Similarity as a Step Rule: Parameter-Free Root Finding Derived from a Similarity Function

V. Sitnikov

Abstract

Bisection places the next iterate at the midpoint because it uses no information about the function's value. We replace that rule with a *similarity* rule: the step is proportional to how *dissimilar* the current outcome $f(x)$ is to the expectation 0, measured by a similarity function $S\colon \mathbb{R} \to [0,1]$ with $S(0) = 1$. We derive the contraction factor of the resulting iteration in closed form and confirm it numerically to four digits. Three consequences follow and are all verified. First, the two-point normalisation of S over the bracket reproduces the false-position formula identically, so a similarity rule calibrated by the bracket is not a new method. Second, any similarity function fixed in advance yields at best linear convergence at a rate that is a property of the problem rather than of the method; on eight standard problems, five to eight of them fail. Third, the rate vanishes exactly when the similarity's scale equals the local scale of the value axis; reading that scale off the data — a divided difference, not a fitted coefficient — turns the step into $|f|/|f'|$ and the convergence superlinear.

The derivation fixes the method completely, before any benchmark is consulted: a similarity whose approach to 1 is linear, the finest available data-derived scale, and a safeguard that projects the step onto the admissible interval rather than replacing it. That configuration carries a proven worst case of twice bisection plus one step and contains no tuning constant of any kind. On the 154 problems of Alefeld, Potra and Shi it needs 2806 evaluations of f against 2820 for TOMS 748 — a margin of 0.5%, which is a tie — and 2890 for Brent, whom it beats on 102 problems to 40 head to head. On the 45 problems of Chandrupatla it beats Brent 30 to 4 and TOMS 748 31 to 8, and loses to ITP. The result is parity with methods built on tuned constants, obtained without any. The ranking of the nine similarity functions across the whole grid matches the derived law without exception. Reading the same similarity in one more place — which three points the local model is built on, selected by the shape of their configuration rather than by recency — lowers the Alefeld–Potra–Shi count to 2527, below every method measured here including Chandrupatla's 2785, with the guarantee intact and still no tuning constant; reading it instead on the choice of far end gives 5511 over both sets. Where the similarity is read turns out to matter more than which similarity it is.

1 Introduction

A bracketing root finder must decide, at every step, where inside the bracket to evaluate next. Bisection answers "in the middle", which is the correct answer when nothing is known. Every improvement since has answered "where a local model of f says the root is": a secant in false position, an inverse quadratic in Brent [2], a cubic in TOMS 748 [1], an interpolate–truncate–project cycle in ITP [4]. Each of those answers carries constants: Illinois halves the stale ordinate, Brent accepts an interpolation subject to three empirical inequalities, ITP is governed by κ_1, κ_2, n_0.

This paper asks a different question. Suppose the only thing available is a *similarity function* — a map S from an outcome to $[0,1]$, where 1 means "identical to what was wanted" and 0 means "not similar at all", in the sense used for parameter-free control in [6]. Can a root finder be built out of S alone, and how does it compare with methods built out of polynomial models and tuned constants?

The answer is sharp in both directions, and the negative half is as informative as the positive half.

2 The similarity step rule

Definition 1 (Similarity step)**.** Let $[a, b]$ bracket a root, x be the current iterate and x_{far} the opposite end. Let $S\colon \mathbb{R} \to [0, 1]$ with $S(0) = 1$ be a similarity function and $\sigma > 0$ a scale. The next iterate is

$$x' \;=\; x \;+\; \big(1 - S(f(x)/\sigma)\big)\,(x_{\text{far}} - x). \tag{1}$$

The rule is exactly the informal statement "step further when the outcome is dissimilar to the expectation, and less far when it is similar". Bisection is the degenerate member $S \equiv \frac{1}{2}$: a similarity function that carries no information about f.

2.1 The two-point normalisation is false position

The most obvious way to make (1) scale-free is to normalise the similarity by the two ends of the bracket, $S(a) = |f(b)|/(|f(a)| + |f(b)|)$ and $S(b) = 1 - S(a)$, and to place the next iterate at the similarity-weighted average of the ends.

Proposition 1. $S(a)\,a + S(b)\,b \;=\; \dfrac{af(b) - bf(a)}{f(b) - f(a)}$.

That is the false-position formula. The identity is elementary but decisive: a similarity rule whose scale is taken from the bracket is not a new method, and it inherits false position's stalling failure. In our measurements plain false position solves 127 of the 154 Alefeld–Potra–Shi problems, against 153 for bisection.

2.2 The contraction law

Write $e_n = x_n - r$ for the error, $D = x_{\text{far}} - r$, and let c be the slope of $1 - S$ at the origin, $1 - S(z) = c\,z + O(z^2)$.

