WHEN DOES THE MOVE ALGEBRA GUIDE SEARCH? A TWO-CERTIFICATE CALCULUS FOR OPERATOR NON-COMMUTATIVITY IN LOCAL SEARCH

Anonymous authors
Paper under double-blind review

ABSTRACT

Local search evaluates the objective f at neighboring states to decide where to move or restart. Can the move set itself help make those decisions? We measure how far apart two moves take a state when applied in opposite orders. This non-commutativity defect κ needs no objective evaluations; its counterpart κ_f compares f at the two resulting states. We identify two conditions necessary to credit a gain to this signal. Under coupling, differences in f track the geometric disagreement. Under non-redundancy, κ_f cannot be recovered from the current and one-step neighbor values already available to standard local search. We test these conditions on three classes. On SAT, the signal is an exact function of one-step values. On TSP with 2-opt moves, κ is constant across tours. Both conditions hold on NK landscapes, yet standard NK shows no speed-up: states near the optimum do not carry more signal than average states. On an engineered NK family that satisfies this additional condition, κ_f-weighted draws from a basin-balanced restart pool require about 40% fewer restarts than uniform draws and 10–19% fewer than an equal-evaluation random two-step control. These counts exclude pool-scoring cost. Two of the three pre-search conditions can be decided before search, identifying cases where a gain cannot be credited to new information from the move set.

1 INTRODUCTION

Local search moves from a state x to a neighbor σx, using moves σ from a fixed set Σ such as bit flips, 2-opt reversals, or clause repairs (Hoos & Stützle, 2018). The objective f guides which neighbor to accept and when and where to restart. Simulated annealing and tabu search compare neighboring values of f (Kirkpatrick et al., 1983; Glover, 1989); evolutionary algorithms select on f (Droste et al., 2002); and landscape analysis summarizes f across the neighborhood (Weinberger, 1990; Ochoa et al., 2008). These methods use the move set to define the neighborhood. We ask what its structure tells the search.

The constraint is the evaluation budget. Local search spends most of it on the *one-step values* $f(\sigma x)$, $\sigma \in \Sigma$. Whether two moves interact can be hidden from those values and appear only two steps away. The structure of Σ might identify which two-step states deserve an evaluation. We seek to allocate those extra queries and to tell in advance when they cannot pay off.

For a state x and moves σ_i, σ_j, let $p = \sigma_i\sigma_j x$ and $q = \sigma_j\sigma_i x$ be the results of the two orders (Figure 1a). The *non-commutativity defect* $\kappa(x)$ averages their squared distance over move pairs. It vanishes if every pair commutes and requires no evaluation of f. The companion signal $\kappa_f(x)$ averages $(f(p) - f(q))^2$, at two evaluations per pair. UCDS makes decisions using κ_f; κ helps identify when those evaluations are useful. Section 3 shows that κ_f captures the antisymmetric part of the second-order structure of f, which symmetric Hessian summaries miss.

Yet a measurable signal need not help. Four questions must be addressed before κ_f can speed up search (Figure 1b): (1) *geometry*: does move order matter ($\kappa > 0$)? (2) *relevance*: does f respond, so that κ_f tracks κ? (3) *novelty*: does κ_f reveal anything beyond the one-step values? (4) *exploitation*: is the signal stronger near the optimum than at an average state? Questions 2

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

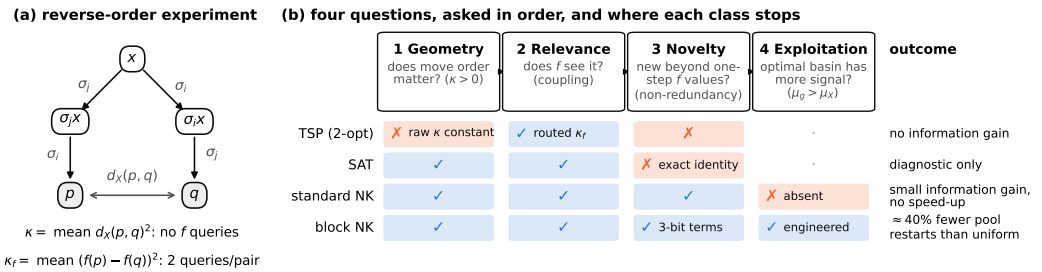

Figure 1: (a) Apply moves σ_i and σ_j to a state x in both orders. The defect κ is the mean squared distance between the two results p and q and needs no evaluation of f; κ_f is the mean squared difference $f(p) - f(q)$. (b) Four questions that must all be answered positively before κ_f can speed up restarts, and where each problem class stops. Questions 2 and 3 are the certificates of Section 4; question 4 is the exploitation condition of Lemma 2. Marks for questions 1–3 are theorems (Section 5); marks for question 4 are measured (Section 7). Passing questions 1–3 guarantees extra information, not speed.

and 3 form our *coupling* and *non-redundancy* certificates. Both are necessary to attribute a gain to information from the signal (Section 4). Benchmarks alone cannot settle question 3: a redundant signal may correlate with good decisions while adding no information the baseline lacks.

Section 5 answers these questions for three classes. On SAT, κ counts violated and fragile clauses with exact coupling, but an exact identity expresses that count as a function of the one-step values. On TSP with the full 2-opt neighborhood, κ is the same for every tour, and κ_f can be computed from the public city coordinates. On NK landscapes both certificates hold. The experiments confirm each prediction (Section 7): on standard NK the signal adds a little information but no speed-up, because question 4 fails; on an engineered NK family where it holds, κ_f-weighted draws from a basin-balanced restart pool require about 40% fewer restarts than uniform draws and 10–19% fewer than an equal-budget random control, not counting the cost of scoring the pool.

