A GEOMETRIC VARIATIONAL VIEW OF COMPETITIVE BOUNDARIES IN METRIC LEARNING

Anonymous authors
Paper under double-blind review

ABSTRACT

We study metric learning in tasks where predictions are selected from a learned score landscape and correctness is defined by task-specific tolerances. We develop a geometric bridge from these tolerances to parameter-space learning directions, providing a task-derived reference for evaluating comparison-based updates. Specifically, we quantify the extent of acceptable hypotheses that outrank all unacceptable alternatives, turning the model's realization of task tolerance into a regional objective. Applying Hadamard's first-variation formula, we derive the steepest direction for first-order regional expansion and connect it to the direction governing region formation. These directions allow the task utility of individual training comparisons and their aggregation by metric losses to be assessed within a common framework. Empirical analyses on real-world datasets support the theoretical predictions and link regional expansion to improved task performance. Together, these results establish a geometric bridge between metric-learning updates and task-defined regions of correct prediction.

1 INTRODUCTION

Consider a score-based prediction task in which a model parameterized by θ assigns a score $F_\theta(x, y)$ to each candidate hypothesis $y \in \mathcal{Y}$ for an input x LeCun et al. (2006); Tsochantaridis et al. (2005). Fixing x yields the score landscape $F_\theta(x, \cdot) : \mathcal{Y} \to \mathbb{R}$, with prediction $\hat{y} = \arg\max_{y \in \mathcal{Y}} F_\theta(x, y)$.

Across many score-based prediction tasks, candidate outputs are commonly evaluated according to task-specific criteria rather than exact equality to a nominal target Taskar et al. (2003); Sattler et al. (2018). We denote the resulting set of valid hypotheses by the *acceptable region* $\mathcal{G} \subseteq \mathcal{Y}$. Under the argmax rule, correct selection requires that at least one hypothesis in $\mathcal{G}$ outrank the competing hypotheses outside it.

This set-valued notion of correctness extends the large-margin view of comparing a target output against competing outputs Tsochantaridis et al. (2005). For each acceptable hypothesis $y \in \mathcal{G}$, its *competitive margin* therefore can be defined as $m_\theta(y) = F_\theta(x, y) - \sup_{z \in \mathcal{Y} \setminus \mathcal{G}} F_\theta(x, z)$. A positive competitive margin means that y itself outranks every unacceptable hypothesis. This induces the *winning region* $\mathcal{R}_\theta = \{y \in \mathcal{G} : m_\theta(y) > 0\}$, the subset of task-valid hypotheses that are actually competitive under the current score landscape.

Two models may therefore both rank the nominal ground-truth hypothesis correctly, yet induce substantially different winning regions, with $\mathcal{R}_\theta$ covering a large portion of $\mathcal{G}$ in one landscape but only a small neighborhood around the nominal state in the other. This distinction shifts the task objective from merely making the nominal hypothesis win to enlarging the portion of the acceptable region that remains competitive against unacceptable alternatives, an extent quantified by the *winning-region measure* $W_\theta := \mu(\mathcal{R}_\theta)$, where μ is a reference measure over the hypothesis domain. This makes explicit the dependency $\theta \to F_\theta(x, \cdot) \to m_\theta \to \mathcal{R}_\theta \to W_\theta$. (Rumelhart et al., 1986)

Metric learning provides a practical mechanism for shaping the landscape $F_\theta(x, \cdot)$ by constructing updates to θ from sampled comparisons between an acceptable hypothesis and a competing hypothesis outside the acceptable region Hadsell et al. (2006); Schroff et al. (2015); Song et al. (2015); Sohn (2016). Each comparison contributes a parameter-space update direction, while sampling, mining, and metric-loss weighting determine how these directions are selected and aggregated into the model update Manmatha et al. (2017); Wang et al. (2019); Sun et al. (2020). These mechanisms

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.

operate on comparison-level quantities, whereas the task objective considered here is the regional quantity W_θ. Because all comparison directions act through the shared parameters θ, the effect of an individual direction is generally not confined to the hypotheses in that comparison: it can alter scores elsewhere in the landscape and interact with the effects of other comparisons. Consequently, comparisons with similar pairwise violations need not induce similar changes in the winning region.

