Mixed-Variable Optimisation as a Metric Product Space: Transient Categorical Geometry and a Hierarchy of Local Optimality

J. F. A. Madeira*

Abstract

Mixed-variable optimisation problems involving continuous, integer, and categorical variables are ubiquitous Li et al. [2013], yet the mathematical treatment of their search spaces remains largely ad hoc. In particular, categorical variables are typically treated as structureless—finite sets of labels with no order, distance, or geometry. We argue that this assumption is both unnecessary and limiting. We propose that categorical variables can admit an *operational (transient)* combinatorial geometry through cyclic-ordering-induced structures, and that mixed search spaces can be formalised as weighted Cartesian products of metric spaces. We introduce a hierarchy of local optimality—instantaneous, robust, and categorical—capturing the degree to which a local optimum is independent of the chosen geometry on the categorical components. We prove that the weighted product metric remains valid under dynamic rotation of the categorical component, establish a tight coverage bound showing that $\lceil (k-1)/2 \rceil$ cyclic orderings are both sufficient and minimal for full categorical adjacency coverage, and prove uniform metric equivalence across all ordering rotations. We show that standard epoch-based descent schemes under recurrent ordering schedules naturally produce accumulation points satisfying robust—and, under coverage, categorical—local optimality. The framework provides a structural foundation for mixed-variable optimisation independent of any particular algorithmic paradigm.

Keywords: mixed-variable optimisation; categorical variables; metric product spaces; transient geometry; local optimality hierarchy; derivative-free optimisation

1 Introduction

Continuous and integer variables in optimisation have well-understood geometry: Euclidean spaces and lattice structures provide natural notions of distance, neighbourhood, and local optimality. Categorical variables—finite sets of unordered labels such as material types, solver configurations, or neural-network architectures—have no *canonical* geometric structure. The standard treatment assigns them the trivial discrete metric ($d(a, b) = 1$ whenever $a \neq b$), a modelling convention commonly adopted in mixed-variable optimisation and surrogate modelling Hutter et al. [2011], Ru et al. [2020], Garrido-Merchán and Hernández-Lobato [2020], effectively declaring all non-identical categories equidistant. This creates a structural asymmetry in mixed-variable optimisation: continuous and integer components admit intrinsic geometric structure, while categorical components are typically handled through ad hoc encodings or component-wise random mutations.

This paper argues that the assumed structurelessness of categorical variables is a *convention*, not a necessity. Categories remain ontologically unordered—no fixed ordering is intrinsic to the variable. However, they can admit *transient operational geometry*: a temporary combinatorial

*IDMEC, Instituto Superior Técnico, Universidade de Lisboa, Portugal.
ISEL – Instituto Superior de Engenharia de Lisboa, Instituto Politécnico de Lisboa, Portugal.
E-mail: `aguilarmadeira@tecnico.ulisboa.pt`

structure, induced by a cyclic ordering, that is rotated over time so that no single ordering is privileged. This separation between ontological nature and operational structure is the conceptual foundation of the present work.

Building on this principle, we develop a geometric framework for mixed search spaces. The main contributions are:

(i) A formalisation of *transient geometry* on categorical variables via cyclic-ordering-induced metrics that rotate over time (Section 4.2).

(ii) A rigorous model of mixed spaces as *weighted Cartesian products of metric spaces*, valid under dynamic rotation of the categorical component (Section 4.3).

(iii) A *hierarchy of local optimality*—instantaneous, robust, and categorical—capturing the extent to which optimality is independent of the chosen categorical geometry, together with a monotone robustness result (Section 4.5).

(iv) A *coverage theorem*: cyclic adjacency achieves full categorical coverage in $\lceil (k-1)/2 \rceil$ orderings for a k-level variable, with a tight bound based on a classical Hamiltonian decomposition (Section 4.6).

(v) An *algorithmic interpretation* showing that recurrent ordering schedules naturally lead to robust—and, under coverage, categorical—local optimality (Section 4.8).

(vi) A *uniform metric equivalence* result: all dynamic metrics induce the same topology, addressing the concern that ordering rotation "changes the space" (Section 5.3).

The central contribution is that categorical variables, despite lacking canonical order, admit a rigorous geometric treatment through transient structures—and that this treatment leads to a well-defined hierarchy of local optimality with quantified coverage bounds. In particular, the framework distinguishes three levels of optimality (instantaneous, robust, and categorical), proves that the strongest level is reachable in a finite and explicitly bounded number of ordering rotations (Proposition 4.31), and shows that recurrent descent schemes naturally produce accumulation points satisfying categorical local optimality (Section 4.8). These are structural results about the mixed search space itself, independent of any particular algorithmic realisation.

Related work and positioning. Mixed-variable black-box optimisation has been approached from several angles. In derivative-free direct search, pattern search and MADS-type frameworks have been extended to handle categorical and mixed-variable problems Audet and Dennis [2006], Abramson et al. [2009], Audet and Le Digabel [2024], Vicente and Custódio [2012]. In surrogate-based and Bayesian optimisation, a central difficulty is the definition of kernels and distance functions for categorical and hybrid domains; representative approaches include multi-armed-bandit decompositions for continuous–categorical inputs Ru et al. [2020], diffusion-kernel constructions for hybrid spaces Deshwal et al. [2021], and mixed-variable Gaussian processes with latent or hierarchical structures Garrido-Merchán and Hernández-Lobato [2020], Pelamatti et al. [2019], Saves et al. [2024]. In combinatorial surrogate modelling, distance-based models for permutations and indefinite-kernel treatments have been proposed Zaefferer et al. [2014], Zaefferer and Bartz-Beielstein [2020]. From a landscape perspective, the role of adjacency and metric choice in shaping landscape features is classical Stadler [2002], Ochoa et al. [2011], Tomassini et al. [2008], and permutation metrics have been systematically studied in the context of search landscape analysis Schiavinotto and Stützle [2007], Irurozki et al. [2019]. A common thread in all these approaches is that categorical components are handled either through fixed encodings (integer, one-hot, binary) or through per-type operators (random swap, exhaustive enumeration), while the mixed search space is assembled informally. We note that mixed-distance constructions have a long history—Gower's general similarity coefficient Gower [1971] combines heterogeneous variable types, and MADS-based frameworks Audet and Dennis [2006], Audet and Le Digabel [2024] define poll neighbourhoods on mixed domains. However, these approaches use *fixed* distance or neighbourhood definitions on the categorical component (typically the discrete metric or exhaustive enumeration); in particular, they do not address minimal coverage bounds or uniform metric equivalence under dynamic neighbourhood rotation. The recent CatMADS framework Audet et al. [2025] extends MADS to mixed-variable constrained blackbox problems via distance-induced categorical neighbourhoods. In CatMADS, a categorical distance is constructed—either user-defined or learned from an initial design of experiments—and neighbourhoods consist of the m closest categorical components with respect to this fixed distance. The framework introduces several types of local minima characterised by interactions between variable types and establishes convergence results including Clarke stationarity for the continuous variables. In contrast, the present work does not assume a fixed or data-dependent categorical metric. Instead, categorical variables are endowed with transient cyclic metrics that rotate over time, with no privileged ordering. The resulting hierarchy of local optimality is defined by the degree of geometric coverage, and we establish explicit coverage bounds together with uniform metric equivalence across rotations. The two frameworks are complementary: CatMADS provides algorithmic convergence guarantees within the MADS paradigm, while the present work provides a structural metric foundation for mixed spaces under dynamic categorical geometry. To the best of our knowledge, no existing approach formalises robustness of categorical local optimality under systematically *varying* neighbourhood structures within a unified metric-product setting, nor endows categorical variables with a *dynamic* metric that rotates over time with explicit coverage bounds, convergence guarantees, and topological invariance within a product metric space. The present work provides a complementary structural perspective by introducing transient operational metrics on categorical variables and developing the associated theory. In particular, the contribution is not a new mixed distance, but the formalisation of a *rotating* categorical geometry together with coverage and convergence guarantees.

The framework is illustrated through the Deterministic Neighbourhood Rotation (DNR) mechanism Madeira [2026] for derivative-free optimisation; see Audet and Hare [2017], Conn et al. [2009], Audet and Dennis [2006] for general references. DNR serves as a concrete instantiation of transient geometry. However, the theory is not specific to DFO and applies to any method operating on mixed search spaces—evolutionary, hybrid, surrogate-based, or otherwise.

The remainder of the paper is organised as follows. Section 2 reviews the geometric properties of variable types. Section 3 examines whether categorical variables are genuinely structureless. Section 4 develops the formal framework, including the product metric, optimality hierarchy, coverage theorem, and convergence result. Section 5 establishes continuity and topological invariance under ordering rotation. Section 6 compares the framework with classical encodings. Section 7 provides an illustrative example. Section 8 discusses broader implications and connections. Section 9 lists open questions, and Section 10 concludes.

2 Variable Types and Their Geometric Properties

We briefly review the geometric status of the principal variable types encountered in mixed-variable optimisation.

Continuous variables. A continuous variable takes values in an interval $[a, b] \subset \mathbb{R}$ (or more generally in $\mathbb{R}^{n_c}$). The Euclidean metric provides a canonical distance, and the topology is the standard one. Neighbourhoods are balls of radius ε, local optimality is defined via these balls, and gradient-based or direct-search methods exploit the smoothness (or at least continuity) of the objective with respect to this metric.

Integer variables. An integer variable takes values in a subset of $\mathbb{Z}$. The natural metric is the restriction of the absolute-value metric $d(i, j) = |i - j|$, and the natural neighbourhood consists of ± 1 moves. Integer variables inherit an ordering and a notion of "between," so interpolation and rounding are meaningful operations.

Discrete numerical variables. Some variables take values in a finite ordered set (e.g., $\{0.1, 0.5, 1.0, 2.0\}$), inheriting a metric from the real line. These are geometrically similar to integer variables: they possess an ordering, a distance, and a neighbourhood structure.

Categorical variables. A categorical variable takes values in a finite unordered set $C = \{c_1, \ldots, c_k\}$. There is no intrinsic ordering, no notion of "between," and no canonical distance beyond the trivial discrete metric ($d(c_i, c_j) = 0$ if $i = j$, 1 otherwise). Under this modelling convention, all distinct categories are treated as equidistant. This representation is commonly adopted—explicitly or implicitly—in mixed-variable optimisation frameworks and surrogate modelling, often via fixed integer or one-hot encodings Hutter et al. [2011], Ru et al. [2020], Garrido-Merchán and Hernández-Lobato [2020]. While effective in practice, it provides no notion of graded locality within the categorical component.

The asymmetry. The first three variable types share a common geometric pattern: a metric space with a meaningful distance that induces neighbourhoods, local optimality concepts, and search operators. Categorical variables, under the trivial metric or fixed encodings, lack an intrinsic neighbourhood structure beyond global enumeration. This asymmetry forces mixed-variable methods to treat categorical components differently—typically through random mutations, exhaustive enumeration, or representation-based heuristics Hutter et al. [2011], Ru et al. [2020], Pelamatti et al. [2019]—breaking the geometric unity of the search space.

Table 1 summarises the conventional geometric status of each variable type.

Table 1: Geometric properties of variable types (conventional view).

	Ordering	Distance	Neighbourhood	Locality
Continuous	Yes	Euclidean	ε-ball	Natural
Integer	Yes	$\|i - j\|$	± 1	Natural
Discrete num.	Yes	Inherited	Adjacent levels	Natural
Categorical	No	Trivial	Trivial ($\{x\}$ or C)	None

3 Are Categorical Variables Really Unstructured?

The classification in Table 1 reflects the *conventional* treatment of categorical variables. We now question whether this treatment is a necessary consequence of their nature, or merely a choice.

The practical observation. In applications, categories are seldom arbitrary labels. A material variable with levels {aluminium, steel, titanium, carbon fibre} encodes real physical differences: some materials are closer in weight, others in stiffness, others in cost. A solver-configuration variable with levels {direct, iterative-CG, iterative-GMRES, multigrid} encodes algorithmic relationships: the two iterative methods share more structure with each other than with the direct solver. These relationships are not captured by the trivial discrete metric. Thus, the apparent structurelessness is not necessarily inherent in the categories themselves, but may result from the modelling choice.