Contributions. (1) *A move-set signal:* κ and κ_f, where κ_f captures order-dependent second-order structure of f that Hessian-style summaries discard (Prop. 1, Thm. 1). (2) *A test for when the signal can help:* under an explicit model of what a baseline observes, the two certificates provide necessary conditions for crediting a gain to the signal rather than to information the baseline already had (Thms. 2–3, Lem. 1, Prop. 2). (3) *Answers for three classes, including two negative verdicts.* Before any experiment, the theory predicts that on TSP and SAT no gain can be credited to new information from the signal, and that on NK it adds information but does not guarantee speed (Thms. 4–7, Prop. 4). (4) *A method and a controlled test.* UCDS (Unified Curvature-Driven Search) uses κ_f to bias restarts; on a basin-balanced pool of the engineered family it requires fewer restarts than uniform, shuffled-signal, equal-budget random, and endpoint-permutation controls, whereas on standard instances it yields no speed-up (Section 7).

2 Related Work

Related work measures landscape structure, chooses operators, or evaluates two-step moves. Our question is whether move order reveals information beyond one-step values.

Landscape analysis. Ruggedness, autocorrelation, fitness-distance correlation, and local optima networks describe how f varies over a fixed neighborhood (Weinberger, 1990; Hordijk, 1996; Ochoa et al., 2008), and elementary-landscape (Walsh) analyses decompose f into eigenfunctions of the neighborhood graph (Reidys & Stadler, 2002). These tools hold the move set fixed and summarize f symmetrically across moves. They do not test what the order-dependent interaction adds beyond one-step values.

Operator selection and two-step moves. Adaptive operator selection and hyper-heuristics choose among move types by their observed effect on f (Burke et al., 2013), and Pei et al. (2023) relate operators by the correlation of the local optima they reach. Second-order neighborhoods (Torralbo et al., 2023) and pair-flip sampling in MaxSAT (Zheng et al., 2023) evaluate f two moves away to escape local optima. They use observed performance to select operator relations or two-step evaluations. Our test asks whether those evaluations add information unavailable one step away.

Stochastic local search heuristics. WalkSAT and GSAT score a variable by how many clauses its flip would repair or break (Selman et al., 1992; 1994), and Lin–Kernighan restricts 2-opt moves to candidate lists (Lin & Kernighan, 1973; Helsgaun, 2000). Both approaches use exactly the one-step information that defines our baseline class.

Geometry and hardness. Riemannian optimization and persistent homology use a given manifold or the sublevel sets of f (Absil et al., 2008; Carlsson, 2009), and Lie-bracket control moves along commutators of known vector fields (Dürr et al., 2013); we instead read geometry from the failure of discrete moves to commute. The overlap gap property states that in some random ensembles near-optimal solutions are either close together or far apart, with no intermediate distances, which provably defeats local and stable algorithms (Gamarnik & Sudan, 2014; Rahman & Virág, 2017; Gamarnik, 2021); clustering results play a similar role for random SAT (Achlioptas & Coja-Oghlan, 2008; Mézard et al., 2002; Ibrahimi et al., 2012). These results establish hard structure; our question is how a search might detect useful move interactions.

3 Setting and the Signal

This section fixes the setting, defines the signal, and identifies the part of the structure of f that it captures. An instance consists of a finite state space X with a metric d_X, a finite move set $\Sigma = \{\sigma_1, \ldots, \sigma_d\}$ of maps $X \to X$, and an objective $f : X \to \mathbb{R}$ to be minimized. A search algorithm knows (X, d_X, Σ) and learns about f only by evaluating it at states. We write $\sigma_i\sigma_j x$ for $\sigma_i(\sigma_j(x))$.

Definition 1 (Non-commutativity defect)**.** For $x \in X$, the non-commutativity defect is defined as:

$$\kappa(x) = \frac{1}{\binom{d}{2}} \sum_{1 \le i < j \le d} d_X\big(\sigma_i\sigma_j x,\ \sigma_j\sigma_i x\big)^2.$$

Intuitively, $\kappa(x)$ measures how much the order of moves matters at x; it is zero exactly when every pair commutes at x, and computing it requires only move applications and distances.

Definition 2 (f-order sensitivity)**.** With $p_{ij} = \sigma_i\sigma_j x$, $q_{ij} = \sigma_j\sigma_i x$, and $\Delta f_{ij}(x) = f(p_{ij}) - f(q_{ij})$, the f-order sensitivity is:

$$\kappa_f(x) = \frac{1}{\binom{d}{2}} \sum_{i<j} \Delta f_{ij}(x)^2.$$

In words, κ_f measures how strongly the objective responds to a change in move order. It costs two evaluations of f per pair. In practice, we estimate both quantities from n random pairs, with Hoeffding-type error bounds (Proposition 5, Appendix G).

Example. Let $X = \{0,1\}^N$ with Hamming distance, and let Σ contain the bit flips F_i and the adjacent swaps S_i, which exchange bits i and $i{+}1$. Flips commute with one another, so with flips alone $\kappa \equiv 0$: the signal is a property of the move set, not solely of the problem. Overlapping swaps fail to commute. If $(x_i, x_{i+1}, x_{i+2}) = (0,0,1)$, then $p = S_iS_{i+1}x$ has $(1,0,0)$ and $q = S_{i+1}S_ix$ has $(0,1,0)$ in these positions, so $d_X(p,q) = 2$ and the pair adds 4 to the sum defining κ without any call to f. The state q is a neighbor of x, whereas p is not: $f(p)$ is a two-step value that no one-step evaluation reveals. Whether κ_f is useful therefore depends on whether such differences $f(p) - f(q)$ are informative, which is precisely what the certificates in Section 4 test.