Proposition 2 (Contraction law)**.** *For a simple root and e_n small, rule* (1) *satisfies*

$$e_{n+1} \;=\; e_n \left(1 - \frac{c\,|f'(r)|\,D}{\sigma}\right) + O(e_n^2), \qquad \textit{rate} = \left|1 - \frac{c\,|f'(r)|\,D}{\sigma}\right|. \tag{2}$$

Derivation. $|f(x_n)| = |f'(r)|\,|e_n| + O(e_n^2)$, hence $1 - S(f(x_n)/\sigma) = c\,|f'(r)|\,|e_n|/\sigma + O(e_n^2)$, and $x_{\text{far}} - x_n = D - e_n$. Substituting into (1) and discarding second-order terms gives (2). □

Table 1 checks (2) against measurement on $f(x) = x^2 - 2$ over $[0, 2]$ with $\sigma = |f(0)|$, for which $|f'(r)| = 2\sqrt{2}$, $D = 2 - \sqrt{2}$ and $\sigma = 2$.

Corollary 1 (A chosen similarity is not enough)**.** *For fixed S and fixed σ the rate in* (2) *depends on $|f'(r)|$, D and σ, none of which is under the method's control. Convergence is at best linear and occurs only when the product lands in $(0, 2)$.*

Empirically, with σ taken as $|f|$ at the starting point, five to eight of eight standard problems fail outright for the six similarity functions tested.

Corollary 2 (Faster is not better)**.** *Increasing c moves the rate away from 1 in either direction. Similarities with unbounded slope — $1 - \sqrt{|z|}$ has $c \approx 3.2 \cdot 10^3$ at the resolution used, $1 - |z|^{1/3}$ has $c \approx 4.6 \cdot 10^4$ — overshoot as readily as flat ones stagnate, and fail on five and six of the eight problems respectively.*

Corollary 3 (The scale must come from the data)**.** *The rate is zero exactly when $c\,|f'(r)|\,D/\sigma = 1$, i.e. $\sigma = c\,|f'(r)|\,D$. Substituting this σ back into* (1) *with $c = 1$ gives $x' - x = |f(x)|/|f'|$, the Newton step. The similarity rule is therefore superlinear precisely when its scale is the* local scale of the value axis*, and that scale is not available a priori: it must be read off the evaluations.*

Table 1: The contraction law against measurement. c is the slope of $1 - S$ at the origin. Similarities that approach 1 quadratically have $c = 0$ and are predicted to stagnate.

$S(z)$	c	predicted rate	measured rate
$1/(1+\lvert z\rvert)$	1	0.1716	0.1716
$e^{-\lvert z\rvert}$	1	0.1716	0.1716
$1/(1+z^2)$	0	1.0000	0.9357
e^{-z^2}	0	1.0000	0.9444
$1/\sqrt{1+z^2}$	0	1.0000	0.9327
$\max(0, 1-\lvert z\rvert)$	1	0.1716	1.0000 (saturates)

3 The method

Corollaries 1–3 fix every choice in the method before any benchmark is consulted. We record them here as a prediction, and Section 5.3 tests it against the full grid.

Similarity: any with $c \neq 0$. The law makes the shape of S almost irrelevant and its slope at the origin decisive. We take the plainest member, the ratio form $S(z) = \max(0, 1 - |z|)$.

Scale: the finest data-derived one. We set $\sigma = |\hat{f}'| \cdot |x_{\text{far}} - x|$ with $\hat{f}'$ a divided difference of the most recent evaluations. A divided difference is a computation, not a fitted coefficient: nothing is chosen, regressed or tuned. The two-point difference has slope error $O(h)$ and gives the secant order $\varphi \approx 1.618$; the three-point difference — the derivative at x_n of the parabola through the last three evaluations, $\hat{f}' = f[x_{n-1}, x_n] + f[x_{n-2}, x_{n-1}, x_n](x_n - x_{n-1})$ — has error $O(h^2)$ and lifts it. Measured medians are 1.584 and 1.944, against 1.809 for Brent and 2.654 for TOMS 748.

3.1 The safeguard: project, do not replace

An interpolation-only rule has no worst-case bound. The obvious safeguards *replace* the computed step with a midpoint whenever some progress test fails, and all of them are ruinous here, for a reason that is specific to a fast method: a step that collapses the *error* need not halve the *bracket*, so a bracket-based test fires on exactly the steps that were best. With the three-point scale, replacing on that test drops the observed order from 1.94 to 1.00 and the Alefeld–Potra–Shi total from 2806 to 7260 — worse than bisection.

The remedy is to keep as much of the computed step as the budget allows. Let $n_{\max} = 2\lceil \log_2(w_0/\varepsilon_x) \rceil$ be twice the bisection step count for the initial width w_0 and x-tolerance ε_x, and let $\epsilon = \varepsilon_x/2$. At step j, with the bracket $[\ell, h]$, define the allowance and the projection

$$\alpha_j = 2^{\,n_{\max}-j}\epsilon, \qquad x \;=\; \text{clamp}\big(x_{\text{sim}},\; h - \alpha_j,\; \ell + \alpha_j\big). \tag{3}$$

The next bracket is at most $\max(x - \ell,\, h - x)$, so (3) gives $w_{j+1} \leq \alpha_j$ directly, and induction bounds the run by $n_{\max}$ steps in exact arithmetic. In floating point the bracket can land exactly on the boundary, and the termination test then asks for one more; the guarantee stated and verified is therefore $2\lceil \log_2(w_0/\varepsilon_x) \rceil + 3$ evaluations.