In this work, we shift the analysis from comparison-level scalars to parameter-space directions, asking *which directions produce favorable first-order changes in the task quantity W_θ*. Our insight is that, once a winning region exists, its extent need not change solely through an increase in competitive height. The same change in W_θ may arise from variations in the local width, curvature, or more general shape of the competitive landscape. Despite these different deformation modes, their first-order effect on winning-region extent is governed by the motion of the competitive boundary. Motivated by this insight, we introduce boundary variation as the geometric interface between parameter-space updates and task-level changes in W_θ Delfour & Zolésio (2011); Sokołowski & Zolésio (1992). This allows different landscape deformations to be evaluated through a common first-order task quantity. Among parameter-space directions of equal magnitude, the direction producing the largest first-order increase in W_θ then defines the task-optimal direction for region expansion.

Building on these views, we develop a geometric framework that connects metric-learning updates to downstream task geometry, with three main components: **(1)** We formulate downstream task tolerance geometrically over the hypothesis space, representing correct prediction through the region in which acceptable hypotheses outperform competing alternatives and using its formation and extent to characterize how the score landscape realizes task tolerance. **(2)** We derive first-order task-optimal directions in parameter space from the winning-region geometry, using boundary variation to characterize the direction of maximal first-order expansion of an existing winning region and competitive-height variation to characterize its formation counterpart. **(3)** Based on these task-optimal directions, we formulate task utility as a common reference for evaluating individual comparison directions, their loss-dependent aggregation, and the relation between comparison hardness and downstream task utility. In this work, we use visual localization as the primary empirical setting, where task-defined pose tolerances naturally induce a set of valid hypotheses and therefore provide a concrete setting for studying winning-region geometry.

2 BACKGROUND

Score-based hypothesis selection under task tolerance. Many prediction problems can be expressed as hypothesis selection over a learned score landscape (LeCun et al., 2006; Florence et al., 2022; Murphy et al., 2021). Given an input x, a model assigns scores over a hypothesis space $\mathcal{Y}$, and prediction is obtained by selecting or searching for a high-scoring hypothesis. This formulation appears in retrieval and localization as well as energy- or probability-field models such as IBC (Florence et al., 2022) and Implicit-PDF (Murphy et al., 2021). Existing work has largely focused on learning, representing, or efficiently searching such score landscapes (Han et al., 2023). In many downstream tasks, however, correctness does not correspond to a unique hypothesis: a set of nearby states may satisfy the task tolerance (Deuser et al., 2023). This changes the relevant geometric object from a single maximizer to the behavior of the landscape over a task-defined acceptable set $\mathcal{G}$.

Geometry of task tolerance. Task tolerance turns correctness from a pointwise condition into a regional one: predictions should remain valid over a set of admissible hypotheses rather than only at a nominal target. Robust representation learning provides one established way to operationalize such regional correctness: it first specifies a set of admissible perturbations and then encourages the learned prediction or representation to remain valid throughout that set Madry et al. (2017); Tsipras et al. (2018); Cohen et al. (2019). Viewed through our score-landscape formulation, both perspectives share the goal of making task-valid hypotheses within $\mathcal{G}$ outrank hypotheses outside the acceptable region. The key difference is that robust representation learning typically seeks to enforce this condition throughout the prescribed region, corresponding to the full-coverage case $\mathcal{R}_\theta = \mathcal{G}$. Our formulation, by contrast, treats the winning region as a model-dependent subset of the acceptable region, $\mathcal{R}_\theta \subseteq \mathcal{G}$, and studies its expansion toward this full-coverage limit. To characterize this expansion, we draw on level-set geometry and shape calculus. Since $\mathcal{G}$ is fixed while the model-induced $\mathcal{R}_\theta$ varies with the score landscape, parameter updates move its interior

competitive boundary Γ_θ. Classical shape-variation results relate this boundary motion to the first-order change of regional measure through boundary integrals (Delfour & Zolésio, 2011; Sokołowski & Zolésio, 1992). We use this machinery to connect parameter-space learning directions to winning-region variation.