Because κ reads only the moves and the metric, one might wonder whether it is curvature of f in disguise. It is not:

Proposition 1 (Hessian independence). *There is no function Φ such that $\kappa(x) = \Phi(\nabla^2 f(x))$ for all instances, and none such that $\kappa(x) = \Phi(\|\nabla^2 f(x)\|)$ for all instances.*

Concretely, Appendix G gives one instance with a vanishing Hessian and $\kappa > 0$, and another with an arbitrarily large Hessian and $\kappa \equiv 0$. What κ_f does measure is the following:

Theorem 1 (Antisymmetric characterization). *Let $H_f(x; i, j) = f(\sigma_i\sigma_j x) + f(x) - f(\sigma_i x) - f(\sigma_j x)$ be the discrete Hessian. Then $\Delta f_{ij}(x) = H_f(x; i, j) - H_f(x; j, i)$. Consequently $\kappa_f(x)$ is the mean square of the antisymmetric part of $H_f(x; \cdot, \cdot)$, and $\kappa_f(x) = 0$ whenever $H_f(x; \cdot, \cdot)$ is symmetric.*

Schwarz's theorem makes the continuous Hessian symmetric. Theorem 1 shows that κ_f measures precisely the part lost under that symmetry, so a symmetric Hessian summary cannot detect it. If σ_i is the time-h flow of a vector field V_i, then $\Delta f_{ij} = h^2([V_j, V_i]f)(x) + O(h^3)$: in this smooth limit, κ_f measures f along Lie brackets, which vanish for coordinate directions.

4 Two Certificates: When Can κ_f Help?

Section 3 introduced a signal that ignores f (κ) and one that queries it (κ_f). In this section, we ask when κ_f can improve search: Section 4.1 gives the relevance condition, Section 4.2 the novelty condition, and Section 4.3 the exploitation condition that converts information into fewer restarts.

4.1 The First Certificate: Coupling

The first certificate asks whether f responds to the geometric disagreements that κ measures.

Definition 3 (Visible-mass coupling). Fix $0 \le c \le C < \infty$ and $\theta \in (0, 1]$. A state x is (c, θ, C)-*visible* if every pair satisfies $|\Delta f_{ij}(x)| \le C\, d_X(p_{ij}, q_{ij})$, and the *visible* pairs, namely those for which $|\Delta f_{ij}(x)| \ge c\, d_X(p_{ij}, q_{ij})$, carry at least a θ-fraction of the curvature mass $\sum_{i<j} d_X(p_{ij}, q_{ij})^2$. A class is (c, θ, C)-coupled on $R \subseteq X$ if every $x \in R$ is (c, θ, C)-visible.

Each parameter controls a distinct possibility: (i) the ceiling C, a Lipschitz constant of f at the commutator endpoints, depends on how steep f is there; (ii) the floor c, the smallest response $|\Delta f|/d_X$ that counts as visible, depends on how strongly f reacts to reordering; and (iii) the fraction θ depends on how much curvature mass falls on invisible pairs, such as ties with $\Delta f = 0$.

Theorem 2 (The bridge). *At every (c, θ, C)-visible state x, $c^2\theta\, \kappa(x) \le \kappa_f(x) \le C^2\kappa(x)$.*

In effect, whenever f detects a fixed fraction of the geometric disagreement, the f-free defect pins the f-dependent one between two positive multiples of itself. The inequality itself is elementary (proof in Appendix G); the substantive work is to establish that the floor (c, θ) is positive in concrete classes, which Section 5 carries out. Visibility also gives an exact characterization of failure:

Theorem 3 (Failure characterization). (a) *If all moves commute at x, then $\kappa_f(x) = 0$ for every f.* (b) *$\kappa(x) > 0$ does not imply $\kappa_f(x) > 0$.* (c) *For finite Σ and $\kappa(x) > 0$, $\kappa_f(x) > 0$ if and only if x is (c, θ, C)-visible for some $c, \theta > 0$ and some C.*

For part (b), take the instance from Proposition 1 with $\kappa > 0$ and choose f to take the same value at the two endpoints. Part (c) essentially requires that some reordering change the objective: visibility, rather than non-commutativity alone, couples κ to f. (When $\kappa(x) = 0$, the visibility condition holds vacuously, which is why part (c) assumes $\kappa(x) > 0$.)

Estimating the floor. A probe of n random pairs yields ratios $r_k = |\Delta f_k|/d_k$ for pairs with $d_k > 0$. UCDS uses $\hat{c} = \mathrm{quantile}_{0.1}\{r_k\}$ as a plug-in floor at $\theta = 0.9$, and $\hat{C} = \mathrm{quantile}_{0.9}\{r_k\}$ as a confidence weight rather than as the Lipschitz constant C. This quantile is unweighted, whereas Definition 3 weights pairs by d_X^2. Fortunately, on NK and SAT every non-commuting pair has the same distance (Theorems C.1 and D.1), so the two constructions coincide; on TSP, $\hat{c}$ is only a heuristic. Appendix A adds that branch points, where the two move orders lead to different basins, each contribute a lower bound to κ.

4.2 THE SECOND CERTIFICATE: NON-REDUNDANCY

Coupling ensures that κ_f tracks f, but it does not ensure that κ_f tells the search anything new. To make this distinction precise, we must first specify what a search is allowed to know.