Three remarks on why this introduces nothing.

- **The doubling is not a tuning constant.** It is the factor already claimed in the bound. Zero slack — $n_{\max} = \lceil \log_2(w_0/\varepsilon_x) \rceil$, a worst case of exactly bisection — is degenerate and was measured as such: a worst case of exactly bisection demands a halving at *every* step, so the projection has no freedom, the radius on a unit bracket is 0.05 from the first step, and the guard behaves as bisection.

- **This is ITP's projection stage without ITP's truncation stage.** All three of ITP's constants live in the truncation, whose purpose is to repair regula falsi; the similarity step needs no repair, so the truncation and every constant with it are dropped.
- **Writing the clamp in the form** (3) **rather than as a radius $\alpha_j - w_j/2$ about the midpoint is not cosmetic.** That radius is a difference of two nearly equal numbers; the cancellation made the computed radius too large by a relative $8.9 \cdot 10^{-5}$, which broke the invariant on eleven Chandrupatla problems.

4 Experimental protocol

Problem sets. The 154 problems of Alefeld, Potra and Shi [1] and the 45 of Chandrupatla [3], taken programmatically from SciPy's `scipy.optimize._tstutils` so that no bracket or root is transcribed by hand.

Cost model. Evaluations of f. A run counts as solved when it reports convergence and lands within 10^{-6} of the published root; x-tolerance 10^{-12}; budget 200 evaluations. *Failures are charged the full budget*, so no method can buy a low total by giving up early, and totals are reported over all problems rather than over the subset a method happened to solve. Rankings are by failures first, then by total.

Fairness to the competitors. Three deliberate concessions. (i) ITP is natively written against a half-width ϵ and terminates on $b - a \leq 2\epsilon$; feeding the common tolerance in as ϵ would let it stop at twice its competitors' bracket width, so $\epsilon = \varepsilon_x/2$ is used. (ii) ITP is reported both at the authors' recommended constants and as a *per-problem oracle*, the best of a (κ_1, n_0) grid chosen separately for every problem — not an achievable method, included so that the state of the art cannot be said to have been handicapped. (iii) Our implementation of ITP is validated against the method's own theorem, that it never needs more than $n_{1/2} + n_0$ iterations, on all 154 problems for $n_0 \in \{0, 1, 2, 5\}$.

Reproducibility. All numbers come from a single Linux run. The test functions call `libm`, whose last digit differs between platforms; the same SciPy version scores 2819 for TOMS 748 on Windows and 2820 on Linux for the same set. Every number stated in this paper is checked against the run data by a script that is part of the test suite.

5 Results

The grid is nine similarity functions × four scales × six safeguards = 216 configurations, plus ten baselines. The configuration proposed in Section 3 — ratio similarity, three-point scale, projection safeguard — is one entry of that grid, fixed by the theory rather than chosen from the ranking.

On the problems both solve, the proposed configuration beats Brent 102/11/40 (wins/ties/losses), TOMS 748 86/16/51, ITP at its recommended constants 132/3/18, the per-problem ITP oracle 121/4/28, and bisection 152/1/0.

The set's nominal winner is a similarity configuration that solves one problem more than anything else, 41 of 45, at a worse total; ranking by failures first puts it on top, and we report it as such rather than by the total that would flatter our proposal. The proposed configuration is tenth by total and beats Brent 30/4/4 and TOMS 748 31/1/8, while losing to ITP 17/0/23 and to the ITP oracle 16/0/24.

Table 2: Alefeld–Potra–Shi, 154 problems, out of 226 configurations.

#	configuration	tuned	solved	total evals	$> 2\times$
1	**proposed: ratio, 3-point, projection**	**no**	153/154	**2806**	0
3	TOMS 748	yes	153/154	2820	0
4	Brent	yes	153/154	2890	0
11	ITP, per-problem oracle	yes	153/154	3260	0
18	Ridders	no	153/154	3313	0
31	ITP, recommended constants	yes	153/154	3957	0
61	Illinois	yes	153/154	7045	0
65	bisection	no	153/154	7530	0
206	false position	no	127/154	10560	–

Table 3: Chandrupatla, 45 problems, brackets up to 10^{12} wide.