The separation principle. We distinguish two levels of description:

- **Ontological level:** Categorical variables have no *fixed*, *intrinsic* order. This is a property of the variable itself and does not change.
- **Operational level:** At each phase of the algorithm, a temporary geometry can be imposed on the categories—for instance, via a cyclic ordering—providing a working notion of distance and neighbourhood.

The key principle is:

> *The operational structure of the search space need not coincide with the ontological nature of the variable.*

Formally, the *ontological* structure of a variable refers to properties intrinsic to its type—for a categorical variable, the absence of any canonical total order on C. The *operational* structure refers to the choice of metric used to define neighbourhoods during optimisation; this choice need not reflect the ontological type. Categories remain ontologically unordered. But operationally, they can admit a transient, rotating geometry that provides structure without privileging any single ordering.

What this is not. This approach does *not* convert categories into ordinal variables. An ordinal encoding imposes a fixed order that is never revisited. Transient geometry imposes a temporary order that rotates over time, ensuring that no ordering is privileged and that the categorical nature is preserved in the long run.

In the next section, we formalise this idea by endowing categorical sets with cyclic-ordering-induced metrics and embedding them into a weighted product metric for the mixed space.

4 Formal Framework

We now develop the mathematical foundations. We begin with standard definitions, then introduce transient geometry for categorical spaces, the weighted product metric for mixed spaces, and the hierarchy of local optimality.

Notation. We use π for elements of the symmetric group (permutations) and σ for *cyclic orderings*—equivalence classes of permutations under circular rotation. Since only the cyclic structure matters for our neighbourhood definition (there is no distinguished "first" category), the operational object throughout is σ, not π.

4.1 Preliminaries

Definition 4.1 (Metric space)**.** A *metric space* is a pair (X, d) where X is a non-empty set and $d\colon X \times X \to [0, \infty)$ satisfies, for all $x, y, z \in X$:

(i) $d(x, y) = 0$ if and only if $x = y$ (identity of indiscernibles);
(ii) $d(x, y) = d(y, x)$ (symmetry);
(iii) $d(x, z) \le d(x, y) + d(y, z)$ (triangle inequality).

Definition 4.2 (Component spaces)**.** We consider three types of component spaces:

(a) **Continuous:** $(\mathbb{R}^{n_c}, d_{\text{cont}})$ with $d_{\text{cont}}(x, y) = \|x - y\|_2$.
(b) **Integer:** $(\mathbb{Z}^{n_i}, d_{\text{int}})$ with $d_{\text{int}}(x, y) = \|x - y\|_1$.
(c) **Categorical:** For each categorical variable ℓ, a finite set $C_\ell = \{c_1, \ldots, c_{k_\ell}\}$ with $k_\ell \ge 2$. The *discrete (trivial) metric* is

$$d_0(c_i, c_j) = \begin{cases} 0, & i = j, \\ 1, & i \neq j. \end{cases}$$

Definition 4.3 (Symmetric group and cyclic orderings)**.** The *symmetric group* S_k is the set of all bijections $\pi\colon \{1, \ldots, k\} \to \{1, \ldots, k\}$, with $|S_k| = k!$. A *cyclic ordering* of $C = \{c_1, \ldots, c_k\}$ is an equivalence class of permutations under circular rotation. Two permutations π, π' define the same cyclic ordering if π' is a cyclic shift of π. The number of distinct cyclic orderings is $(k-1)!/2$ (for unoriented cycles) or $(k-1)!$ (for oriented cycles); we use oriented cycles throughout. We denote a cyclic ordering by σ and write the induced sequence as $(c_{\sigma(1)}, c_{\sigma(2)}, \ldots, c_{\sigma(k)})$, understood cyclically. We denote by $\mathcal{C}_k$ the set of all oriented cyclic orderings on a k-element set; its cardinality is $(k-1)!$.

Example 4.4 (Cyclic orderings for $k = 4$)**.** Let $C = \{1, 2, 3, 4\}$. The symmetric group S_4 has $4! = 24$ permutations. Consider the permutation

$$\pi = (1, 2, 3, 4).$$

All cyclic shifts of π,

$$(1, 2, 3, 4), \quad (2, 3, 4, 1), \quad (3, 4, 1, 2), \quad (4, 1, 2, 3),$$

define the same *oriented cyclic ordering*. They belong to the same equivalence class under circular rotation.

Hence one cyclic ordering corresponds to 4 permutations. Since $|S_4| = 24$, the number of distinct oriented cyclic orderings is

$$\frac{4!}{4} = (4 - 1)! = 6.$$

4.2 Transient Geometry on Categorical Spaces

Definition 4.5 (Cyclic-adjacency neighbourhood)**.** Let $C = \{c_1, \dots, c_k\}$ be a categorical set with $k \geq 3$, and let σ be a cyclic ordering on C, inducing the cyclic sequence $(c_{\sigma(1)}, c_{\sigma(2)}, \dots, c_{\sigma(k)})$ with $c_{\sigma(k+1)} \equiv c_{\sigma(1)}$ and $c_{\sigma(0)} \equiv c_{\sigma(k)}$.

The *cyclic-adjacency neighbourhood* of category $c_{\sigma(j)}$ under σ is

$$\mathcal{N}_\sigma(c_{\sigma(j)}) \;=\; \big\{c_{\sigma(j-1)},\; c_{\sigma(j+1)}\big\}, \qquad j \in \{1, \dots, k\},$$

where $j \pm 1$ is interpreted cyclically in $\{1, \dots, k\}$. Every category has exactly two neighbours under every cyclic ordering.

Remark 4.6 (Design rationale)**.** The adoption of cyclic adjacency is not arbitrary. Among all simple graphs on a k-element set, the cycle graph Cyc_k is the minimal non-trivial regular structure: it has constant degree 2, introduces no distinguished vertices, and preserves complete symmetry of categories. Any sparser connected graph would either disconnect the space or break the uniform degree property; any denser regular graph (e.g., the complete graph K_k) would increase the degree to $k-1$, effectively collapsing the notion of minimal perturbation into near-enumeration. The cycle therefore represents the smallest symmetric combinatorial geometry that induces a meaningful locality structure while keeping neighbourhood size constant. Indeed, for $k \geq 3$ the cycle is the unique connected 2-regular graph on k vertices.

This choice is analogous to the ± 1 neighbourhood for integer variables and the ε-ball for continuous variables. The cyclic convention (rather than linear) avoids boundary effects and ensures that every category has exactly two neighbours under every ordering, regardless of its position—a *structural neutrality* that introduces no bias toward any particular category.

This minimality underlies both the coverage result (Proposition 4.31) and the linear bound $\lceil (k-1)/2 \rceil$ on the number of epochs required for full adjacency validation.

We emphasise that the framework itself is not restricted to cyclic structures. Richer transient geometries—such as weighted complete graphs or permutation metrics including Kendall or Cayley distance—can be incorporated within the same product-metric formalism. A generalisation to radius $r > 1$ (yielding $2r$ neighbours per category) is likewise straightforward. We adopt the cycle with $r = 1$ as a canonical minimal instance, balancing symmetry, locality, and constructive coverage; the convergence results (Sections 4.6–4.8) hold for any fixed r.

Example 4.7. Let $C = \{1, 2, 3, 4, 5\}$ and consider the cyclic ordering $\sigma = (1, 2, 3, 4, 5)$. The cyclic-adjacency neighbourhoods are:

$$\mathcal{N}_\sigma(1) = \{2, 5\}, \qquad \mathcal{N}_\sigma(2) = \{1, 3\}, \qquad \mathcal{N}_\sigma(3) = \{2, 4\}.$$

Note that the cyclic convention avoids boundary effects: category 1 has neighbours $\{2, 5\}$, not only $\{2\}$.

Remark 4.8 (The case $k = 2$)**.** For $k_\ell = 2$, the cyclic metric coincides with the trivial discrete metric (distance 0 or 1) and every category is the unique neighbour of the other. The transient geometry machinery is therefore vacuous for binary categorical variables; the framework becomes substantive for $k \geq 3$.

Definition 4.9 (Cyclic-ordering-induced metric on C)**.** Given a cyclic ordering σ on $C = \{c_1, \ldots, c_k\}$, define the *cyclic distance*

$$d_{\text{cat}}^{(\sigma)}(c_i, c_j) \;=\; \min\bigl(|\sigma^{-1}(i) - \sigma^{-1}(j)|,\; k - |\sigma^{-1}(i) - \sigma^{-1}(j)|\bigr),$$

where $\sigma^{-1}(i)$ is the position of c_i in the sequence induced by σ. This is the shortest arc length on the cycle of k elements, taking values in $\{0, 1, \ldots, \lfloor k/2 \rfloor\}$. We identify σ with the bijection $\sigma : \{1, \ldots, k\} \to \{1, \ldots, k\}$ mapping positions to category indices, so that $\sigma^{-1}(i)$ denotes the position of category c_i in the ordering.

Proposition 4.10 ($d_{\text{cat}}^{(\sigma)}$ is a metric)**.** *For every cyclic ordering σ, the function $d_{\text{cat}}^{(\sigma)}$ is a metric on C.*

Proof. The cyclic distance is the shortest-path metric on the cycle graph Cyc_k with vertices $\{1, \ldots, k\}$ and edges $\bigl\{\{j, j+1 \bmod k\} : j = 1, \ldots, k\bigr\}$, relabelled via σ. The cycle Cyc_k ($k \geq 3$) is connected (each vertex has degree 2 and the graph forms a single cycle), so shortest-path distances between all pairs of vertices are finite. Shortest-path distances on any connected graph form a metric: identity of indiscernibles and symmetry are immediate; the triangle inequality follows because the shortest path from a to c cannot exceed the sum of shortest paths $a \to b$ and $b \to c$. $\square$

Remark 4.11 (Well-definedness of the cyclic distance)**.** Although a cyclic ordering σ is formally an equivalence class of permutations under circular rotation, the distance $d_{\text{cat}}^{(\sigma)}(c_i, c_j)$ is well defined on this class. Indeed, if σ' is a cyclic shift of σ, then $\sigma'^{-1}(i) = \sigma^{-1}(i) + r \pmod{k}$ for some fixed shift $r \in \{0, \ldots, k-1\}$, so $|\sigma'^{-1}(i) - \sigma'^{-1}(j)| = |\sigma^{-1}(i) - \sigma^{-1}(j)|$ modulo k, and the arc-minimum is unchanged. Hence $d_{\text{cat}}^{(\sigma)}$ depends only on the cyclic ordering, not on the representative permutation chosen.

Example 4.12. Let $C = \{1, 2, 3, 4, 5\}$ with cyclic ordering $\sigma = (1, 2, 3, 4, 5)$. The cyclic distances from category 1 are:

$$d_{\text{cat}}^{(\sigma)}(1,2) = 1, \qquad d_{\text{cat}}^{(\sigma)}(1,5) = 1,$$
$$d_{\text{cat}}^{(\sigma)}(1,3) = 2, \qquad d_{\text{cat}}^{(\sigma)}(1,4) = 2.$$

The maximum possible distance is $\lfloor k/2 \rfloor = 2$, corresponding to opposite points on the cycle.

Definition 4.13 (Ordering schedule and epochs)**.** An *ordering schedule* is a sequence $\mathcal{S} = (\sigma_1, \sigma_2, \sigma_3, \ldots)$ of cyclic orderings on C.

Each integer $t \geq 1$ defines an *epoch*, during which the categorical neighbourhood structure is determined by the ordering σ_t. At epoch t, the categorical metric is $d_{\text{cat}}^{(\sigma_t)}$.

We denote by $\Sigma_t = \{\sigma_1, \ldots, \sigma_t\}$ the set of *visited orderings* up to epoch t.