Query model. A search knows the *structure*: the triple (X, d_X, Σ) together with any instance data declared public. Which data are public is a modelling choice that we call the *oracle boundary*. We treat SAT formulas and TSP city coordinates as public because solvers read them directly, and NK tables as hidden because NK serves as a black-box benchmark. The search learns about f only through queries $x \mapsto f(x)$.

Definition 4 (First-order field)**.** The first-order field at x collects the one-step values:

$$\mathcal{I}_1(x) = \big(f(x),\ (f(\sigma x))_{\sigma\in\Sigma}\big).$$

In practical terms, $\mathcal{I}_1(x)$ is the objective information obtained by evaluating the current state and all its immediate neighbors.

Definition 5 (Σ-redundancy)**.** A signal s is Σ-*redundant* on a class $\mathcal{C}$ of instances if there is a function Ψ, with no restriction on its computational power, such that $s(x) = \Psi\big(x, \mathrm{structure}, \mathcal{I}_1(x)\big)$ for every instance in $\mathcal{C}$ and every $x \in X$. Otherwise, s is *query-novel*.

Definition 5 asks whether one-step queries and public data already determine a signal. The objective enters this test only through $\mathcal{I}_1$. The raw defect κ is always redundant because it uses no f; the relevant certificate concerns κ_f, whose computation may require two-step queries.

Lemma 1 (Simulation lemma)**.** *Let Π_1 be the class of search policies whose access to f is limited to zeroth- and first-order queries, with unrestricted memory and computation. If s is Σ-redundant on $\mathcal{C}$, every policy using s can be simulated, decision by decision, by a policy in Π_1.*

In words, a redundant signal cannot explain a gain over Π_1, which contains the standard baselines (Lemma B.1). It may still save computation, as the $O(m)$ SAT counter does. The result has two qualifications. Because Ψ is unrestricted, declaring instance data public makes every signal redundant, even the optimum's location; verdicts for public-data classes thus depend on the oracle boundary. Also, novelty is local: after moving, a Π_1 policy can reach two-step values by spending more queries.

Proposition 2 (Two-certificate necessity)**.** *Consider a search policy that uses the signal only through κ_f-weighted decisions. A gain of this policy over Π_1 can be attributed to the signal only if* (i) $\kappa_f \not\equiv 0$ *on the states it visits and* (ii) κ_f *is query-novel.*

By Theorem 3(c), condition (i) means that some visited state with $\kappa > 0$ is visible, whereas condition (ii) is the novelty certificate (proof in Appendix G); (i) depends on the landscape and (ii) on the oracle boundary. We do not claim sufficiency: Section 7 presents a class in which both hold yet no speed-up follows.

4.3 EXPLOITATION

Information translates into fewer restarts only under one additional condition, which concerns where the signal is concentrated.

Lemma 2 (Curvature-restart formula)**.** *Let uniform restarts require $E_{\mathrm{rand}} = |X|/|B_g|$ restarts in expectation to hit the optimal basin B_g, and suppose biased restarts are drawn from $q_\alpha(x) \propto 1 + \alpha\kappa_f(x)$ with $\alpha > 0$. If μ_g and μ_X are the means of κ_f over B_g and over X, then biased restarts require $E_{\mathrm{bias}} = E_{\mathrm{rand}}\,(1 + \alpha\mu_X)/(1 + \alpha\mu_g)$ restarts in expectation.*

Biased restarts therefore outperform uniform ones exactly when $\mu_g > \mu_X$, meaning that the optimal basin carries more signal than the average state. We call this the *exploitation condition*; it is question 4 in Figure 1b.

Proposition 3 (Break-even gate)**.** *If probing multiplies the cost of each biased restart by $1 + \beta$, biased restarts have lower expected cost than uniform ones if and only if $\alpha(\mu_g - \mu_X) > \beta(1 + \alpha\mu_X)$; for large α, this simplifies to $\mu_g > (1 + \beta)\mu_X$.*

The left side measures the benefit, governed by how far the optimal basin out-signals the average state; the right side measures the cost, governed by the probe price β.

5 Deciding the Certificates in Three Classes

So far, the signal, the certificates, and the exploitation condition have been stated for an arbitrary move set. Whether the certificates hold depends on the problem class, and this section determines them for TSP, NK, and SAT. Two verdicts are negative, allowing the theory to predict in advance where the signal cannot help (Figure 1b; details in Table S1, Appendix A).

TSP. States are tours of n cities, d_X counts the edges on which two tours differ, and f is tour length. A positional 2-opt move $R_{a,b}$ reverses the tour segment in positions $a{+}1, \ldots, b$.

Theorem 4 (Flatness)**.** *Let $\Sigma_{\mathrm{pos}} = \{R_{a,b} : 0 \le a < b \le n-1\}$. There is a constant $\kappa_0(n)$, depending only on n, such that $\kappa(x) = \kappa_0(n)$ for every tour x.*

The reason is that a positional move never consults the identity of the city occupying any position; for example, $\kappa_0(9) = 24.22$ for every nine-city tour.

Corollary 1 (Routing is forced)**.** *Under Σ_{pos}, raw κ does not distinguish any two tours.*

This flatness result overturned our initial TSP success criterion: κ does not separate boundary from interior local optima. We recorded that null prediction before the experiments, which confirmed it (Section 7). Discriminative use must instead go through κ_f, whose Euclidean coupling is of constant order (Theorem E.3), or through candidate-restricted neighborhoods that break positional symmetry, as Lin–Kernighan candidate lists do. Unfortunately, κ_f fails the second certificate: city coordinates are public, so κ_f is determined by the state and the structure without any query. By Lemma 1, no gain can then be credited to it as information about f. The verdict follows from the declared boundary; if coordinates were hidden, the channel could not be recovered from one-step values (Theorem E.3).