#	configuration	tuned	solved	total evals
1	similarity, $1 - \sqrt{\lvert z \rvert}$, bracket scale, projection	no	**41/45**	4516
2	ITP, per-problem oracle	yes	40/45	2611
3	similarity, $e^{-\lvert z \rvert}$, bracket scale, both guards	no	40/45	2733
5	ITP, recommended constants	yes	40/45	2857
6	bisection	no	40/45	3090
10	**proposed: ratio, 3-point, projection**	**no**	40/45	**3220**
127	TOMS 748	yes	40/45	4000
189	Brent	yes	38/45	4101

5.1 What the guarantee costs

Nothing on one set, and less than nothing on the other. Against the identical rule with no safeguard at all:

	solved	total evals	runs over the $2\times$ bound
APS, no safeguard	153/154	2806	0
APS, projection	153/154	2806	0
Chandrupatla, no safeguard	35/45	4104	10
Chandrupatla, projection	**40/45**	**3220**	**0**

On the canonical set the projection never binds and the totals are identical to the last evaluation. On the wide-bracket set it solves five more problems for 21% fewer evaluations. This is the opposite of the usual trade, and it is the point of projecting rather than replacing: the two safeguards that replace cost 7260 and 7239 on Alefeld–Potra–Shi, against 2806.

5.2 Wall-clock time, and why it is not the cost model

Evaluations of f are the cost model because in any real problem f is the expensive part. Wall-clock time measures something narrower — the method's own arithmetic per step — and it is implementation- and language-bound. We report it because a reader will ask, and because what it shows is instructive.

On the 199 problems as they stand, every f being a formula of a few operations, the proposed rule is the *slowest* of the pure-Python methods compared, at 32.8 ms against 18.4 for Brent, 21.1 for ITP, 14.2 for Ridders and 22.9 for bisection — while using the fewest evaluations of all of

them, 4991 against 5932, 5675, 6421 and 9474. Bisection, needing nearly twice the evaluations, finishes in two thirds of our time.

That inversion is the point. Making f progressively more expensive by a controlled amount of arithmetic and timing the sweep directly — nothing subtracted, nothing modelled — the proposed rule overtakes ITP once one call to f costs about $9\,\mu$s and Brent at about $20\,\mu$s, of the order of a hundred arithmetic operations on this machine. Below that the ranking is decided by per-step overhead; above it, by evaluations.

Three caveats, all limitations of the measurement rather than of the methods. Timings vary by tens of percent between runs even on an idle machine. An attempt to separate our per-step cost into "algorithm" and "Python call overhead", by timing a flattened rewrite of the same rule, produced a variant that diverged from the original on 42 of the 199 problems and was discarded rather than quoted; the question stays open. And an earlier version of this measurement estimated a per-evaluation overhead by subtracting the measured cost of f, which returned *negative* overheads for four of six methods — an impossibility that condemned the estimator, and the reason nothing is subtracted here.

5.3 The law across the whole grid

Corollaries 1–3 predict an ordering of the nine similarity functions and of the four scales. Table 4 reports it. The prediction holds without exception for every data-derived scale: the three similarities with $c = 1$ occupy the top three places and the three with $c = 0$ the bottom three. For the chosen scale the two groups are indistinguishable (16102 against 15345), which is consistent — with a chosen scale nothing converges well and the ordering carries no information.

Table 4: Mean total evaluations over both sets, step guard, grouped by the slope c of $1 - S$ at the origin.

scale	$c = 1$ (linear approach)	$c = 0$ (quadratic approach)
3-point derived	**6809**	22294
2-point derived	6906	22167
bracket	10263	19137
start (chosen)	16102	15345

Ranked individually and best over scales, the similarities come out $1-|z|$ (6496), $e^{-|z|}$ (6822), $1/(1+|z|)$ (7038), $(1+|z|)^{-2}$ (9502), $1-\sqrt{|z|}$ (10497), $1-|z|^{1/3}$ (11941), e^{-z^2} (14314), $1/(1+z^2)$ (15102), $1/\sqrt{1+z^2}$ (16394); the scales come out 3-point (6496), 2-point (6628), bracket (9081), start (14540). Both orderings are those the law predicts, and both were predicted before the grid was run.

6 Where the evaluations go

Splitting each run at the geometric mean of the initial width and the attainable tolerance — the log-midpoint of the journey, at which bisection by construction has spent half its evaluations — separates the *approach* from the *endgame*. On Alefeld–Potra–Shi the two-point rule spends 2235 on the approach and 668 on the endgame against Brent's 2175 and 515: its deficit is almost entirely the endgame, i.e. the order, which is what motivated the three-point scale. On Chandrupatla it spends 1068 and 1161 against Brent's 1343 and 1358, winning both phases.

Two natural explanations of the Chandrupatla advantage were tested and *refuted*. Brent discards 53.3% of its interpolations to its safeguard on the set it wins and only 34.0% on the set it loses, so safeguard rate is not the mechanism; and on Chandrupatla Brent is the bolder of the two — 56.3% of its probes land in the outer tenths of the bracket against our 41.8% — so step length is not the mechanism either.

The mechanism is step *quality*. Reconstructing each method's bracket from the signs it observed, the median probe removes 0.500 of the bracket for every method on Alefeld–Potra–Shi, but on Chandrupatla it removes 0.275 for the similarity rule against 0.058 for Brent and 0.068 for TOMS 748 — and that comparison already flatters them, since 69 and 141 of their probes respectively fall outside the proven bracket and are excluded, against none of ours. High-order interpolation is a wager that f is locally polynomial. When the wager pays, order 1.84 or 2.65 wins the endgame; when it fails, the probes are wasted. The similarity rule makes the weakest available assumption — a single local slope — and degrades gracefully.