Remark 4.14 (On the schedule)**.** No restriction is imposed on the choice of schedule. It may be deterministic (e.g., a low-discrepancy sequence of orderings as in Madeira [2026]), random, or adaptive. The theoretical results in Sections 4.3–4.6 hold for *any* schedule. For multiple categorical variables, we use *ordering vectors* $\boldsymbol{\sigma}_t = (\sigma_t^{(1)}, \ldots, \sigma_t^{(n_{\text{cat}})})$, one cyclic ordering per categorical component (Definition 4.16).

4.3 The Weighted Product Metric

Definition 4.15 (Mixed search space)**.** Let $\mathcal{X}_{\text{cont}} = \mathbb{R}^{n_c}$, $\mathcal{X}_{\text{int}} = \mathbb{Z}^{n_i}$, and $\mathcal{X}_{\text{cat}} = C_1 \times \cdots \times C_{n_{\text{cat}}}$ where each C_ℓ is a finite categorical set with $k_\ell \geq 2$ levels. The *mixed search space* is

$$\mathcal{X} \;=\; \mathcal{X}_{\text{cont}} \times \mathcal{X}_{\text{int}} \times \mathcal{X}_{\text{cat}}.$$

A point $x \in \mathcal{X}$ is written $x = (x_{\text{c}}, x_{\text{i}}, x_{\text{cat}})$ with $x_{\text{cat}} = (x_{\text{cat},1}, \ldots, x_{\text{cat},n_{\text{cat}}})$.

Definition 4.16 (Aggregated categorical metric)**.** Let $p \in [1, \infty)$ be fixed (the same value as in Definition 4.17 below). Given cyclic orderings $\sigma^{(\ell)}$ on each categorical variable C_ℓ for $\ell = 1, \ldots, n_{\text{cat}}$, define

$$d_{\text{cat}}^{(\boldsymbol{\sigma})}(x_{\text{cat}}, y_{\text{cat}}) \;=\; \left(\sum_{\ell=1}^{n_{\text{cat}}} \big(d_{\text{cat}}^{(\sigma^{(\ell)})}(x_{\text{cat},\ell},\, y_{\text{cat},\ell}) \big)^p \right)^{1/p},$$

where $\boldsymbol{\sigma} = (\sigma^{(1)}, \ldots, \sigma^{(n_{\text{cat}})})$ and $p \in [1, \infty)$. Each categorical variable has its own cyclic ordering, allowing *independent rotation*.

Definition 4.17 (Weighted product metric)**.** Let $w_c, w_i, w_{\text{cat}} > 0$ be positive weights and $p \in [1, \infty)$. At epoch t with ordering vector $\boldsymbol{\sigma}_t$, the *weighted product metric* on $\mathcal{X}$ is

$$d_t(x, y) \;=\; \Big(w_c \, d_{\text{cont}}(x_{\text{c}}, y_{\text{c}})^p + w_i \, d_{\text{int}}(x_{\text{i}}, y_{\text{i}})^p + w_{\text{cat}} \, \big(d_{\text{cat}}^{(\boldsymbol{\sigma}_t)}(x_{\text{cat}}, y_{\text{cat}}) \big)^p \Big)^{1/p}. \tag{1}$$

Theorem 4.18 (Product metric validity)**.** *If* $(\mathcal{X}_{\text{cont}}, d_{\text{cont}})$*,* $(\mathcal{X}_{\text{int}}, d_{\text{int}})$*, and* $(\mathcal{X}_{\text{cat}}, d_{\text{cat}}^{(\boldsymbol{\sigma})})$ *are metric spaces and* $w_c, w_i, w_{\text{cat}} > 0$*, then* $(\mathcal{X}, d)$ *with* d *defined by* (1) *is a metric space for every* $p \in [1, \infty)$*.*

Proof. Let $x, y, z \in \mathcal{X}$ and write $\mathbf{u} = (w_c^{1/p} d_{\text{cont}}(x_{\text{c}}, y_{\text{c}}),\; w_i^{1/p} d_{\text{int}}(x_{\text{i}}, y_{\text{i}}),\; w_{\text{cat}}^{1/p} \, d_{\text{cat}}^{(\boldsymbol{\sigma})}(x_{\text{cat}}, y_{\text{cat}}))$ so that $d(x, y) = \|\mathbf{u}\|_p$.

Identity of indiscernibles. $d(x, y) = 0$ iff $\|\mathbf{u}\|_p = 0$ iff each component is zero. Since weights are positive, this holds iff $d_{\text{cont}} = d_{\text{int}} = d_{\text{cat}}^{(\boldsymbol{\sigma})} = 0$, i.e., $x = y$.

Symmetry. Each component metric is symmetric, hence $\mathbf{u}$ is symmetric in (x, y).

Triangle inequality. Define $\mathbf{v}$ and $\mathbf{w}$ analogously for the pairs (x, z) and (z, y). The component-wise triangle inequality gives $u_j \leq v_j + w_j$ for each j. Since $u_j, v_j, w_j \geq 0$, monotonicity of the ℓ^p norm on the non-negative orthant gives $\|\mathbf{u}\|_p \leq \|\mathbf{v} + \mathbf{w}\|_p$. Applying the Minkowski inequality in $\mathbb{R}^3$ endowed with the ℓ^p norm then yields

$$\|\mathbf{u}\|_p \;\leq\; \|\mathbf{v} + \mathbf{w}\|_p \;\leq\; \|\mathbf{v}\|_p + \|\mathbf{w}\|_p,$$

which yields $d(x, y) \leq d(x, z) + d(z, y)$. □

Remark 4.19 (On the product construction)**.** The weighted ℓ^p product of metric spaces is a classical construction in metric geometry. The novelty here is not the product itself but rather *what is being combined*: a dynamic categorical component $(\mathcal{X}_{\text{cat}}, d_{\text{cat}}^{(\boldsymbol{\sigma}_t)})$ whose metric changes at each epoch. Theorem 4.18 ensures that the product remains a valid metric space under every such change; the non-trivial consequences—topological invariance (Theorem 5.3), coverage (Theorem 4.28), and algorithmic robustification (Section 4.8)—arise from the interplay between the dynamic component and the static ones. We note that the choice of ℓ^p aggregation is not essential for the topological results: any equivalent product norm (e.g., a different value of p, or a weighted ℓ^∞ product) yields the same topology on $\mathcal{X}$, by standard norm equivalence in finite-dimensional product spaces, and hence the same convergence and coverage properties. The specific value of p affects only the quantitative geometry—neighbourhood shapes and metric constants—not the qualitative structure of the framework.

Proposition 4.20 (Dynamic metric validity)**.** *Let $\mathcal{S} = (\boldsymbol{\sigma}_1, \boldsymbol{\sigma}_2, \dots)$ be any ordering schedule. Then $(\mathcal{X}, d_t)$ is a metric space for every epoch t.*

Proof. Immediate from Theorem 4.18: at each epoch t, the component $d_{\mathrm{cat}}^{(\boldsymbol{\sigma}_t)}$ is a metric by Proposition 4.10, so the product is a metric. □

4.4 Neighbourhoods in the Product Space

Definition 4.21 (Product neighbourhood)**.** Given a point $x \in \mathcal{X}$ and epoch t, the ε-*neighbourhood* is $B_t(x, \varepsilon) = \{y \in \mathcal{X} : d_t(x, y) \leq \varepsilon\}$.

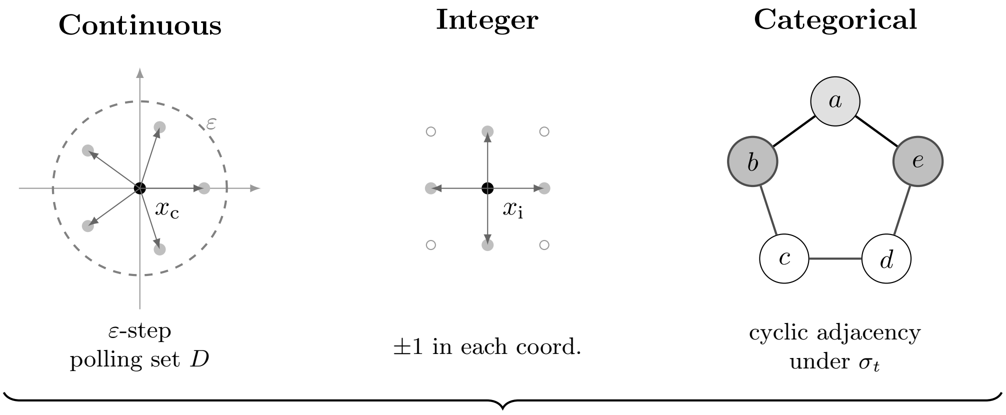
For algorithmic purposes, we also define the *elementary neighbourhood* as the set of points differing from x by a minimal move in *exactly one* component:

$$\mathcal{N}_t(x) \;=\; \mathcal{N}_{\mathrm{cont}}(x) \;\cup\; \mathcal{N}_{\mathrm{int}}(x) \;\cup\; \mathcal{N}_{\mathrm{cat}}^{(\boldsymbol{\sigma}_t)}(x),$$

where:

- $\mathcal{N}_{\mathrm{cont}}(x)$: points of the form $x_{\mathrm{c}} + \delta d$ (with all other coordinates unchanged), where $D \subset \mathbb{R}^{n_c}$ is a *fixed* finite set of unit polling directions (e.g., coordinate directions $\{\pm e_1, \dots, \pm e_{n_c}\}$, or a positive spanning set) and $\delta > 0$ is a *fixed* step-size parameter. Both D and δ remain constant across epochs; in particular, the convergence results in Section 4.8 concern optimality at this fixed resolution, not asymptotic stationarity as $\delta \to 0$. Extending the analysis to vanishing step sizes would require Clarke-type arguments and lies beyond the scope of the present framework;
- $\mathcal{N}_{\mathrm{int}}(x)$: points differing from x in one integer coordinate by ± 1;
- $\mathcal{N}_{\mathrm{cat}}^{(\boldsymbol{\sigma}_t)}(x)$: points differing from x in one categorical variable, to one of its two cyclic-adjacent neighbours under the current ordering $\sigma_t^{(\ell)}$.

Remark 4.22 (Uniform minimal-perturbation logic)**.** In each component, the neighbourhood is defined by the *smallest non-trivial perturbation*: ε-step in continuous, ± 1 in integer, and cyclic adjacency in categorical. This uniform logic across variable types is a direct consequence of the product metric structure (Fig. 1).



$$\mathcal{N}_t(x) \;=\; \mathcal{N}_{\mathrm{cont}}(x) \;\cup\; \mathcal{N}_{\mathrm{int}}(x) \;\cup\; \mathcal{N}_{\mathrm{cat}}^{(\boldsymbol{\sigma}_t)}(x)$$

Figure 1: Elementary neighbourhood in the product space. In each component, the neighbourhood is defined by the smallest non-trivial perturbation: ε-step along polling directions in the continuous component, ± 1 moves in the integer lattice, and cyclic adjacency under the current ordering σ_t in the categorical component. Dark nodes: current point; grey nodes: neighbours.

4.5 Hierarchy of Local Optimality

Let $f\colon \mathcal{X} \to \mathbb{R}$ be the objective function (to be minimised).

Definition 4.23 (Instantaneous local optimum)**.** A point $x^* \in \mathcal{X}$ is an *instantaneous local optimum* under ordering vector $\boldsymbol{\sigma}$ if

$$f(x^*) \;\le\; f(y) \qquad \forall\, y \in \mathcal{N}_{\boldsymbol{\sigma}}(x^*).$$

This is the weakest level: it depends on the current cyclic ordering.

Definition 4.24 (Robust local optimum)**.** Given visited ordering vectors $\Sigma_t = \{\boldsymbol{\sigma}_1, \dots, \boldsymbol{\sigma}_t\}$, a point x^* is a *robust local optimum of order* t if

$$f(x^*) \;\le\; f(y) \qquad \forall\, y \in U_t(x^*) \;\coloneqq\; \bigcup_{\boldsymbol{\sigma} \in \Sigma_t} \mathcal{N}_{\boldsymbol{\sigma}}(x^*).$$

This depends on the process history: optimality is validated against all visited geometries.