NK. States are bit strings in $\{0,1\}^N$ with Hamming distance, the moves are the flips and adjacent swaps from the example in Section 3, and $f(x) = \frac{1}{N}\sum_k g_k(x_k, \ldots, x_{k+K})$ sums table contributions over windows of $K{+}1$ consecutive bits, with table entries drawn i.i.d. from $U[0,1]$ (Kauffman, 1993).

Theorem 5 (NK coupling)**.** *On adjacent NK landscapes, let $\bar{\rho}$ be the ratio of a typical visibility floor to the Lipschitz ceiling (Appendix C). The expected defect and, as $N \to \infty$, this ratio are:*

$$\mathbb{E}[\kappa] = \frac{11N-14}{(2N-1)(N-1)}, \qquad \bar{\rho}(K) = \frac{\sqrt{11K+23}}{\sqrt{264}\,(K+1)} = \Theta(K^{-1/2}).$$

Moreover, for every fixed state and every $\varepsilon \in (0, \frac{1}{2})$, the expected visible mass at visibility threshold $\varepsilon/(2N)$ is at least $1-2\varepsilon$.

The formula says that coupling weakens like $K^{-1/2}$ as interaction width grows, independently of N ($\sqrt{264} \approx 16.25$). The expectation is over random tables at a fixed state; concentration gives high-probability per-state guarantees (Appendix C). Proposition 2 requires only one visible visited state.

Proposition 4 (NK novelty)**.** *On adjacent NK landscapes whose tables are accessible only through f, κ_f is query-novel.*

The proof changes one table entry seen by a two-step endpoint but by no one-step state. Thus $\mathcal{I}_1(x)$ stays fixed while $\kappa_f(x)$ changes (machine witness with $N{=}10$, $K{=}2$). This is the extra information illustrated in Section 3. Only swap–swap (braid) pairs can access such entries: every flip–swap pair has both endpoints in the one-step ball, with $\{F_jS_jx, S_jF_jx\} = \{F_jx, F_{j+1}x\}$ (Lemma C.6), so $2(N{-}1)$ of the $3N{-}4$ non-commuting pairs contribute only redundant terms. If the tables were public, κ_f would be redundant: novelty is a property of the query interface rather than of the problem alone. The raw defect has the wall-counter form $\kappa(x) = \frac{8(N-1)+4T(x)}{(2N-1)(N-1)}$, where $T(x)$ counts non-constant windows (x_i, x_{i+1}, x_{i+2}), namely positions next to a domain wall (Theorem C.1).

SAT. States are assignments in $\{0,1\}^n$ of a 3-SAT formula with m clauses, each over three distinct variables, and $f(x) = V(x)$ counts violated clauses. The moves are the flips F_i and, for each clause C, a repair move r_C that flips the lowest-index variable of C when C is violated and acts as the identity otherwise.

Algorithm 1 UCDS-Core

Input: space X, moves Σ, objective f, start x_0, budget B
Output: best state found

assert a non-commuting pair exists in Σ (else warn $\kappa \equiv 0$; greedy+restart)
$x \leftarrow x_0$; $best \leftarrow x_0$
while $B > 0$ **do**
  if time to probe **then**
    sample n pairs $\rightarrow (p_k, q_k, d_k)$; $\hat{\kappa} \leftarrow \overline{d_k^2}$
    evaluate f on $\{d_k > 0\} \rightarrow \Delta f_k$; $\hat{\kappa_f} \leftarrow \overline{\Delta f_k^2}$
    $r_k \leftarrow |\Delta f_k|/d_k$; $\hat{c} \leftarrow \text{quantile}_{0.1}(r_k)$; $\hat{C} \leftarrow \text{quantile}_{0.9}(r_k)$
  end if
  regime $\leftarrow$ R1 if $\hat{\kappa} \le \kappa_{\min}$; otherwise R2 if $\hat{c} \ge c_{\min}$; otherwise R3
  if R1 or R3 **then**
    $x \leftarrow$ best-improvement neighbor (Thm. 3)
  else
    $x \leftarrow \arg\max_k [w\,\hat{\kappa_f}(\sigma_k x) - f(\sigma_k x)]$
  end if
  on stagnation: restart $\sim q_\alpha \propto 1 + \alpha\hat{\kappa_f}\hat{C}$ w.p. $1-\varepsilon$, else uniform (Lem. 2)
  every epoch: if $\alpha(\hat{\mu}_g - \hat{\mu}_X) \le \beta(1 + \alpha\hat{\mu}_X)$ shrink probes (Prop. 3)
end while
return *best*

Theorem 6 (SAT counter)**.** *For every x, $\kappa(x) = \big(3V(x) + U(x) + R(x)\big)/\binom{n+m}{2}$, where $U(x)$ counts clauses with exactly one true literal and $R(x)$ counts non-commuting pairs of repair moves.*

In words, the counter assigns three units to a violated clause and one to a fragile clause, echoing WalkSAT's break count. Coupling holds with floor $c = 1$ for every pair whose endpoints differ in f (Appendix D).

Theorem 7 (SAT redundancy)**.** (a) *For every assignment x:*

$$3V(x) + U(x) = 6f(x) + \sum_{j=1}^{n} \big[f(F_j x) - f(x)\big].$$

(b) *The full defect κ, including R, is a function of* $(x, \text{structure})$.