7 Where the similarity is read

Everything above reads the similarity in one place: the *length* of the step, $(1 - S(z))$ times the distance to the far end. But a bracketing iteration takes several decisions per step, and the similarity was not consulted in any of the others: which end counts as far (decided by distance), which points the local slope is estimated from (decided by recency), and whether the step is worth taking at all (decided after the fact, by the safeguard). Each of those is a place where a similarity can be read, and the placement turns out to matter more than the form of S.

Three readings were added, each parameter-free and each measured on the same protocol; every number below is re-checked against its sweep file by `verify_article.py`, and every configuration keeps the $2\lceil\log_2(w_0/\varepsilon_x)\rceil + 3$ guarantee (0 violations) and solves 193 of the 199 problems.

1. **Trust before the step.** The step trusts a local slope, so the question of trust is whether that slope has settled: if two consecutive estimates have opposite signs the local model is not established, and the evaluation is spent halving instead. A sign is compared, not a magnitude.

2. **Which points the model uses.** For a linear f the ratio of positions $\xi = (x-u)/(v-u)$ and the ratio of values $\varphi = (f(x)-f(u))/(f(v)-f(u))$ coincide, so $|\varphi - \xi|$ measures how unlike a linear configuration a triple is. Among the declared candidate triples the one minimising it is *selected* — an argmin over what is already in hand, with no threshold.

3. **Which end is far.** Read by dissimilarity instead of distance: the end whose $|f|$ is larger is the one less similar to the target.

configuration (all with the guarantee, 193/199)	APS-154	CHA-45	both
proposed method of §3 (credit)	3568	3020	6588
projection variant	2806	3220	6026
(1) trust before the step	2848	3074	5922
(1)+(2) points chosen by shape	**2527**	3108	5635
(1)+(3) far end by dissimilarity	2726	**2785**	**5511**
(1)+(2)+(3)	2884	2859	5743
Chandrupatla	2785	2423	5208
TOMS 748	2820	4000	6820
ITP (oracle over its constants)	3260	2611	5871

Table 5: Reading the similarity in one more place is worth more than changing S. Evaluations of f; failures charged the full budget.

Two results here are worth stating plainly. First, the canonical set improves from 2806 to **2527**, which is below every method measured, including Chandrupatla's 2785, and it is obtained with the worst-case guarantee intact and still without a tuning constant. Second, on the two sets together the 5511 of reading (3) passes ITP's oracle (5871) but not Chandrupatla (5208): the

remaining gap is entirely on the wide-bracket families, where Chandrupatla needs 2423 against our 2683.

The readings do not compose. Taken together, (2) and (3) give 2884 and 2859 — worse than either alone. Two selections that are each an improvement interfere, which is a reason to measure composition rather than assume it.

7.1 The similarity function belongs to the place, not to the method

Everything above varies *where* the similarity is read while keeping one function S for the whole method. That is itself an assumption, and it is measurable. Proposition 1 says the two-point normalisation is false position; what it does not say is *why* false position stalls. Write $|f| \sim Ke^m$ near the current point, $e = |x - r|$. Then the residual ratio against the far end needs no slope at all,

$$z \;=\; \frac{|f(x)|}{|f(x_{\text{far}})|} \;=\; \Big(\frac{e}{E}\Big)^m, \qquad \text{so} \qquad z^{1/m} \;=\; \frac{e}{E}, \tag{4}$$

and under $S = 1/(1+z)$ the step fraction is $z/(1+z)$, which is exactly inverse linear interpolation. False position therefore reads a dimensionless ratio *at order one* on an approach of order m: its three-century-old stall is a mismatch between the similarity function and the place it is read in. The exponent $1/m$ is not a constant — $m = 1/(1 - ff''/f'^2)$ comes from three evaluations already paid for, and is 1 whenever the power model does not apply.

Measured over the 252 cells of the grid, the correction behaves exactly as that reading predicts and not at all as a universal improvement. Paired inside each (function, safeguard) cell of the far reading it is cheaper in 29 of 60 cells — indistinguishable from a coin. But in the one pairing where the far reading *is* exact inverse interpolation, $S = 1/(1 + z)$, it is cheaper in all seven safeguards:

safeguard	APS m=1	APS measured m	CHA m=1	CHA measured m
both	4522	**3675**	3387	**3216**
bracket	4606	**3542**	3571	**3303**
credit	5190	**4686**	3013	**2884**
project	6932	**6364**	4794	**4559**
safe	5987	**5229**	4108	**3923**
step	6435	**5551**	4257	**4022**
none	10676	**10043**	8265	**7794**

Table 6: The measured order in the step place, on the far reading. Cheaper in 7 of 7 safeguards for the similarity that matches the place, and in 29 of 60 cells over all nine functions.

