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.

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 $\sigma = 2$.

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)

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.*

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 5.2 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 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.

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),

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

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 Negative results

1. **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.

2. **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.

3. **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.

4. **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.

5. **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).

8 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.

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.