The identity reconstructs the counter from six times the objective plus the one-step differences SAT solvers already compute. Part (b) extends this redundancy to the full defect. By Lemma 1, no gain can be credited to it as information about f. Unlike the TSP verdict, part (a) is independent of the oracle boundary: it recovers the counter from one-step values alone, without reading the formula. Overall, coupling holds in all three classes but only NK passes non-redundancy, which is why the ordering by coupling strength (SAT, TSP, NK) is the reverse of the ordering by usefulness.

6 UCDS

The certificates now guide the search rule. UCDS (Algorithm 1) consults the signal at three points, each tied to a result above.

When the moves essentially commute ($\hat{\kappa} \le \kappa_{\min}$), the search takes the best neighbor (R1). When the signal is visible ($\hat{c} \ge c_{\min}$), it ranks neighbors y by $w\hat{\kappa_f}(y) - f(y)$ with a tuned weight $w > 0$ (R2). Otherwise, the signal is blind to f, and the search again takes the best neighbor (R3); by Theorem 3(b), this protects it from high but useless κ. Restarts are drawn with probability proportional to $1 + \alpha\hat{\kappa_f}\hat{C}$, following Lemma 2 for a non-negative score. The implementation approximates that rule on a budget-charged pool of 128 states (Appendix F). The gate in Proposition 3 reduces probing when the estimated benefit falls below its cost, with $\hat{\mu}_g$ estimated over the incumbent's basin. UCDS never uses raw κ on TSP and uses the SAT counter only as a diagnostic. Thresholds are probe quantiles; all other parameters are tuned with the same 20-configuration budget as every baseline, and all moves and evaluations, including probes, are charged to one budget (Appendix F).

7 EXPERIMENTS

The experiments test the predictions from Sections 4 and 5: the closed forms, the restart gain when all four questions hold, controls that isolate its source, and standard instances where question 4 fails.

Setup. All methods share one budget counting both move applications and evaluations of f, use paired seeds for 30 runs, and are tuned with 20 configurations on disjoint instances. The baselines are random-restart best-improvement hill climbing (RHC), simulated annealing with geometric cooling (SA; Kirkpatrick et al. 1983), adaptive SA with acceptance rate kept between 0.15 and 0.45, tabu search over recently visited states (Glover, 1989), a (1+1)-EA with Poisson(1) mutations retained when no worse (Droste et al., 2002), and iterated local search with random-move perturbations (ILS; Lourenço et al. 2003). We report Vargha–Delaney $\hat{A}$ effect sizes, namely the probability that a UCDS run or batch beats one from another method, Holm-corrected p-values, and predictions frozen beforehand in a ledger (Table S3, Appendix F).

Closed forms. Monte Carlo measurements of $\mathbb{E}[\kappa]$ and $\mathbb{E}[\kappa_f]$ on random NK instances ($N \in \{20, 50\}$, $K \in \{2, 4, 8\}$) match Theorem 5 within 0.1% and 1.6%, with no fitted parameters; the measured wall-versus-interior gap in κ stays below its analytic bound (Figure S1, Appendix C).

Block NK: all four questions answered. The engineered family (Theorem C.3) splits $N = mL$ bits into m non-interacting L-bit blocks. Adjacent-bit disagreements are penalized, runs of three zeros are rewarded, and per-block enumeration proves that the 1-flip optima are exactly the 2^m block-constant strings and that all-zero is the unique global optimum. The 3-bit term creates novelty: without it, κ_f follows from one-step values, while with it 26–30% of braid pairs are query-novel (Proposition C.7). Restarts use a pool with 16 states per basin, so uniform draws succeed with probability exactly 2^{-m}, and the exploitation condition holds exactly (9.9 rather than 16 expected restarts for m=4, L=8; Appendix F). Biasing draws by $\kappa_f \hat{C}$ cuts restarts by 37.5–40.9% versus uniform across $m \in \{4, 8\}$, $L \in \{8, 16\}$ (Figure 2). A shuffled-signal control matches uniform, and UCDS beats it by 38–50% ($\hat{A} \geq 0.99$). Because pool scoring ($3N-4$ pair queries per state) is excluded, these numbers measure the restart distribution rather than end-to-end speed. The construction followed four documented failures (Appendix F).

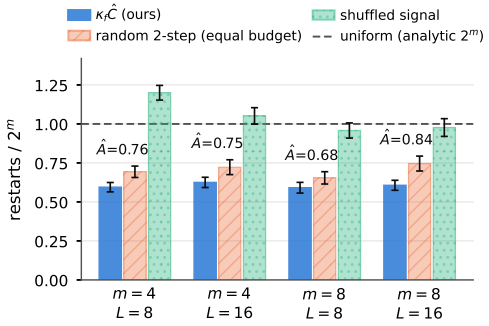

Figure 2: Restarts to reach the optimum on block NK, relative to the uniform expectation 2^m (mean of 30 paired batches of 50 runs; bars are 95% intervals). $\hat{A}$ compares UCDS with the equal-budget random control.

Equal budget, random allocation. Beating uniform restarts establishes information, but not efficient allocation; arbitrary two-step evaluations might work equally well. A third control receives the same pair queries per state ($3N-4$, each costing four applications and two evaluations) but spends them on random two-step paths. UCDS still wins every cell by 9.7–18.7% ($\hat{A}$ 0.68–0.84, Holm $p \leq 0.016$), the budget-matched margin; the 40% gain is versus uniform draws.