Definition 4.25 (Categorical local optimum)**.** A point x^* is a *categorical local optimum* if

$$f(x^*) \;\le\; f(y) \qquad \forall\, y \in U_\infty(x^*) \;\coloneqq\; \bigcup_{\boldsymbol{\sigma}} \mathcal{N}_{\boldsymbol{\sigma}}(x^*),$$

where the union is over *all* possible ordering vectors $\boldsymbol{\sigma} \in \prod_{\ell=1}^{n_{\mathrm{cat}}} \mathcal{C}_{k_\ell}$ (Definition 4.3). This is the strongest level: optimality holds under *every possible* cyclic ordering. It is independent of any particular choice of geometry on the categorical components. Since each categorical component C_ℓ admits finitely many cyclic orderings ($|\mathcal{C}_{k_\ell}| = (k_\ell - 1)!$), the product $\prod_\ell \mathcal{C}_{k_\ell}$ is finite and $U_\infty(x^*)$ is a finite union of finite sets.

Proposition 4.26 (Monotonicity of robust optimality)**.** *Let x^* be a robust local optimum of order t. Then:*

(i) x^* is a robust local optimum of every order $s < t$.

(ii) If x^* is a robust local optimum of order t but not *of order $t+1$, then there exists $y \in \mathcal{N}_{\boldsymbol{\sigma}_{t+1}}(x^*) \setminus U_t(x^*)$ with $f(y) < f(x^*)$.*

Proof. *(i)* $\Sigma_s \subseteq \Sigma_t$ implies $U_s(x^*) \subseteq U_t(x^*)$. Hence $f(x^*) \le f(y)$ for all $y \in U_t(x^*)$ implies the same for all $y \in U_s(x^*)$.

(ii) If x^* is not robust of order $t+1$, there exists $y \in U_{t+1}(x^*)$ with $f(y) < f(x^*)$. Since $f(x^*) \le f(y)$ for all $y \in U_t(x^*)$, such y must lie in $U_{t+1}(x^*) \setminus U_t(x^*) \subseteq \mathcal{N}_{\boldsymbol{\sigma}_{t+1}}(x^*) \setminus U_t(x^*)$. □

Corollary 4.27 (Robustness as monotone validation)**.** *As t increases, the set of robust local optima of order t can only* shrink *or remain the same. Points that survive validation against more geometries have strictly stronger optimality guarantees. The rotation of orderings thus acts as a* progressive robustness filter*: it does not merely explore—it validates (Fig. 2).*

4.6 Coverage and the Categorical Level

The following results establish that the categorical level of optimality is well defined and reachable, and quantify the rate at which robust optimality approaches it.

Theorem 4.28 (Full coverage by cyclic adjacency)**.** *Let $C = \{c_1, \dots, c_k\}$ with $k \ge 3$ and let $\mathcal{N}_\sigma$ denote the cyclic-adjacency neighbourhood (Definition 4.5). Then for any $a \in C$,*

$$\bigcup_{\sigma} \mathcal{N}_\sigma(a) \;=\; C \setminus \{a\},$$

where the union is over all cyclic orderings of C.

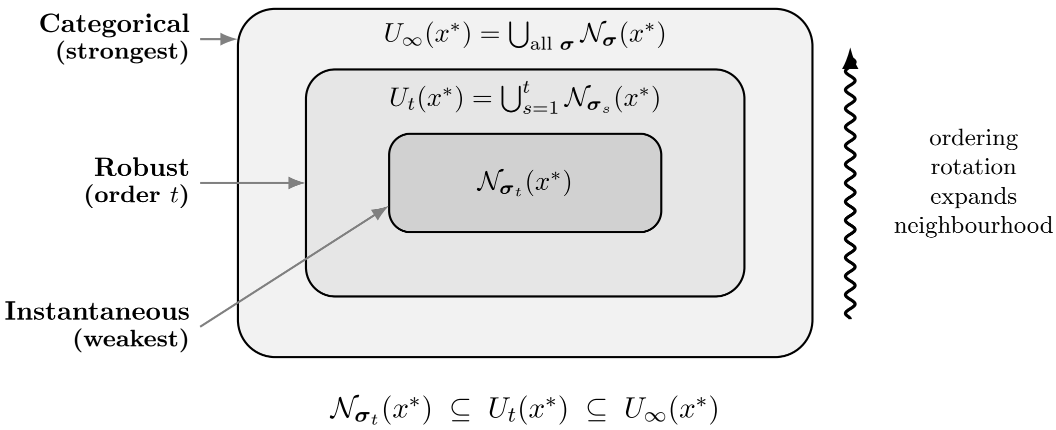


Figure 2: Hierarchy of local optimality as progressive neighbourhood expansion. Each ordering rotation enlarges the neighbourhood against which a candidate is validated. A point that remains locally optimal as the union grows from instantaneous ($\mathcal{N}_{\boldsymbol{\sigma}_t}$) through robust ($U_t$) to categorical ($U_\infty$) has been tested against increasingly diverse geometries. (The inclusion refers to neighbourhood sets; the corresponding sets of local optima shrink as the neighbourhood expands.)

Proof. Let $b \in C \setminus \{a\}$. Construct a cyclic ordering σ that places a and b in adjacent positions: set $\sigma(1)$ to the index of a and $\sigma(2)$ to the index of b, then fill the remaining $k-2$ positions arbitrarily. Under this σ, the category b is a cyclic-adjacent neighbour of a, so $b \in \mathcal{N}_\sigma(a)$. Since b was arbitrary, the union covers $C \setminus \{a\}$. The reverse inclusion $\mathcal{N}_\sigma(a) \subseteq C \setminus \{a\}$ is immediate. Equivalently, this follows from the classical fact that any prescribed edge of K_k can be extended to a Hamiltonian cycle, hence realised as an adjacency under some cyclic ordering. $\square$

Corollary 4.29 (Meaning of categorical optimality)**.** *If x^* is a categorical local optimum (Definition 4.25), then $f(x^*) \leq f(y)$ for every $y \in \mathcal{X}$ that differs from x^* by a minimal move in exactly one component, namely:*

- *in one categorical component (to any other level),*
- *or in one integer coordinate (by ± 1),*
- *or in the continuous component by one polling move (as defined in Definition 4.21).*

In other words, the categorical level implies optimality against all single-component elementary moves *in the mixed space, without reference to a specific cyclic ordering.*

Proof. By definition, categorical local optimality means $f(x^*) \leq f(y)$ for all $y \in U_\infty(x^*)$, where $U_\infty(x^*)$ is the union of elementary neighbourhoods over all ordering vectors. By Theorem 4.28, the union of cyclic-adjacency neighbourhoods over all cyclic orderings yields all categorical levels except the current one. The integer and continuous elementary move sets are fixed across orderings by Definition 4.21. Hence $U_\infty(x^*)$ contains all single-component elementary moves listed above, and the claim follows. $\square$

Remark 4.30 (Categorical optimality and enumeration)**.** Corollary 4.29 shows that a categorical local optimum has been tested against *all* single-component elementary moves. In principle, the same test could be performed by exhaustive enumeration of all categorical levels at the current point. The contribution of the transient geometry framework is not the *destination* (testing all neighbours) but the *path*: it provides a structured, progressive mechanism—with explicit coverage bounds (Proposition 4.31)—that reaches full coverage through a sequence of small, geometrically grounded neighbourhoods, rather than through exhaustive enumeration. In particular, the degree of each categorical neighbourhood remains constant (equal to two), independent of k, whereas exhaustive enumeration requires $k-1$ evaluations per variable. This distinction becomes practically significant when the number of categorical levels is large or when multiple categorical variables are present, since the structured approach integrates naturally with descent-based search.

Proposition 4.31 (Minimum orderings for full coverage)**.** *Let $C = \{c_1, \ldots, c_k\}$ with $k \geq 3$. The minimum number of cyclic orderings needed so that every pair (a, b) with $a \neq b$ appears as cyclic-adjacent neighbours under at least one ordering is*

$$\left\lceil \frac{k-1}{2} \right\rceil .$$

Proof. Each cyclic ordering induces a Hamiltonian cycle on k vertices and therefore contributes exactly k distinct unordered adjacency pairs. Since K_k has $\binom{k}{2} = k(k-1)/2$ edges, any family of cyclic orderings that covers all pairs must satisfy $M \cdot k \geq \binom{k}{2}$, hence $M \geq \lceil (k-1)/2 \rceil$.

To show achievability, we appeal to a classical result on Hamilton decompositions of complete graphs (Walecki construction; see [Alspach et al., 1990, §3, Theorem 1]). The complete graph K_k admits a Hamilton decomposition for every k, in the sense that when $\deg(K_k)$ is odd the decomposition consists of Hamiltonian cycles together with one perfect matching.

If k is odd, then $\deg(K_k) = k - 1$ is even and the decomposition yields exactly $(k-1)/2$ edge-disjoint Hamiltonian cycles partitioning $E(K_k)$. Each such cycle corresponds to a cyclic ordering whose adjacency pairs cover k distinct edges of K_k. Hence $(k-1)/2 = \lceil (k-1)/2 \rceil$ cyclic orderings suffice for full coverage.

If k is even, then $\deg(K_k) = k - 1$ is odd and the decomposition consists of $(k-2)/2$ edge-disjoint Hamiltonian cycles together with one perfect matching F. These cycles cover all edges of K_k except those in F. Since K_k is complete, any perfect matching

$$F = \{(u_i, v_i)\}_{i=1}^{k/2}$$

is contained in the Hamiltonian cycle

$$(u_1, v_1, u_2, v_2, \ldots, u_{k/2}, v_{k/2}),$$

obtained by interleaving the matching edges with connecting edges of K_k. Because K_k is complete, every consecutive pair in this sequence is joined by an edge, so the displayed sequence indeed defines a Hamiltonian cycle containing all edges of F. Adding this cycle produces a total of

$$(k-2)/2 + 1 = k/2 = \lceil (k-1)/2 \rceil$$

cyclic orderings whose adjacency pairs cover all $\binom{k}{2}$ edges of K_k. Unlike the odd case, the final Hamiltonian cycle may share edges with the preceding $(k-2)/2$ cycles; this is not an obstacle, since for coverage purposes it suffices that it contains all edges of F. The family therefore covers every edge of K_k but is not necessarily edge-disjoint when k is even. □

Remark 4.32 (Convergence rate of robustness)**.** Proposition 4.31 provides a concrete bound on the number of epochs needed for robust optimality to equal categorical optimality (in each categorical component): $\lceil (k_\ell - 1)/2 \rceil$ for a variable with k_ℓ levels. For independent rotation of n_{cat} categorical variables, the bound becomes $\max_\ell \lceil (k_\ell - 1)/2 \rceil$ if the schedule is designed to cover all variables simultaneously. The cumulative neighbourhood size $|U_t(x^*)|$ is non-decreasing in t and reaches its maximum when full coverage is achieved.

Since each categorical variable C_ℓ has its own cyclic ordering $\sigma_t^{(\ell)}$ (Definition 4.16), the schedules can be designed independently: the coverage requirement for C_ℓ is that $\lceil (k_\ell - 1)/2 \rceil$ suitably chosen orderings appear among the visited ordering vectors. In a product schedule where all components rotate simultaneously, each epoch advances coverage in *every* categorical variable at once, so the overall number of epochs for full product coverage is $\max_\ell \lceil (k_\ell - 1)/2 \rceil$—determined by the variable with the most levels. Variables with fewer levels achieve coverage earlier and their subsequent orderings provide redundant (but harmless) validation.

4.7 Role of Weights in the Product Metric

Proposition 4.33 (Weight-dependent neighbourhood composition)**.** *In the ε-neighbourhood $B_t(x, \varepsilon)$ of the product metric* (1)*, the maximum allowable perturbation in component j (holding other components fixed) is $\varepsilon_j^{\max} = \varepsilon / w_j^{1/p}$. Larger weights impose tighter constraints on the corresponding component; smaller weights allow larger perturbations. The weights thus control the* relative resolution *of the neighbourhood across variable types.*

Proof. If only component j varies, $d_t(x, y) = (w_j\, d_j^p)^{1/p} = w_j^{1/p}\, d_j$. The bound follows from $d_t \leq \varepsilon$. □

Remark 4.34 (Practical interpretation)**.** The weights w_c, w_i, w_{cat} are design parameters, analogous to step sizes or trust-region radii. They can be: (i) set by the user based on domain knowledge; (ii) normalised by variable ranges or number of levels; (iii) learned adaptively during the optimisation. The framework is agnostic about the specific choice, but Proposition 4.33 makes explicit how the choice shapes the search neighbourhood.