Two limits must be stated with it. The estimator is *biased low*: on a pure cube, with no noise at all, it returns 2.830 rather than 3, because a parabola through three points carries only the second difference. The bias is towards $m = 1$, that is towards the uncorrected reading, which is the harmless direction, but it is a bias. And the correction does not rescue the bracket reading as a method: the best cell it reaches on the wide-bracket set is 2884, against 2683 for reading (3) above and 2423 for Chandrupatla. In the *slope* reading the same exponent is harmful, 2806 → 2854, which is the same statement read from the other side — a function that does not belong to a place does not become right by being applied there.

7.2 Six problems of the two sets are scored unsolved for every method

Chasing the remaining gap to Chandrupatla led to a property of the benchmarks themselves. On the five problems of the wide-bracket family `fun7`, and on the single canonical problem `aps.13.00`, the declared root is 0 while f vanishes *identically* on a neighbourhood of it in

floating point: $f(0 \pm t) = 0$ for every t up to at least 3.2×10^{-2}, because the exponential factor underflows — at the returned points $|f(x)|$ is below the smallest normal double. Every method therefore stops after 8 to 28 evaluations at a point where $f(x) = 0$ *exactly* — a genuine root — and every method is then scored as having failed, because the harness asks for $|x - \text{root}| \leq 10^{-6}$ while the returned point sits up to 2.8×10^{-2} away.

The consequence is arithmetic. Each such problem charges the full budget of 200 evaluations to every method: about 1150 evaluations of every reported two-set total — ours, Chandrupatla's, TOMS 748's, ITP's, Brent's, bisection's — are this artefact, and the "153 of 154" and "40 of 45" that every method reports are statements about the scoring rather than about the methods. Re-scored with the criterion "a root is a root" ($|x - \text{root}| \leq 10^{-6}$ *or* $f(x) = 0$), the same runs solve 199 of 199 for every method except Brent's 197, and the totals fall as follows.

	APS-154		CHA-45	
method	as scored	re-scored	as scored	re-scored
ours, points by shape	**2527**	**2337**	3108	2159
ours, projection variant	2806	2634	3220	2357
Chandrupatla	2785	2593	**2423**	**1487**
TOMS 748	2820	2636	4000	3088
Brent	2889	2707	4099	3222
ITP	3957	3767	2857	1907
bisection	7530	7338	3090	2136

Table 7: The same runs under the harness criterion and under "a root is a root". The ranking is unchanged; the magnitudes are not, and neither is the size of the remaining gap on the wide-bracket set: 253 evaluations, or 17%, not 11%.

7.3 Where the wide-bracket limit actually is

The remaining gap looks diffuse — per family +57, +54, +46, +43, +35 and smaller — but it has one address. The `fun1` family is the same function with brackets of growing width, and our cost against Chandrupatla's across them is 7/8, 12/12, 19/15, 59/24, 63/44: we tie or win while the bracket is narrow and lose progressively as it widens. The endgame is not the cause — our observed order on those very problems is the higher one (3.63 against 2.02 on the 2×10^4-wide bracket). The cause is the rate of progress far from the root: 59 evaluations on a 2×10^4-wide bracket is what bisection itself needs ($\lceil \log_2(2 \times 10^{16}) \rceil \approx 54$), i.e. there we advance by one halving per evaluation while Chandrupatla advances by about 2.3.

It is not the guarantee that caps us. With the safeguard removed entirely the same problems cost 46, 57, 61 and 68 against bisection's 57, 77, 45 and 45 — better on the wide brackets, *worse than bisection* on `fun3.1` and `fun5.1` — and still far from Chandrupatla's 24 and 44.

Nor is it that the step fails to be relative to the bracket. With the ratio similarity $1 - S = \min(1, |f|/(|f'|D))$, so the step is exactly $\min(D, |f|/|f'|)$: a *fraction* of the distance to the far end, and the similarity is that fraction. Nor does the iterate creep along one side without moving the proven bracket: measured probe by probe, 91.5% of our evaluations on the 2×10^4-wide bracket do shrink the bracket, slightly *more* often than Chandrupatla's 87.5%.

What differs is how much each probe removes. Ours shrinks the bracket by 0.93 halvings per evaluation there, Chandrupatla's by 2.26 — a factor of about two against a factor of about five. On a wide bracket of a convex function the Newton step systematically undershoots, because a linear model of the value axis cannot see the curvature that separates the iterate from the root, so our probe lands well short and the new bracket is about half the old one; a model that carries the curvature — inverse quadratic interpolation — lands far closer and removes about four fifths.

So the limit is stated exactly, and it is the *order of the model the similarity is read against*, not the placement of the reading and not the guarantee. Three parameter-free ways round it

have now been measured and all three fail in the same way, buying the wide brackets at the price of the canonical set: a fixed fraction of the bracket (`cbrt`), a logarithmic change of variable, and a curvature-corrected scale (`derived3_m`, whose derivation holds but whose estimator of the local order fires on 27.9% of steps where the truth is a simple root). The fourth way is to model the curvature directly, which is inverse quadratic interpolation — a different method rather than another reading of this one.