Isolating the pairing. The random control also changes endpoint coverage. A fourth control keeps the same endpoints and changes only the pairings, scoring $p_a = \sigma_i \sigma_j x$ against $q_{\pi(a)}$ for a random derangement π; both arms therefore evaluate the same states at the same cost. Commutator pairing wins all four cells by 9.9–21.3% ($\hat{A}$ 0.708–0.937, Holm $p \leq 0.0088$), and a same-seed rerun is byte-identical. Controls resampling pairs from the same support fail the preregistered gate, localizing the benefit to exact reverse-order pairings.

Standard NK: information without speed-up. On standard random NK, both certificates hold but question 4 fails, and signal-on and signal-off results are identical. The remaining test is informa-

tional: under frozen protocol X-1 (20 instances per domain, 200 candidate pairs each; Appendix F), we ask whether the signal predicts the better short greedy rollout beyond one-step values. It does slightly on NK (partial Spearman 0.052, 95% CI $[0.013, 0.093]$), not on TSP (CI $[-0.040, 0.055]$), while ΔAUC is null to slightly negative in both (Figure 1b).

TSP: attributing the gain. Across 24 matched-budget TSP instances, routed UCDS has the best median tour on 23, though candidate lists do most of the work: signal-off B6 remains within about 1% (Table S2, Appendix F). The isolated signal effect is near 1% (UCDS beats B6 on 18/24, Holm $p = 0.008$; a raw-κ variant wins on 19/24). Since the channel is redundant, this margin cannot be credited to new information about f; it may be a cheaper first-order summary. We compare general-purpose metaheuristics, not specialized solvers (Helsgaun, 2000; Applegate et al., 2006).

Predicted nulls. Raw κ has AUC 0.500 for boundary versus interior TSP optima over 1280 labeled tours, as Corollary 1 requires, whereas kNN-restricted neighborhoods restore discrimination. On random 3-SAT, the counter stays within the null's 15% band for $\alpha \in [0.5, 6]$ and peaks at $\alpha = 1$, far from $\alpha \approx 4.27$ (Mitchell et al., 1992): a counting artifact rather than a hardness marker. A frozen coupling estimate predicts winning TSP families (Spearman $\rho = 0.53$) but not instances within families, and the scale is stable across size ($\hat{\rho} = 0.069$–0.076, $n = 50$–500). Robustness checks appear in Appendix F.

8 Conclusion and Limitations

We asked whether the move set informs local search beyond one-step values of f. Four questions organize the answer, including two certificates that can be decided per class before search. As predicted, SAT and TSP provide no creditable new information, NK carries some information, and the experiments agree. When exploitation was engineered, the signal reduced balanced-pool restarts under controls for budget, coverage, and pairing.

Limitations remain. The performance evidence concerns one engineered family and a basin-balanced pool with scoring cost excluded; it establishes an improved restart distribution, not end-to-end speed, while standard NK yields information without speed-up. We did not disentangle query-novel braid pairs from redundant flip–swap pairs or compare with one-step restart rules, which the zero reward may partly favor. Novelty is boundary-relative and informational. The estimate $\hat{c}$ matches Definition 3 only for equal pair distances (NK, SAT), sufficiency is unproved, and the pairing effect uses one construction and a fixed pool. A natural next step is to predict favorable W-model regimes in IOHprofiler (Weise & Wu, 2018; Doerr et al., 2020) from the certificates and test UCDS on held-out instances.

AI USE STATEMENT

In this work, we used generative AI tools for three purposes: to help draft initial versions of the manuscript text, to assist in developing research ideas during the ideation stage, and to assist with the experimental work, including experiment code. We have reviewed all AI-assisted work. AI-drafted text was revised by the authors, AI-assisted ideas were evaluated and selected by the authors, and AI-assisted experiment code was checked against the closed-form regression tests and the frozen logs released with the paper (Appendix F), from which every reported number is reproduced. We take responsibility for the final content of this work, including text, claims or artifacts produced with the aid of generative AI.

REPRODUCIBILITY STATEMENT

Complete statements and proofs of the theoretical results are in Appendices A–E and G, each numerical closed form carrying its machine-verification note. Appendix F specifies instances, budgets, per-decision query accounting, seeds, the incremental-value protocol, the equal-budget and endpoint-permutation attribution protocols, and the block-restart development log. The anonymous supplementary material contains the code, closed-form regression tests, the frozen prediction ledger, the NK query-novelty witness, the configuration, status report and manifest of every reported run, seed files, and per-figure regeneration scripts (Figures 1 and 2 are regenerated by the scripts in the supplementary folder `figure_scripts`).

REFERENCES

P-A Absil, Robert Mahony, and Rodolphe Sepulchre. *Optimization algorithms on matrix manifolds*. Princeton University Press, 2008.

Dimitris Achlioptas and Amin Coja-Oghlan. Algorithmic barriers from phase transitions. In *2008 49th Annual IEEE Symposium on Foundations of Computer Science*, pp. 793–802. IEEE, 2008.

David L. Applegate, Robert E. Bixby, Vašek Chvátal, and William J. Cook. *The Traveling Salesman Problem: A Computational Study*. Princeton University Press, 2006. ISBN 9780691129938.

Edmund K Burke, Michel Gendreau, Matthew Hyde, Graham Kendall, Gabriela Ochoa, Ender Özcan, and Rong Qu. Hyper-heuristics: A survey of the state of the art. *Journal of the Operational Research Society*, 64(12):1695–1724, 2013.

Gunnar Carlsson. Topology and data. *Bulletin of the American Mathematical Society*, 46(2):255–308, 2009.