4.8 Algorithmic Interpretation Under Recurrent Orderings

The optimality hierarchy introduced in Definitions 4.23–4.25 has a natural algorithmic interpretation under standard descent schemes.

Consider an epoch-based method that, at each epoch t, explores the elementary neighbourhood $\mathcal{N}_{\boldsymbol{\sigma}_t}(x)$ and accepts only improving moves. Assume:

(A1) f is lower semicontinuous and bounded below, with compact sublevel sets in the product topology;

(A2) each elementary neighbourhood is finite (which follows directly from Definition 4.21: D is a fixed finite set of polling directions, ± 1 moves are finite in number, and each C_ℓ is a finite set);

(A3) the ordering schedule visits a finite set of ordering vectors recurrently (each ordering recurs infinitely often).

Under these assumptions, the sequence of objective values is non-increasing and therefore convergent. Each epoch terminates at a point that satisfies local optimality with respect to the active ordering vector. Because every ordering recurs infinitely often, any accumulation point must inherit these local optimality certificates for all recurrent geometries. (Since the elementary neighbourhoods are finite and f is lower semicontinuous, local optimality inequalities $f(x^*) \leq f(y)$ for all $y \in \mathcal{N}_{\boldsymbol{\sigma}}(x^*)$ are preserved along convergent subsequences: if $x^{(t_j)} \to x^*$ and $f(x^{(t_j)}) \leq f(y)$ for every fixed y in the finite neighbourhood, lower semicontinuity of f gives $f(x^*) \leq \liminf_{j\to\infty} f(x^{(t_j)}) \leq f(y)$ in the limit.) Consequently, accumulation points satisfy robust local optimality in the sense of Definition 4.24.

If, in addition, the visited ordering vectors achieve full categorical coverage (Proposition 4.31), accumulation points are categorical local optima.

A full convergence theory integrating vanishing step sizes and Clarke-type stationarity for continuous components constitutes a natural extension of the present structural framework.

Remark 4.35 (Epoch length as a design parameter)**.** The number of function evaluations performed under a fixed ordering (the *epoch length*) is a design parameter not addressed by the present geometric framework. Short epochs increase the rate of coverage and accelerate robustness validation, while long epochs allow deeper local search within each geometry but delay validation across geometries. The choice is therefore method-dependent and algorithmic rather than geometric; in particular, DNR Madeira [2026] explicitly implements scheduled epochs and provides a concrete instantiation of this trade-off.

5 Continuity Under Ordering Rotation

A natural objection to the transient geometry framework is: "When the cyclic ordering changes, are you not jumping between different spaces?" We show that the answer is no, in a strong sense. Three mechanisms ensure continuity: neighbourhood overlap, ordering recurrence, and—most importantly—uniform metric equivalence.

5.1 Neighbourhood Overlap

When the ordering changes from σ_t to σ_{t+1}, the neighbourhood of a category may partially persist.

Proposition 5.1 (Expected neighbourhood overlap)**.** *Let σ, σ' be two cyclic orderings of $C = \{c_1, \ldots, c_k\}$ chosen independently and uniformly at random. For a fixed category $a \in C$, the expected size of the overlap $\mathcal{N}_\sigma(a) \cap \mathcal{N}_{\sigma'}(a)$ is*

$$\mathbb{E}\big[|\mathcal{N}_\sigma(a) \cap \mathcal{N}_{\sigma'}(a)|\big] \;=\; \frac{4}{k-1}.$$

For $k = 5$, this is 1; for $k = 10$, it is approximately 0.44.

Proof. Fix σ. Since σ' is drawn uniformly from $\mathcal{C}_k$, by the rotational symmetry of the uniform distribution over cyclic orderings, every element of $C \setminus \{a\}$ is equally likely to occupy each of the $k-1$ non-a positions in any representative of σ'. Hence each element of $C \setminus \{a\}$ has equal probability $2/(k-1)$ of being adjacent to a in σ' (there are two adjacent positions out of $k-1$ available). A specific neighbour of a under σ is therefore also a neighbour under σ' with probability $2/(k-1)$. By linearity of expectation over the two neighbours under σ, the expected overlap is $2 \cdot 2/(k-1) = 4/(k-1)$. □

Remark 5.2. For small k, the overlap is substantial (e.g., expected overlap of 1 for $k = 5$). For large k, overlap becomes rare, but the coverage bound of Proposition 4.31 ensures that full coverage is still achieved in $O(k)$ orderings. This calculation is illustrative rather than prescriptive; practical schedules need not be uniform.

5.2 Ordering Recurrence

The set of cyclic orderings on k elements is finite: $(k-1)!$ for oriented cycles. Any schedule that cycles through a predetermined list of orderings revisits each geometry periodically. This ensures that:

- No geometry is used only once and then "lost."
- Robust optimality (Definition 4.24) is periodically re-validated.
- The algorithm's behaviour is ultimately periodic in its geometric structure.

5.3 Uniform Metric Equivalence

The strongest answer to the "jump" objection is topological: all dynamic metrics induce the *same topology* on $\mathcal{X}$.

Theorem 5.3 (Uniform equivalence of dynamic metrics)**.** *Let d_t and d_s be weighted product metrics on $\mathcal{X}$ corresponding to ordering vectors $\boldsymbol{\sigma}_t$ and $\boldsymbol{\sigma}_s$ respectively. Then there exist constants $0 < m \le 1 \le M$ depending only on the categorical cardinalities $k_1, \ldots, k_{n_{\mathrm{cat}}}$ such that*

$$m\, d_s(x, y) \;\le\; d_t(x, y) \;\le\; M\, d_s(x, y) \qquad \forall\, x, y \in \mathcal{X},\ \forall\, s, t. \tag{2}$$

Explicitly, $M = \max_\ell \lfloor k_\ell/2 \rfloor$ and $m = M^{-1}$. Note that the bound is independent of the particular ordering vectors $\boldsymbol{\sigma}_t$ and $\boldsymbol{\sigma}_s$.

Proof. For any $x, y \in \mathcal{X}$, write

$$d_t(x,y)^p \;=\; A(x,y) + B_t(x,y),$$

where

$$A(x,y) \;=\; w_c\, d_{\mathrm{cont}}(x_{\mathrm{c}}, y_{\mathrm{c}})^p + w_i\, d_{\mathrm{int}}(x_{\mathrm{i}}, y_{\mathrm{i}})^p,$$

and

$$B_t(x,y) \;=\; w_{\mathrm{cat}}\left(d_{\mathrm{cat}}^{(\boldsymbol{\sigma}_t)}(x_{\mathrm{cat}}, y_{\mathrm{cat}})\right)^p.$$

The term $A(x,y)$ is independent of the ordering. We now bound B_t/B_s via a component-wise argument. Recall from Definition 4.16 that $\left(d_{\mathrm{cat}}^{(\boldsymbol{\sigma}_t)}(x_{\mathrm{cat}}, y_{\mathrm{cat}})\right)^p = \sum_{\ell=1}^{n_{\mathrm{cat}}}\left(d_{\mathrm{cat}}^{(\sigma_t^{(\ell)})}(x_{\mathrm{cat},\ell}, y_{\mathrm{cat},\ell})\right)^p$. For each categorical component C_ℓ with k_ℓ levels, the cyclic distance between two distinct categories lies in $\{1, \ldots, \lfloor k_\ell/2 \rfloor\}$. Hence for any fixed pair of distinct categories in component ℓ, and any two orderings $\sigma_t^{(\ell)}$, $\sigma_s^{(\ell)}$,

$$d_{\mathrm{cat}}^{(\sigma_t^{(\ell)})}(x_{\mathrm{cat},\ell}, y_{\mathrm{cat},\ell}) \;\le\; \lfloor k_\ell/2 \rfloor \;\le\; M \cdot d_{\mathrm{cat}}^{(\sigma_s^{(\ell)})}(x_{\mathrm{cat},\ell}, y_{\mathrm{cat},\ell}),$$

where $M = \max_\ell \lfloor k_\ell/2 \rfloor$ and we used the fact that any non-zero cyclic distance is at least 1 (since the cyclic metric takes integer values in $\{0, \ldots, \lfloor k_\ell/2 \rfloor\}$). Note that $M \ge 1$ since $k_\ell \ge 2$ implies $\lfloor k_\ell/2 \rfloor \ge 1$; this ensures in particular that $A \le M^p A$ and hence the bound $A + M^p B_s \le M^p(A + B_s)$ holds. Raising to the p-th power and summing over ℓ,

$$\sum_\ell \left(d_{\mathrm{cat}}^{(\sigma_t^{(\ell)})}(x_{\mathrm{cat},\ell}, y_{\mathrm{cat},\ell})\right)^p \;\le\; M^p \sum_\ell \left(d_{\mathrm{cat}}^{(\sigma_s^{(\ell)})}(x_{\mathrm{cat},\ell}, y_{\mathrm{cat},\ell})\right)^p,$$

which gives $B_t(x,y) \le M^p\, B_s(x,y)$, and symmetrically $B_s(x,y) \le M^p\, B_t(x,y)$. If $x_{\mathrm{cat}} = y_{\mathrm{cat}}$, then $B_t(x,y) = B_s(x,y) = 0$ for all s, t and the bound is trivial.

Since $A(x,y) \ge 0$, we obtain

$$A + B_t \;\le\; A + M^p\, B_s \;\le\; M^p(A + B_s),$$

and symmetrically $A + B_s \le M^p(A + B_t)$. Taking p-th roots yields

$$\frac{1}{M}\, d_s(x,y) \;\le\; d_t(x,y) \;\le\; M\, d_s(x,y),$$

which establishes uniform equivalence with $M = K$ and $m = 1/M$. □

Corollary 5.4 (Topological invariance)**.** *All dynamic metrics $\{d_t\}_{t\ge1}$ induce the same topology on $\mathcal{X}$. In particular:*

(i) A sequence (x_n) converges under d_t if and only if it converges under d_s, for any s, t.

(ii) Open sets, closed sets, and compactness are independent of the ordering.

(iii) The dynamics introduced by transient geometry are geometric *(affecting distances and neighbourhoods), not* topological *(the underlying structure is invariant).*

Proof. Uniform equivalence of metrics implies they generate the same topology (a standard result in metric space theory). □

Remark 5.5 (Strength of the invariance)**.** Theorem 5.3 provides a definitive answer to concerns about "jumping between spaces." The space $\mathcal{X}$ is always the same set, with the same topology. Only the *metric geometry*—which distances are small and which are large—changes under rotation. This is analogous to equipping the same smooth manifold with different Riemannian metrics: the underlying topology remains unchanged, while the metric geometry varies. The constants m and M quantify how much the geometry can change: for $k = 5$, $M = 2$ and $m = 1/2$, so distances can at most double or halve.

We note that, while each categorical component is finite, the mixed space $\mathcal{X}$ is not: the continuous component $\mathcal{X}_{\mathrm{cont}} \subseteq \mathbb{R}^{n_c}$ is typically unbounded. The equivalence constants m and M are uniform across all epochs and independent of the ordering schedule—a property that does not follow from finite-dimensional compactness alone.

6 Comparison with Classical Encodings

Several established strategies exist for handling categorical variables in optimisation. We compare the transient geometry framework with the most common ones, highlighting what each approach assumes and what it sacrifices.

Integer (ordinal) encoding. Each category is mapped to an integer: $c_i \mapsto i$. This imposes a *fixed* linear order on the categories, creating an artificial notion of "between" (e.g., category c_2 lies between c_1 and c_3) and inducing a locality structure that depends entirely on the chosen labelling. Results become label-dependent: relabelling the categories changes the geometry and potentially the algorithm's behaviour. Transient geometry avoids this by rotating the ordering, so that no single assignment is privileged.