On that set our shape reading (2159 re-scored) is *worse* than bisection (2136); only the dissimilarity reading (1740) is clearly ahead, which is the same statement seen from the other end: where the step cannot march, the reading that helps least is the one that only improves the local model. Every statement in these two subsections is reproduced by `diag_rescore_cha.py`, `diag_rescore_aps.py`, `diag_fun7_root.py` and `diag_cha_gap.py`, whose outputs are kept beside them.

8 Negative results

1. **Placement beats form, and the wrong placement is expensive.** The same "refuse before the step" idea, with refusal decided by *promised bracket progress* instead of by the shape of the model, costs 4449 on the canonical set against 2527 — a difference of 1922 evaluations for one choice of criterion. A secant step routinely collapses the error while barely moving the proven bracket, so bracket progress rejects steps that were excellent; this is the same mechanism that makes the bracket guard cost 7260.

2. **Making the scale independent of the bracket does not help either.** The natural objection to the derived scale is that it defeats the very principle of a similarity step: with $\sigma = |f'|D$ the normalised residual is $|f|/(|f'|D)$, which on the 2×10^4-wide bracket of $x^3 - 2x - 5$ at $x = 10^4$ equals 0.167, so $S = 0.833$ — "similar" — and the step is small although the root is some five thousand step-lengths away. The similarity is reading what fraction of the *bracket* the Newton step is, not how far the iterate is from the root. Selecting the scale between the two quantities already computed, $\sigma = \min(|f'|D, \max(|f(A)|, |f(B)|))$ — no new constant, a `min` over what is in hand — does give a full-bracket step far away and Newton's step near the root, and on eight hand-picked wide problems it improves 403 evaluations to 390. Over the whole grid it is worse on both sets: 3288 on the canonical set against 2726, and 3111 on the wide-bracket set against 2683. Its apparent extra solves (194 instead of 193) are two `fun7` problems where it lands inside the flat zero region close enough to satisfy the harness — the artefact of §7.2, not a capability.

3. **Reading the similarity at the measured order does not cure false position.** Since false position's stall is a similarity read at order one (§7.1), the obvious hope is that reading it at the measured order is the parameter-free version of the classical fix. It is not. Unguarded, the correction takes the bracket reading from 127/154 and 10676 to 128/154 and 10043 — one problem and six per cent — where Illinois, at the price of one constant (halve the stale ordinate), reaches 153/154 and 7045. A sixth of the gap is not the gap. The correction is real and it is in the right place; it is simply much smaller than a constant that was chosen for exactly this failure. *Our* 10676 and the 10560 of Table 2 are not the same algorithm — our iterate starts at the midpoint of the bracket, plain false position at an end — and the two must not be differenced.

4. **A logarithmic change of variable does not buy the wide brackets.** Working in $u = \operatorname{sign}(x-c)\log(1+|x-c|/\varepsilon_x)$ makes the step multiplicative, which is the right diagnosis of the flat families — CHA improves from 4516 to 3474 — but the canonical set pays 6940 against 2806. A march needs a *ratio* of distances while an identity on narrow brackets needs an *absolute* unit, and no parameter-free unit provides both.

5. **The prototype-aggregation form of the method loses.** A richer construction — prototypes over *pairs* of evaluated points, softmax weights over the whole memory, similarity on both the value and the abscissa axis — leads on eight hand-picked problems but loses decisively on the real set: 0 wins, 6 ties and 147 losses against Brent. The simplest form of the idea won.

6. **Every safeguard that replaces the step destroys the order.** Bracket-based and step-based replacement both drop the three-point rule from order 1.94 to 1.00 and cost 7260 and 7239 evaluations on Alefeld–Potra–Shi. Only projection preserves both. We record this because it is the trap the construction naturally falls into.

7. **Zero slack in the projection is degenerate.** A worst case of exactly bisection demands a halving at every step and leaves the projection no freedom at all.

8. **Regimes exist, at the level of a function family, and cannot be exploited by a switch.** The cube-root similarity beats the proposed rule on all five members of Chandrupatla's `fun4` — 32/32/40/47/59 against 89/93/113/125/153, a factor of 2.6 — and consistency across a family at different parameters is the signature of structure rather than luck. But the same similarity over the whole Chandrupatla set costs 4099 against 3220, so a switch chosen per benchmark is worse than no switch at all: 6905 against 6026 over both sets. The scale axis offers only 5.3% headroom even to an oracle handed the best scale per problem, two thirds of it inside a single family, and the winners are not separated by anything the method can see before it commits — the proposed scale wins on bracket widths from 10^0 to 10^{12}.