Carola Doerr, Furong Ye, Naama Horesh, Hao Wang, Ofer M. Shir, and Thomas Bäck. Benchmarking discrete optimization heuristics with IOHprofiler. *Applied Soft Computing*, 88:106027, 2020. doi: 10.1016/j.asoc.2019.106027.

Stefan Droste, Thomas Jansen, and Ingo Wegener. On the analysis of the (1+ 1) evolutionary algorithm. *Theoretical Computer Science*, 276(1-2):51–81, 2002.

Hans-Bernd Dürr, Miloš S. Stanković, Christian Ebenbauer, and Karl Henrik Johansson. Lie bracket approximation of extremum seeking systems. *Automatica*, 49(6):1538–1552, 2013. doi: 10.1016/j.automatica.2013.02.016.

David Gamarnik. The overlap gap property: A topological barrier to optimizing over random structures. *Proceedings of the National Academy of Sciences*, 118(41):e2108492118, 2021.

David Gamarnik and Madhu Sudan. Limits of local algorithms over sparse random graphs. In *Proceedings of the 5th conference on Innovations in theoretical computer science*, pp. 369–376, 2014.

Fred Glover. Tabu search—part i. *ORSA Journal on computing*, 1(3):190–206, 1989.

Keld Helsgaun. An effective implementation of the Lin–Kernighan traveling salesman heuristic. *European Journal of Operational Research*, 126(1):106–130, 2000. doi: 10.1016/S0377-2217(99)00284-2.

Holger H. Hoos and Thomas Stützle. Stochastic local search. In *Handbook of approximation algorithms and metaheuristics*, pp. 297–307. Chapman and Hall/CRC, 2018.

Wim Hordijk. A measure of landscapes. *Evolutionary Computation*, 4(4):335–360, 1996.

Morteza Ibrahimi, Yashodhan Kanoria, Matt Kraning, and Andrea Montanari. The set of solutions of random xorsat formulae. In *Proceedings of the twenty-third annual ACM-SIAM symposium on Discrete Algorithms*, pp. 760–779. SIAM, 2012.

Stuart A. Kauffman. *The Origins of Order: Self-Organization and Selection in Evolution*. Oxford University Press, 1993. ISBN 9780195079517.

Scott Kirkpatrick, C. Daniel Gelatt, Jr., and Mario P. Vecchi. Optimization by simulated annealing. *Science*, 220(4598):671–680, 1983.

Shen Lin and Brian W Kernighan. An effective heuristic algorithm for the traveling-salesman problem. *Operations Research*, 21(2):498–516, 1973.

Helena R. Lourenço, Olivier C. Martin, and Thomas Stützle. Iterated local search. In Fred Glover and Gary A. Kochenberger (eds.), *Handbook of Metaheuristics*, International Series in Operations Research & Management Science, pp. 320–353. Springer, 2003. doi: 10.1007/0-306-48056-5_11.

Marc Mézard, Giorgio Parisi, and Riccardo Zecchina. Analytic and algorithmic solution of random satisfiability problems. *Science*, 297(5582):812–815, 2002.

David G. Mitchell, Bart Selman, and Hector J. Levesque. Hard and easy distributions of SAT problems. In *Proceedings of the Tenth National Conference on Artificial Intelligence*, pp. 459–465. AAAI Press, 1992.

Gabriela Ochoa, Marco Tomassini, Sébastien Vérel, and Christian Darabos. A study of NK landscapes' basins and local optima networks. In *Proceedings of the 10th Annual Conference on Genetic and Evolutionary Computation (GECCO)*, pp. 555–562, 2008.

Jiyuan Pei, Hao Tong, Jialin Liu, Yi Mei, and Xin Yao. Local optima correlation assisted adaptive operator selection. In *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO)*, pp. 339–347, 2023. doi: 10.1145/3583131.3590399.

Mustazee Rahman and Bálint Virág. Local algorithms for independent sets are half-optimal. *The Annals of Probability*, 45(3):1543–1577, 2017. doi: 10.1214/16-AOP1094.

Christian M Reidys and Peter F Stadler. Combinatorial landscapes. *SIAM Review*, 44(1):3–54, 2002.

Bart Selman, Hector J. Levesque, and David G. Mitchell. A new method for solving hard satisfiability problems. In *Proceedings of the Tenth National Conference on Artificial Intelligence*, pp. 440–446, 1992.

Bart Selman, Henry A. Kautz, and Bram Cohen. Noise strategies for improving local search. In *Proceedings of the Twelfth National Conference on Artificial Intelligence*, pp. 337–343. AAAI Press, 1994.

Manuel Torralbo, Leticia Hernando, Ernesto Contreras-Torres, and Jose A. Lozano. On the use of second order neighbors to escape from local optima. In *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO)*, pp. 375–383, 2023. doi: 10.1145/3583131.3590473.

Edward Weinberger. Correlated and uncorrelated fitness landscapes and how to tell the difference. *Biological Cybernetics*, 63(5):325–336, 1990.

Thomas Weise and Zijun Wu. Difficult features of combinatorial optimization problems and the tunable W-model benchmark problem for simulating them. In *Proceedings of the Genetic and Evolutionary Computation Conference Companion (GECCO)*, pp. 1769–1776, 2018. doi: 10.1145/3205651.3208240.

Jiongzhi Zheng, Kun He, and Jianrong Zhou. Farsighted probabilistic sampling: A general strategy for boosting local search MaxSAT solvers. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, pp. 4132–4139, 2023. doi: 10.1609/aaai.v37i4.25529.