One-hot encoding. Categories are embedded in $\{0,1\}^k$ via indicator vectors—a standard choice in algorithm configuration Hutter et al. [2011] and surrogate-based methods. Under the Hamming metric, all distinct categories become equidistant ($d_H = 2$), destroying any notion of locality. The encoding also increases dimensionality from one variable to k. Transient geometry preserves locality at each instant—some categories are closer than others under the current ordering—while rotating it over time.

Binary encodings and QUBO formulations. Categories are mapped to binary variables $\{0,1\}^m$ with $m = \lceil \log_2 k \rceil$, either as an encoding or as part of a QUBO-style formulation. This creates a fixed discrete geometry (the hypercube graph) in which distances depend on the binary representation. While richer than one-hot encoding, the geometry is entirely induced by the chosen binary representation and remains permanently fixed. Transient geometry is analogous in that it provides combinatorial structure, but avoids committing to any single one.

Latent embeddings. Machine-learning approaches (e.g., entity embeddings, variational autoencoders) learn a continuous representation of categories from data Ru et al. [2020], Deshwal et al. [2021]. These can capture complex relationships but are model-dependent, typically stochastic, require training data, and are often difficult to interpret. Transient geometry is model-free, deterministic (when desired), and fully interpretable: the structure is defined by the cyclic ordering, not learned from data.

Transient geometry. The present framework induces a temporary combinatorial metric via a cyclic ordering that rotates over time. At each instant, the geometry is well defined and provides meaningful locality. Over time, no ordering is privileged, preserving the categorical nature of the variable. The approach is deterministic (though stochastic schedules are also compatible with the theory), model-free, and interpretable.

All classical encodings modify the *representation* of the categorical variable. Transient geometry leaves the representation unchanged and modifies only the induced neighbourhood structure. Table 2 summarises the comparison.

Table 2: Comparison of encoding strategies for categorical variables.

Approach	Order	Geometry	Fixed/Dyn.	Determ.	Locality
Integer encoding	Fixed	Artificial	Fixed	Yes	Label-dep.
One-hot	None	Equidistant	Fixed	Yes	Destroyed
Binary / QUBO	Partial	Hypercube	Fixed	Yes	Fixed
Latent embedding	Learned	Learned	Adaptive	Typically no	Learned
Dist.-induced (CatMADS)	None	Distance-based	Fixed (after init.)	User-def./data-dep.	Problem-specific
Transient geom.	**Temporary**	**Combinatorial**	**Dynamic**	**Yes**	**Rotating**

Example 6.1 (Comparing encodings for $k = 5$)**.** Let $C = \{1, 2, 3, 4, 5\}$. Under **integer encoding** ($c_i \mapsto i$), we get $d(1, 2) = 1$ but $d(1, 5) = 4$—categories 1 and 5 appear maximally far apart, an artefact of the labelling. Under **one-hot encoding**, all distinct pairs are equidistant ($d_H = 2$), destroying any notion of locality: moving from 1 to 2 is the same "distance" as moving from 1 to 5. Under **transient geometry** with ordering $\sigma = (1, 2, 3, 4, 5)$, we get $d^{(\sigma)}_{\text{cat}}(1, 2) = 1$ and $d^{(\sigma)}_{\text{cat}}(1, 5) = 1$—both are cyclic neighbours—while $d^{(\sigma)}_{\text{cat}}(1, 3) = 2$. The geometry provides meaningful locality at each instant, and rotating σ ensures no pair is permanently privileged.

The key distinction is between *fixed* and *dynamic* geometry. Integer, one-hot, and binary encodings each commit to a single geometric structure that never changes and whose properties depend on the chosen representation. Latent embeddings adapt geometry through learning, but at the cost of model dependence.

Transient geometry instead rotates through a finite set of explicitly defined combinatorial structures. By Theorem 5.3, this rotation preserves the topology of the search space while modifying its local metric geometry. It therefore combines the rigour of fixed encodings with the flexibility of adaptive methods, without requiring external data or sacrificing determinism.

7 Illustrative Realisation

We illustrate the framework with a concrete example using $k = 5$ categories, then briefly describe how the Deterministic Neighbourhood Rotation (DNR) mechanism Madeira [2026] instantiates the framework within derivative-free optimisation.

7.1 Transient Geometry with Five Categories

Let $C = \{1, 2, 3, 4, 5\}$ with cyclic-adjacency neighbourhoods (radius $r = 1$). Each cyclic ordering σ induces a cycle graph Cyc_5 on C.

Three orderings. Consider the cyclic orderings

$$\sigma_1 = (1, 2, 3, 4, 5), \qquad \sigma_2 = (1, 3, 5, 2, 4), \qquad \sigma_3 = (1, 4, 2, 5, 3).$$

Each defines a Hamiltonian cycle on C with five adjacency pairs.

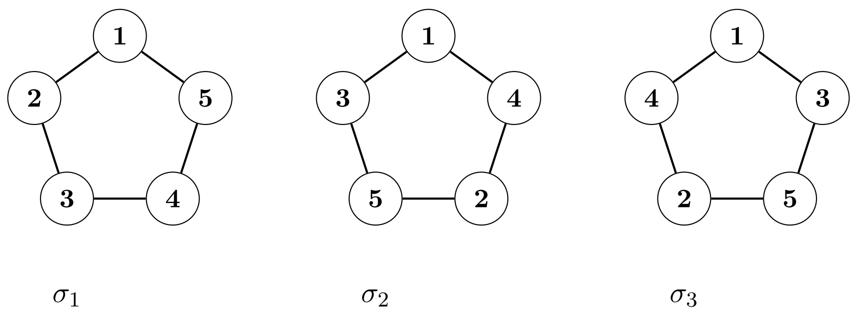


Figure 3: Three cyclic orderings of $C = \{1, 2, 3, 4, 5\}$. Each ordering defines a cycle graph Cyc_5 with five adjacency edges. The vertices are placed at fixed positions to facilitate visual comparison.

Neighbourhood rotation for category 1. The cyclic-adjacent neighbours of category 1 under each ordering are:

$$\mathcal{N}_{\sigma_1}(1) = \{2, 5\}, \qquad \mathcal{N}_{\sigma_2}(1) = \{3, 4\}, \qquad \mathcal{N}_{\sigma_3}(1) = \{3, 4\}.$$

Already after two orderings (σ_1 and σ_2), the union $\mathcal{N}_{\sigma_1}(1) \cup \mathcal{N}_{\sigma_2}(1) = \{2, 3, 4, 5\} = C \setminus \{1\}$ achieves full coverage, consistent with the bound $\lceil (k-1)/2 \rceil = \lceil 4/2 \rceil = 2$ from Proposition 4.31.

The third ordering σ_3 adds no new neighbours for category 1 but contributes to pairwise coverage of other categories.

Global pairwise coverage. The edge sets of the three orderings are:

$$\begin{aligned} E(\sigma_1) &= \big\{\{1,2\},\{2,3\},\{3,4\},\{4,5\},\{1,5\}\big\}, \\ E(\sigma_2) &= \big\{\{1,3\},\{3,5\},\{2,5\},\{2,4\},\{1,4\}\big\}, \\ E(\sigma_3) &= \big\{\{1,4\},\{2,4\},\{2,5\},\{3,5\},\{1,3\}\big\}. \end{aligned}$$

The union $E(\sigma_1) \cup E(\sigma_2)$ already contains all $\binom{5}{2} = 10$ pairs, forming the complete graph K_5. Thus two orderings suffice for full pairwise coverage with $k = 5$, confirming the tight bound. Fig. 4 shows the progressive union.

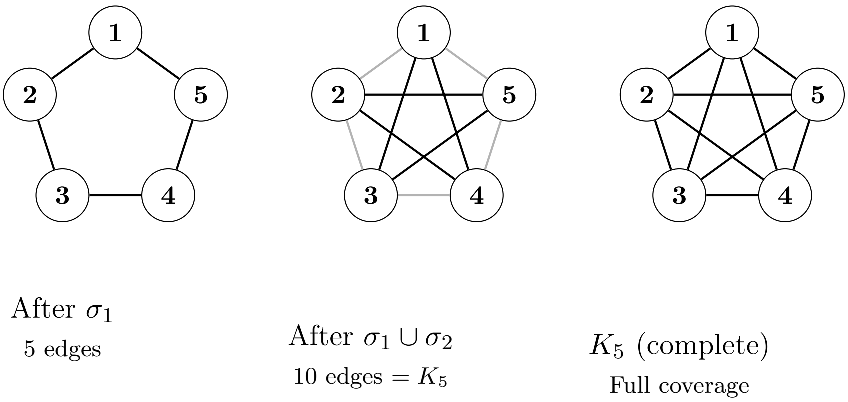


Figure 4: Progressive union of adjacency graphs. Left: cycle from σ_1 (5 edges). Centre: after adding σ_2, all 10 pairs are covered (K_5). Right: the complete graph, confirming full coverage in $\lceil (k-1)/2 \rceil = 2$ orderings. Grey edges are from previous orderings; black edges are newly added.

7.2 Materialising the Optimality Hierarchy

Consider a single categorical variable with $C = \{1, 2, 3, 4, 5\}$ and objective values

$$f(1) = 1, \quad f(2) = 2, \quad f(3) = 0, \quad f(4) = 2, \quad f(5) = 2.$$

Category 1 is an instantaneous—but not robust—optimum. Under $\sigma_1 = (1, 2, 3, 4, 5)$, the neighbours of 1 are $\{2, 5\}$ with $f(2) = f(5) = 2 > 1 = f(1)$. Hence category 1 is an instantaneous local optimum under σ_1. However, under $\sigma_2 = (1, 3, 5, 2, 4)$, the neighbours of 1 are $\{3, 4\}$, and $f(3) = 0 < 1 = f(1)$. Category 1 is *not* a robust local optimum of order 2: the second ordering exposes a better neighbour that was invisible under the first geometry. This is precisely the mechanism of Proposition 4.26 (ii): the improving point $3 \in \mathcal{N}_{\sigma_2}(1) \setminus \mathcal{N}_{\sigma_1}(1)$.

Category 3 is a categorical local optimum. We verify: $f(3) = 0$ is the global minimum of f, so $f(3) \leq f(y)$ for all $y \in C$. In particular, under *any* cyclic ordering σ, every neighbour $y \in \mathcal{N}_\sigma(3)$ satisfies $f(y) \geq 0$. Hence category 3 is an instantaneous optimum under every ordering, a robust optimum of every order, and a categorical local optimum.

Summary. Table 3 shows the hierarchy for this example. The key insight: the rotation of orderings does not merely explore the space—it *invalidates* false optima (category 1 under σ_1) while *confirming* robust ones (category 3).

	f value	Instantaneous optimum (σ_1)	Robust (order 2)	Categorical optimum
Category 1	1	✓	×	×
Category 3	0	✓	✓	✓

Table 3: Optimality hierarchy for the example with $k = 5$ and $f(c)$ as defined in the text. Category 1 passes the weakest test but fails when the geometry rotates; category 3 passes all levels.

7.3 Connection to DNR

The framework developed in Sections 4–5 is self-contained and independent of any particular algorithmic realisation. We now describe, for concreteness, how one specific method—Deterministic Neighbourhood Rotation (DNR) Madeira [2026]—instantiates the abstract framework.

The DNR mechanism proposed in Madeira [2026] for derivative-free optimisation (DFO)—building on the Mesh Adaptive Direct Search framework Audet and Dennis [2006], Abramson et al. [2009] and its recent extension to granular variables Audet and Le Digabel [2024]—is a concrete realisation of the transient geometry framework:

- **Ordering schedule:** DNR uses a deterministic sequence of permutations generated via Sobol-rank indexing, providing low-discrepancy coverage of the cyclic ordering space. This corresponds to a specific choice of schedule $\mathcal{S}$ (Definition 4.13).
- **Epoch structure:** Each epoch in DNR consists of a fixed number of function evaluations under a single cyclic ordering, after which the ordering rotates. This is the epoch-based scheduling of the framework.
- **Independent rotation:** DNR assigns independent permutation sequences to each categorical variable, implementing the per-variable ordering vectors $\boldsymbol{\sigma}_t = (\sigma_t^{(1)}, \ldots, \sigma_t^{(n_{\mathrm{cat}})})$ of Definition 4.16.
- **Robustness validation:** The DNR freezing strategy—which retains the incumbent solution across epochs—naturally implements the monotone robustness filter of Corollary 4.27: a solution that remains best across multiple orderings gains progressively stronger optimality guarantees.