9. **The phase detector: derivation sound, estimator not.** Redoing the contraction law without the assumption $|f| \approx |f'(r)|e$ — taking $|f| \approx Ke^m$ instead — produces the march exponent on its own, $p = L/(L - \ln m)$ with $L = \ln(e/D)$, giving $p = 1$ at a simple root and dropping to 0.50 and 0.23 as the approach flattens, which is the neighbourhood of the cube root that wins those families. Substituted back it collapses to the multiplier m, measurable as $m = 1/(1 - \theta)$ with $\theta = ff''/(f')^2$ from three divided differences and no constant. It preserves the order (2.001) and improves the eight standard problems (82 against 84) — and loses on the full benchmark, 6304 against 6026. The estimator is the failure: on Alefeld–Potra–Shi, where almost every root is simple and the truth is $m = 1$, it reports $m > 1$ on 27.9% of steps with excursions to $m = 34.8$, against 35.8% on the flat families. 27.9 against 35.8 is no separation. A cruder variant, correcting by the fraction of its own promise the previous step kept, fails everything: the order collapses to 0.904 and the standard problems double to 161.

10. **One problem defeats every method**, `aps.13.00`, $f(x) = xe^{-1/x^2}$, all of whose derivatives vanish at the root and which underflows to exactly zero throughout $|x| < 0.086$; every method returns a machine zero at distance up to 0.086 from the root.

11. **The symmetric ratio similarity degenerates.** For $S(x, y) = \min(|x|, |y|)/\max(|x|, |y|)$ with y the target 0, $S \equiv 0$; the only usable reading is $1 - |f|/\sigma$, which saturates and stagnates (Table 1, last row).

9 Conclusion

A root finder can be derived from a similarity function alone, and the derivation shows where the content of such a method actually sits. It is not in the shape of S — nine shapes were tried and their ranking is dictated by a single number, the slope of $1 - S$ at the origin. It is not in making the similarity change faster — that overshoots exactly as readily as flatness stagnates.

It is in the *scale* against which similarity is judged: chosen, the method is brittle and merely linear; taken from the bracket, it is false position and three centuries old; read off the data, it is superlinear. And the safeguard that makes such a method safe must project the step, not replace it: replacing costs the order, projecting costs nothing on one benchmark and improves the method on the other.

The resulting rule has no tuning constant and carries a worst case of twice bisection plus one step. On the canonical Alefeld–Potra–Shi set it is level with TOMS 748 — 2806 evaluations against 2820 is a margin of 0.5% and should be read as a tie — and ahead of Brent by 2.9%, which it also beats 102 problems to 40 head to head. The claim the measurements support is parity with methods that carry tuned constants, not superiority over them.

This is the same movement as in parameter-free similarity control [6], where membership functions are eliminated by normalising the input space from the data rather than by choosing widths. Here the object normalised is the value axis, and the price of choosing it by hand is measured rather than asserted.

Future work

One weakness has a clear diagnosis and no fix inside this framework. On brackets many orders of magnitude wider than the distance to the root, the search must cross those decades, and an *additive* step of $|f|/|f'|$ is sized by the local slope, which knows nothing about how many decades remain. That is why the cube-root similarity, whose step is a fixed *fraction* of the bracket, beats the proposed rule 2.6-fold on those families, and why ITP keeps its lead on the Chandrupatla set.

The regime is therefore multiplicative versus additive movement, not one similarity function versus another — a change of the space the method works in, not of a coefficient inside it. The natural instrument is a logarithmic change of variable, $u = \operatorname{sign}(x-c)\log(1+|x-c|/s)$, in which an additive step becomes a geometric march and the exponent that the cube root fixes at $1/3$ by accident disappears as a free quantity. Two conditions decide whether that is a method or another refuted idea: the origin c and unit s must be read from the bracket rather than chosen, and the transform must be the identity to within rounding on narrow brackets, so that the wide-bracket families are not bought at the price of the canonical set. Every attempt reported in Section 7 failed on exactly that second condition.

References

[1] G. E. Alefeld, F. A. Potra and Y. Shi, "Algorithm 748: Enclosing zeros of continuous functions," *ACM Trans. Math. Softw.*, vol. 21, no. 3, pp. 327–344, 1995.

[2] R. P. Brent, *Algorithms for Minimization without Derivatives*. Englewood Cliffs, NJ: Prentice-Hall, 1973.

[3] T. R. Chandrupatla, "A new hybrid quadratic/bisection algorithm for finding the zero of a nonlinear function without using derivatives," *Adv. Eng. Software*, vol. 28, no. 3, pp. 145–149, 1997.

[4] I. F. D. Oliveira and R. H. C. Takahashi, "An enhancement of the bisection method average performance preserving minmax optimality," *ACM Trans. Math. Softw.*, vol. 47, no. 1, pp. 1–24, 2021.

[5] C. Ridders, "A new algorithm for computing a single root of a real continuous function," *IEEE Trans. Circuits Syst.*, vol. 26, no. 11, pp. 979–980, 1979.

[6] V. Sitnikov, "Parameter-free similarity control," preprint, 2026.

[7] M. Dowell and P. Jarratt, "A modified regula falsi method for computing the root of an equation," *BIT*, vol. 11, pp. 168–174, 1971.