf{I}\Delta_0 + \exp$, which is contained in $\mathsf{I}\Sigma_1$, the first-order part of RCA_0, suffices (compare the much finer analysis of linear algebra in weak theories by Soltys and Cook [160]). This is also why the textbook proofs of duality by induction on $m + n$ cannot be used directly: for non-standard sizes such an induction would range over formulas that quantify over real vectors, i.e. over sets, and RCA_0 has no induction for them. The sign oracle below replaces that induction by a single application of comprehension.

Step 2: the sign oracle. Put the primal problem in equality form with slack variables, $[A\ I]\,(x, s) = b$, $(x, s) \geq 0$. Let $\mathcal{E}$ be the finite list of integer polynomials in the entries of (A, b, c) consisting of: the entries b_i; all minors of the bordered matrices of Phase I (constraint matrix with rows multiplied by ± 1 and artificial columns, right-hand side, Phase I cost row) and of Phase II (constraint matrix, right-hand side, cost row), for every choice of the row signs; and all expressions $\mu_1\mu_2 - \mu_3\mu_4$ with $\mu_1, \ldots, \mu_4$ such minors. The length of $\mathcal{E}$ is bounded by an explicit (exponential) function of $m + n$. The values v_e ($e \in \mathcal{E}$) form a finite sequence of reals, and "$v_e > 0$" and "$v_e < 0$" are Σ^0_1 uniformly in e. By bounded Σ^0_1 comprehension [158, II.3] the finite sets $\{e : v_e > 0\}$ and $\{e : v_e < 0\}$ exist, so the sign function $\sigma\colon \mathcal{E} \to \{-1, 0, 1\}$ with $\sigma(e) = \operatorname{sign}(v_e)$ exists as a finite object. This is the only point at which the proof is not uniform in the data.

Step 3: the symbolic run. The two-phase simplex method with Bland's smallest-index rule [19] is a primitive recursive procedure that takes σ (and m, n) as input and manipulates bases, i.e. finite sets of column indices. At a basis B every quantity the method inspects is the sign of a tableau entry, of a reduced cost, or of a difference of two ratios in a ratio test; by Step 1 each of these signs equals $\sigma(e)$ or $\sigma(e)\sigma(e')$ for explicitly computable $e, e' \in \mathcal{E}$ (for ratios, multiply out the positive denominators). Ties are decided by $\sigma(e) = 0$, and Bland's rule breaks them by the

smallest index. So the run, a finite sequence of bases $B_0, B_1, \dots$ together with the outcome, is defined by primitive recursion from σ. By Δ_0^0 induction (in the parameter σ) on t, the sign data of B_t certify feasibility: $\sigma(\Delta_{B_t}) \neq 0$ and $\sigma(\mu_i)\sigma(\Delta_{B_t}) \geq 0$ for the minors μ_i that give the basic variables by Cramer's rule; the induction step is the usual verification that a pivot chosen by the ratio test preserves feasibility, which is an identity between quotients of minors together with the signs supplied by σ. Because σ agrees with the true signs, each B_t is a basis that is feasible for the real data.

Step 4: termination. Bland's proof that his rule never revisits a basis considers a hypothetical cycle, the largest index that enters and leaves in it, and two tableaux of the cycle; the identity that the cost row of one tableau is orthogonal to the direction defined by the entering column of the other, together with the signs forced by the rule, gives a contradiction. All of this concerns finitely many reals given by quotients of minors and their signs, so it is available in RCA_0 by Steps 1 and 3. Hence no basis repeats, and by the finite pigeonhole principle (provable in $\mathsf{I}\Sigma_1$) each phase stops after at most $\binom{n+2m}{m}$ pivots.

Step 5: Phase I. Multiply the rows with $\sigma(b_i) = -1$ by -1 and add one artificial variable per row. The artificial basis is feasible, and Phase I minimizes the sum of the artificial variables. Phase I cannot be unbounded, since an improving ray would make the sum of the non-negative artificial variables negative. Its optimal value is 0, because the primal is feasible and the value is a lower bound for the sum of artificials at any feasible point. Artificial variables that remain basic at level 0 are pivoted out by degenerate pivots on non-zero entries of their rows in original columns; such entries exist because $[A\ I]$ has rank m, and their non-zeroness is again read off σ. This yields a feasible basis of the sign-flipped system, which is also a basis of the original system $[A\ I]$ with the same basic solution; Phase II is run on the original, unflipped system, so that the dual vector below is expressed in the original coordinates.

Step 6: Phase II and the conclusion. Phase II stops either at an optimal basis, where no column can improve the objective, or with an improving column j whose entries are all non-positive. In the second case the column yields a ray $r \geq 0$ with $Ar \leq 0$ along which the objective increases without bound. This contradicts weak duality: for a dual feasible $y \geq 0$ and every primal feasible x, $c^\top x \leq y^\top Ax \leq y^\top b$, an inequality between finitely many reals. In the first case let B be the final basis. Cramer's rule gives the basic solution x^* and $y^{*\top} = c_B^\top B^{-1}$ as explicit quotients of minors; the optimality condition on the reduced costs says exactly that $A^\top y^* \geq c$ (structural columns) and $y^* \geq 0$ (slack columns), and $c^\top x^* = c_B^\top B^{-1} b = b^\top y^*$ is a formal identity.

Step 7: minimax. Let $P \in \mathbb{R}^{m \times n}$, $m, n \geq 1$, be a payoff matrix (the row player maximizes $p^\top Pq$, the column player minimizes it). Choose a rational $t > 1 + \max_{ij}|P_{ij}|$ (from approximations of the entries) and put $Q = P + t$, which has positive entries. The linear programs $\max\{\mathbf{1}^\top x : Qx \leq \mathbf{1}, x \geq 0\}$ and $\min\{\mathbf{1}^\top y : Q^\top y \geq \mathbf{1}, y \geq 0\}$ are dual to each other and both feasible ($x = 0$, and y a large multiple of $\mathbf{1}$), so by Steps 2–6 they have optimal solutions with a common value θ, and $\theta > 0$ because $Q^\top y^* \geq \mathbf{1}$ forces $y^* \neq 0$. Then $q = x^*/\theta$ and $p = y^*/\theta$ are mixed strategies with $p^\top Q \geq \theta^{-1}\mathbf{1}^\top$ and $Qq \leq \theta^{-1}\mathbf{1}$, so the game Q has value θ^{-1} and the game P has value $\theta^{-1} - t$, attained by p and q. All steps are uniform in m and n. ∎

Remark. Fourier–Motzkin elimination gives an alternative route: one application of bounded Σ^0_1 comprehension to the pre-enumerated (doubly exponential) list of the polynomials that Fourier–Motzkin elimination produces yields Farkas' lemma in RCA_0, and duality follows by the usual homogenization. It avoids Phase I and anticycling at the cost of a larger list.