The DNR paper provides algorithmic details, design principles, and computational experiments. The present framework provides the mathematical structure that DNR instantiates, and shows that the underlying ideas generalise beyond DFO to any method that operates on mixed search spaces.

8 Implications and Connections

The metric product framework with transient geometry has implications for several areas of optimisation research. We discuss the most significant ones and highlight open connections.

8.1 Fitness Landscape Analysis

Classical fitness landscape analysis (FLA) Pitzer and Affenzeller [2012], Ochoa et al. [2011], Tomassini et al. [2008] computes indicators (autocorrelation, ruggedness, neutrality, basin sizes) relative to a fixed neighbourhood. Under transient geometry, these indicators become *geometry-dependent*: their values change with the cyclic ordering.

Two approaches are natural:

- *Per-ordering indicators:* compute FLA indicators under each ordering σ_t and study how they vary. The variance of an indicator across orderings measures how sensitive the landscape structure is to the choice of categorical geometry.

- *Ensemble indicators:* average (or take extrema of) FLA indicators over a set of orderings. This provides a geometry-independent summary of landscape structure.

This suggests a new direction: *meta-landscape analysis*—the study of how landscape features change as a function of the geometry imposed on categorical variables. A landscape whose FLA indicators are stable across orderings is structurally robust; one whose indicators vary widely has geometry-dependent difficulty.

8.2 Connection to Randomised Optimisation

The rotation of orderings has a structural analogy with adaptive mechanisms in randomised optimisation:

- **CMA-ES:** The Covariance Matrix Adaptation Evolution Strategy Hansen and Ostermeier [2001] adapts its covariance matrix, effectively rotating and scaling the search geometry in continuous space. Transient geometry performs an analogous rotation in categorical space. In both cases, the topology is unchanged (Theorem 5.3); only the metric geometry adapts.
- **EDAs:** Estimation of Distribution Algorithms Pelikan et al. [2002] learn a distribution over the search space whose categorical component depends on how categories relate to each other. Transient geometry provides a principled, model-free way to define that relationship.
- **Proposal distributions:** In MCMC and related methods, changing the proposal distribution redefines which moves are "local." Changing the cyclic ordering has an analogous effect: it redefines categorical locality. The difference is that transient geometry rotates deterministically, while proposal adaptation is typically stochastic.

8.3 Operator Design

In many mixed-variable methods, operators are designed *per variable type*: a Gaussian perturbation for continuous variables, a ± 1 move for integers, and a random swap for categoricals. The product metric framework suggests a different approach: operators should be designed to be *consistent with the product metric*, generating perturbations whose size is measured by d_t.

This opens the possibility of *unified operators* that work natively on the product space, rather than combining per-component operators. The weights in the product metric (Proposition 4.33) provide a principled way to balance the contribution of each component.

8.4 Surrogate-Based Methods

Surrogate models (Gaussian processes, random forests) in mixed spaces require a kernel or distance function Garrido-Merchán and Hernández-Lobato [2020], Gower [1971], Zaefferer et al. [2014], Zaefferer and Bartz-Beielstein [2020], Pelamatti et al. [2019]. The product metric (1) provides a principled kernel, but the dynamic categorical component raises a design question: should the surrogate use a *fixed* metric (e.g., the trivial discrete metric), the *current* ordering metric, or an *ensemble* kernel averaging over multiple orderings? The uniform equivalence result (Theorem 5.3) guarantees that all choices induce the same topology, but they differ in their geometry and hence in the surrogate's interpolation behaviour.

9 Open Questions

The transient geometry framework opens several theoretical and algorithmic questions. We highlight five that we consider most significant.

(1) Richer transient structures and combinatorial extensions

Cyclic adjacency provides a minimal symmetric combinatorial geometry (Section 4.2), but richer alternatives exist. Permutation metrics such as Kendall tau or Cayley distance Diaconis [1988], Schiavinotto and Stützle [2007], Irurozki et al. [2019] define different neighbourhood structures on finite sets; it remains open whether rotating through such metrics can yield sharper coverage bounds or stronger optimality guarantees than cyclic adjacency. More broadly, many optimisation problems involve trees, graphs, permutations, or other structured objects as decision variables. Can the product-of-metric-spaces framework accommodate such types? Is there a broader class of dynamic metric spaces in which similar coverage (Theorem 4.28) and robustification (Section 4.8) results hold?

(2) Basins, robustness, and landscape invariants

The hierarchy of local optimality (Definitions 4.23–4.25) naturally induces a corresponding hierarchy of basins of attraction under dynamic metrics. A systematic study of basin robustness—how basins shrink as more orderings are visited, and whether basin persistence across geometries correlates with problem difficulty—constitutes an interesting direction for future work. More generally, uniform metric equivalence (Theorem 5.3) ensures topological invariance, but geometric landscape features—ruggedness, autocorrelation, neutrality—change under rotation. Which features are invariant, and do ensemble indicators averaged over orderings (Section 8.1) provide more reliable landscape characterisations than single-geometry indicators?

(3) Ordering schedules and weight selection

The theory is agnostic to the choice of ordering schedule (Definition 4.13) and product metric weights (Section 4.7). However, these design parameters affect convergence rates and robustness profiles. Do low-discrepancy sequences in permutation space provide provable advantages over random or round-robin schedules? Existing approaches map low-discrepancy points from the unit hypercube to permutations via ranking Madeira [2026]; however, this mapping does not formally preserve the discrepancy properties of the original sequence. The construction of low-discrepancy sequences *native* to the symmetric group—with discrepancy measured relative to a natural metric such as Kendall tau or Cayley distance—remains an open problem with independent combinatorial interest. Can ordering schedules be adapted dynamically while preserving coverage guarantees (Proposition 4.31)? Similarly, can the weights w_c, w_i, w_{cat} be learned adaptively—for instance from early evaluations—without destroying the theoretical guarantees, and is there a principled normalisation across variables with different cardinalities?

(4) Convergence theory beyond local optimality

The present work shows that recurrent descent schemes produce accumulation points satisfying categorical local optimality (Section 4.8). A natural extension is to integrate Clarke stationarity for continuous components Audet and Dennis [2006], Vicente and Custódio [2012] with categorical robustness, yielding a full mixed-variable stationarity theory under dynamic metrics. This likely requires substantially different analytical machinery (e.g., dense polling directions and Lipschitz hypotheses). Can global convergence rates be derived? Can trust-region or model-based frameworks be extended to rotating metric geometries?

(5) Surrogate modelling with dynamic kernels

Surrogate models in mixed spaces Zaefferer and Bartz-Beielstein [2020], Ru et al. [2020] require a kernel or distance function. The product metric (Section 8.4) provides a principled choice, but the dynamic categorical component raises a design question: should the surrogate use the current ordering's metric, a fixed metric, or an ensemble kernel averaging over orderings? How does the choice of kernel geometry affect surrogate accuracy and optimisation performance?

Collectively, these questions span combinatorics, metric geometry, algorithmic design, and landscape analysis, indicating that transient geometry connects several areas of optimisation theory. They suggest that this perspective may extend beyond categorical handling to mixed-variable optimisation viewed more broadly through the lens of metric geometry.

10 Conclusions

This paper has proposed a geometric framework for mixed-variable optimisation in which categorical variables—traditionally treated as structureless—are endowed with a transient combinatorial geometry that rotates over time. The main contributions are:

(i) A *separation principle* distinguishing the ontological nature of a variable (categorical: no fixed order) from the operational structure imposed on it (a temporary cyclic ordering providing distance and neighbourhood).

(ii) A formalisation of *transient geometry* via cyclic-ordering-induced metrics on finite sets, where the ordering rotates according to a prescribed schedule while preserving the categorical nature of the variable.

(iii) A rigorous model of mixed search spaces as *weighted Cartesian products of metric spaces*, with a product metric that remains well defined under rotation of the categorical component.

(iv) A *hierarchy of local optimality*—instantaneous, robust, and categorical—capturing the degree to which a local optimum is independent of the chosen geometry, together with a monotone robustness result showing that the robustness notion strengthens as more orderings are visited.

(v) A *coverage theorem* establishing that $\lceil (k-1)/2 \rceil$ cyclic orderings are sufficient for full categorical adjacency coverage for a k-level variable, with a tight bound based on a classical Hamiltonian decomposition.

(vi) An *algorithmic interpretation* showing that standard epoch-based descent under recurrent orderings naturally produces accumulation points satisfying robust—and, under coverage, categorical—local optimality.

(vii) A *uniform metric equivalence* result proving that all dynamic metrics induced by different orderings generate the same topology, addressing the concern that ordering rotation "changes the space."

The framework shows that categorical variables, despite lacking canonical order, admit a rigorous geometric treatment with quantified coverage bounds. The mixed search space is a mathematically coherent product structure, and categorical variables are not merely a special case to be handled by representation tricks, but components whose geometry can be defined, rotated, and analysed.

The Deterministic Neighbourhood Rotation mechanism Madeira [2026] serves as a concrete instantiation of the framework within derivative-free optimisation, but the theory is not specific to any single algorithmic paradigm. The product metric structure, the optimality hierarchy, and the coverage results apply to any method that operates on mixed search spaces: evolutionary, surrogate-based, hybrid, or otherwise. Algorithmically, transient geometry induces a progressive validation mechanism: as orderings rotate, local optimality is strengthened from instantaneous to robust and, under coverage, to categorical. A full convergence theory integrating Clarke-type stationarity for continuous components constitutes a natural analytical extension of this structural perspective. From a practical standpoint, the transient geometry approach replaces the $k-1$ neighbours per variable required by exhaustive enumeration with a constant-degree neighbourhood (two neighbours per epoch), spreading the coverage cost over $\lceil (k-1)/2 \rceil$ epochs while integrating naturally with descent-based search. The framework therefore shifts categorical handling from representation engineering to geometric design, placing mixed-variable optimisation within a unified metric-structural perspective.

Recent algorithmic developments such as CatMADS Audet et al. [2025] confirm the importance of categorical structure in direct search methods. The present work provides a complementary geometric perspective that formalises this structure independently of any particular algorithmic realisation.

We hope that the geometric perspective developed here—and the open questions it raises—will stimulate further theoretical and algorithmic work on the structure of mixed-variable optimisation problems.

Declarations

Funding. J. F. A. Madeira acknowledges the support of Fundação para a Ciência e a Tecnologia (FCT) through LAETA (project `https://doi.org/10.54499/UID/50022/2025`). This work was developed within the framework of COST Action CA22137 "Randomised Optimisation Algorithms Research Network" (ROAR-NET), supported by COST (European Cooperation in Science and Technology, `https://www.cost.eu/actions/CA22137/`). The author is a member of Working Group 2 (Mixed continuous and discrete optimisation).

Conflict of interest. The author declares that he has no conflict of interest.

Author contributions. The author contributed to conceptualisation, formal analysis, writing, and editing.

Data and code availability. A companion repository with constructive generators and verification scripts will be made public upon acceptance.

References

M. A. Abramson, C. Audet, J. E. Dennis, and S. Le Digabel. OrthoMADS: A deterministic MADS instance with orthogonal directions. *SIAM Journal on Optimization*, 20(2):948–966, 2009. doi: 10.1137/080716980.

B. Alspach, J.-C. Bermond, and D. Sotteau. Decomposition into cycles I: Hamilton decompositions. In G. Hahn, G. Sabidussi, and R. E. Woodrow, editors, *Cycles and Rays*, pages 9–18. Kluwer Academic Publishers, 1990.

C. Audet and J. E. Dennis. Mesh adaptive direct search algorithms for constrained optimization. *SIAM Journal on Optimization*, 17(1):188–217, 2006. doi: 10.1137/040603371.

C. Audet and W. Hare. *Derivative-Free and Blackbox Optimization.* Springer Series in Operations Research and Financial Engineering. Springer, Cham, 2017. doi: 10.1007/978-3-319-68913-5.

C. Audet and S. Le Digabel. The mesh adaptive direct search algorithm for granular and discrete variables. *Optimization and Engineering*, 25:693–715, 2024. doi: 10.1007/s11081-023-09810-0.

Charles Audet, Youssef Diouane, Edward Hallé-Hannan, Sébastien Le Digabel, and Christophe Tribes. CatMADS: Mesh adaptive direct search for constrained blackbox optimization with categorical variables. *arXiv preprint arXiv:2506.06937*, 2025. arXiv:2506.06937 [math.OC].

A. R. Conn, K. Scheinberg, and L. N. Vicente. *Introduction to Derivative-Free Optimization.* MPS-SIAM Series on Optimization. SIAM, Philadelphia, 2009. doi: 10.1137/1.9780898718768.

A. Deshwal, S. Belakaria, and J. R. Doppa. Bayesian optimization over combinatorial structures. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 34, pages 14476–14488, 2021.

P. Diaconis. *Group Representations in Probability and Statistics*, volume 11 of *IMS Lecture Notes—Monograph Series*. Institute of Mathematical Statistics, Hayward, CA, 1988. doi: 10.1214/lnms/1215467407.

E. C. Garrido-Merchán and D. Hernández-Lobato. Dealing with categorical and integer-valued variables in Bayesian optimization with Gaussian processes. *Neurocomputing*, 380:20–35, 2020. doi: 10.1016/j.neucom.2019.11.033.

J. C. Gower. A general coefficient of similarity and some of its properties. *Biometrics*, 27(4): 857–871, 1971. doi: 10.2307/2528823.

N. Hansen and A. Ostermeier. Completely derandomized self-adaptation in evolution strategies. *Evolutionary Computation*, 9(2):159–195, 2001. doi: 10.1162/106365601750190398.

F. Hutter, H. H. Hoos, and K. Leyton-Brown. Sequential model-based optimization for general algorithm configuration. In *Proc. 5th Intl. Conference on Learning and Intelligent Optimization (LION 5)*, volume 6683 of *Lecture Notes in Computer Science*, pages 507–523. Springer, 2011. doi: 10.1007/978-3-642-25566-3_40.

E. Irurozki, B. Calvo, and J. A. Lozano. Mallows and generalized Mallows model for matchings. *Bernoulli*, 25(2):1160–1188, 2019. doi: 10.3150/17-BEJ1017.

Y.-L. Li, Y.-Y. Wah, R. Xia, and T. Cheng. Mixed-variable optimization problems: A survey. *Journal of the Operational Research Society*, 64:1549–1564, 2013. doi: 10.1057/jors.2013.22.

J. F. A. Madeira. Deterministic neighbourhood rotation for categorical variables in derivative-free optimisation. Submitted, 2026.

G. Ochoa, M. Tomassini, S. Vérel, and C. Darabos. A study of NK landscapes' basins and local optima networks. *Journal of Heuristics*, 17(1):75–104, 2011. doi: 10.1007/s10732-010-9145-1.

J. Pelamatti, L. Brevault, M. Balesdent, E.-G. Talbi, and Y. Guerin. Efficient global optimization of constrained mixed variable problems. *Journal of Global Optimization*, 73(3):583–613, 2019. doi: 10.1007/s10898-018-0715-1.

M. Pelikan, D. E. Goldberg, and F. G. Lobo. A survey of optimization by building and using probabilistic models. *Computational Optimization and Applications*, 21(1):5–20, 2002. doi: 10.1023/A:1013500812258.

E. Pitzer and M. Affenzeller. A comprehensive survey on fitness landscape analysis. In *Recent Advances in Intelligent Engineering Systems*, volume 378 of *Studies in Computational Intelligence*, pages 161–191. Springer, Berlin, 2012. doi: 10.1007/978-3-642-23229-9_8.

B. Ru, A. S. Alvi, V. Nguyen, M. A. Osborne, and S. J. Roberts. Bayesian optimisation over multiple continuous and categorical inputs. In *Proceedings of the 37th International Conference on Machine Learning (ICML)*, volume 119 of *Proceedings of Machine Learning Research*, pages 8276–8285. PMLR, 2020.

Paul Saves, Rémi Lafage, Nathalie Bartoli, Youssef Diouane, Jasper Bussemaker, Thierry Lefebvre, John T. Hwang, Joseph Morlier, and Joaquim R. R. A. Martins. SMT 2.0: A surrogate modeling toolbox with a focus on hierarchical and mixed variables Gaussian processes. *Advances in Engineering Software*, 188:103571, 2024. doi: 10.1016/j.advengsoft.2023.103571.

T. Schiavinotto and T. Stützle. A review of metrics on permutations for search landscape analysis. *Computers & Operations Research*, 34(10):3143–3153, 2007. doi: 10.1016/j.cor.2005.11.022.

Peter F. Stadler. Fitness landscapes. In Michael Lässig and Angelo Valleriani, editors, *Biological Evolution and Statistical Physics*, volume 585 of *Lecture Notes in Physics*, pages 183–204. Springer, 2002. doi: 10.1007/3-540-45692-9_10.

M. Tomassini, S. Vérel, and G. Ochoa. Complex-network analysis of combinatorial spaces: The NK landscape case. *Physical Review E*, 78(6):066114, 2008. doi: 10.1103/PhysRevE.78.066114.

L. N. Vicente and A. L. Custódio. Analysis of direct searches for discontinuous functions. *Mathematical Programming*, 133(1–2):299–325, 2012. doi: 10.1007/s10107-010-0429-8.

M. Zaefferer and T. Bartz-Beielstein. Surrogate-assisted discrete optimization. In *High-Performance Simulation-Based Optimization*, volume 833 of *Studies in Computational Intelligence*, pages 275–308. Springer, 2020. doi: 10.1007/978-3-030-18764-4_9.

M. Zaefferer, J. Stork, and T. Bartz-Beielstein. Distance measures for permutations in combinatorial efficient global optimization. In *Parallel Problem Solving from Nature – PPSN XIII*, volume 8672 of *Lecture Notes in Computer Science*, pages 373–383. Springer, 2014. doi: 10.1007/978-3-319-10762-2_37.

Appendix A: Explicit Generation of Covering Cyclic Orderings

This appendix provides a constructive procedure to generate a minimal family of cyclic orderings achieving full categorical adjacency coverage, as stated in Proposition 4.31. The construction is explicit, deterministic, and directly implementable.

A.1 Construction Principle

Let $C = \{1, \dots, k\}$. We seek $M = \lceil \frac{k-1}{2} \rceil$ cyclic orderings whose adjacency pairs cover all $\binom{k}{2}$ unordered pairs.

Case 1: k odd ($k = 2n+1$). Walecki's construction decomposes K_k into exactly $n = (k-1)/2$ edge-disjoint Hamiltonian cycles Alspach et al. [1990]. Fix vertex 1. Arrange the remaining vertices $2, \dots, k$ on a regular $(k-1)$-gon. For each shift $s = 0, \dots, n-1$, rotate the polygon by s positions and trace the zigzag path: starting from vertex 1, alternately visit the top and bottom of the rotated arrangement. Each shift yields a Hamiltonian cycle; the family is edge-disjoint and partitions K_k.

Case 2: k even ($k = 2n$). Fix vertex k and arrange vertices $1, \dots, k-1$ on a regular $(k-1)$-gon. The standard round-robin (circle method) 1-factorisation produces $k-1$ perfect matchings of K_k: in round r ($r = 1, \dots, k-1$), match vertex k with the current top vertex and pair the remaining vertices symmetrically across the polygon; then rotate all polygon vertices by one position.

The $n = k/2$ Hamiltonian cycles are obtained by pairing round r with round $r + (n-1) \pmod{k-1}$, for $r = 1, \dots, n$. A standard property of the circle method pairing ([Alspach et al., 1990, §3, Theorem 1]) ensures that each such union of two matchings forms a single Hamiltonian cycle. The resulting n cycles jointly cover all edges of K_k, yielding a covering family of minimal size

$$n = \frac{k}{2} = \left\lceil \frac{k-1}{2} \right\rceil$$

cyclic orderings with full pairwise coverage.

A.2 Deterministic Generation Algorithm

```
Input: integer k >= 3
Output: M = ceil((k-1)/2) cyclic orderings sigma_1,...,sigma_M

If k is odd (k = 2n+1):
    Fix vertex 1. Place vertices 2,...,k on a (k-1)-gon.
    For s = 0,...,n-1:
        Rotate the polygon by s positions.
        Construct sigma_{s+1} by zigzag traversal from vertex 1:
            alternately visit top and bottom of rotated arrangement.
    Return n cycles.

If k is even (k = 2n):
    Fix vertex k. Place vertices 1,...,k-1 on a (k-1)-gon.
    Compute k-1 perfect matchings via round-robin scheduling:
        In each round, pair vertex k with the top vertex;
        pair remaining vertices symmetrically; rotate by 1.
    For t = 1,...,n:
        Combine matching t with matching t+(n-1) mod (k-1).
        Trace the Hamiltonian cycle from the union of edges.
    Return n cycles.
```

Both constructions are $O(k^2)$ and require only modular arithmetic. No randomisation is involved; the output depends only on k.

A.3 Explicit Cyclic Orderings for $k = 3, \ldots, 10$

Below we list verified minimal covering families. Each family has been computationally confirmed to cover all $\binom{k}{2}$ unordered pairs.

$k = 3$, $M = 1$

$$(1, 2, 3)$$

$k = 4$, $M = 2$

$$(1, 2, 3, 4), \quad (1, 3, 4, 2)$$

$k = 5$, $M = 2$

$$(1, 2, 3, 4, 5), \quad (1, 3, 5, 2, 4)$$

$k = 6$, $M = 3$

$$(1, 2, 3, 4, 5, 6), \quad (1, 3, 5, 2, 6, 4), \quad (1, 2, 4, 3, 6, 5)$$

$k = 7$, $M = 3$

$$(1, 2, 3, 4, 5, 6, 7), \quad (1, 3, 5, 7, 2, 4, 6), \quad (1, 4, 7, 3, 6, 2, 5)$$

$k = 8$, $M = 4$

$$(1, 2, 3, 4, 5, 6, 7, 8), \quad (1, 3, 7, 2, 6, 4, 8, 5),$$
$$(1, 6, 8, 3, 5, 2, 4, 7), \quad (1, 4, 2, 8, 3, 6, 5, 7)$$

$k = 9$, $M = 4$

$$(1, 2, 3, 4, 5, 6, 7, 8, 9), \quad (1, 5, 9, 3, 8, 6, 4, 2, 7),$$
$$(1, 6, 2, 9, 4, 7, 3, 5, 8), \quad (1, 3, 6, 9, 7, 5, 2, 8, 4)$$

$k = 10$, $M = 5$

$$(1,2,3,4,5,6,7,8,9,10), \quad (1,3,6,10,7,5,8,4,2,9),$$
$$(1,6,2,8,3,10,5,9,4,7), \quad (1,4,6,8,10,2,7,9,3,5),$$
$$(1,3,7,4,10,6,9,5,2,8)$$

These constructions are classical in extremal graph theory and yield symmetric families up to relabelling.

A.4 Practical Implications

- The number of required epochs is linear in k:

$$M = \left\lceil \frac{k-1}{2} \right\rceil .$$

- Each ordering induces degree 2 per category, so the neighbourhood size per epoch remains constant.
- Full categorical adjacency coverage is achieved without exhaustive enumeration.
- The construction is deterministic, reproducible, and independent of labelling choices.

This demonstrates that the coverage result of Proposition 4.31 is not merely existential but directly constructive and operationally implementable.

The explicit families for $k = 3, \ldots, 10$ were independently verified by a short script; a companion repository with constructive generators and verification scripts will be made public upon acceptance.