Metric Learning as Geometric Landscape Shaping. Metric learning constructs supervision from pairwise, triplet, or multi-sample comparisons, with different objectives determining which relative similarities should be increased or decreased (Hadsell et al., 2006; Schroff et al., 2015; Song et al., 2015; Sohn, 2016). A substantial line of work has therefore focused on selecting informative comparisons through sampling or hard-example mining (Schroff et al., 2015; Manmatha et al., 2017), and on assigning comparison-dependent weights according to quantities such as similarity, violation, or hardness (Wang et al., 2019; Sun et al., 2020). From a geometric perspective, these mechanisms can be interpreted as comparison-level prescriptions for which relative score differences should be increased or decreased. Such prescriptions characterize how individual comparisons shape local score relations, but do not determine how favorable the resulting parameter-space directions are for the downstream task. This leaves a gap between comparison-level importance, as encoded by similarity, violation, or hardness, and the task-level utility of the corresponding update directions. Our analysis makes this distinction explicit through winning-region geometry.

3 From Winning-Region Extent to Task Performance

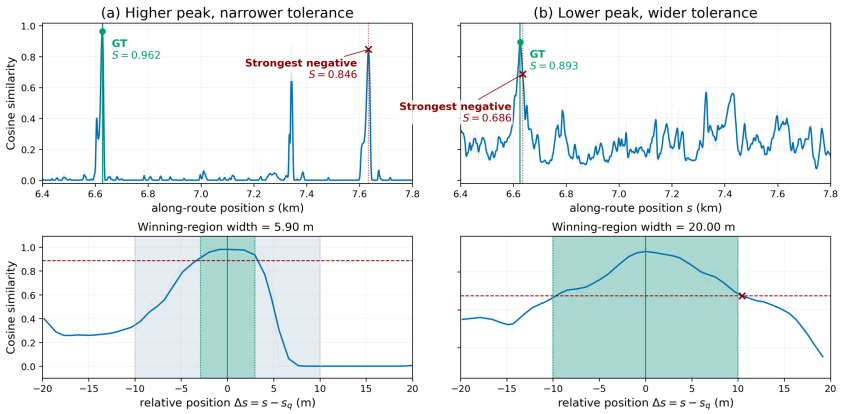


Figure 1: **Why winning-region extent matters.** Schematic one-dimensional score landscapes showing that peak height and winning-region extent need not coincide. A wider winning region provides greater tolerance to finite hypothesis sampling at inference.

In practical score-based prediction, the hypothesis evaluated at inference is generally only an approximation to the underlying continuous state. For example, retrieval and localization operate over discretized reference sets or finite proposal sets, so the available hypotheses need not contain the exact ground-truth state. We now ask how the winning-region extent translates into task performance under such hypothesis approximation.

Table 1: Retrieval performance across winning-region regimes in one-dimensional visual localization. Queries are grouped by winning-region extent, with median statistics reported within each group.

	D_W (m)	P_{hit} (%)	R@1 (%)
G1	1.98	44.64	50.23
G2	4.52	67.04	78.97
G3	7.28	81.64	92.99
G4	8.27	84.43	93.69

Finite hypothesis sampling provides a tractable setting in which this relationship can be quantified. For a candidate set $\mathcal{C}_B = \{Y_b\}_{b=1}^{B}$ of size B, with $Y_b \overset{\text{i.i.d.}}{\sim} \pi$, define the finite-set prediction as $\hat{y}_B = \arg\max_{y \in \mathcal{C}_B} F_\theta(x, y)$. With $W_\theta^\pi := \pi(\mathcal{R}_\theta)$, the probability that at least one candidate falls inside the winning region is

$$P_{\text{hit}}(B) = \Pr(\mathcal{C}_B \cap \mathcal{R}_\theta \neq \varnothing) = 1 - (1 - W_\theta^\pi)^B . \tag{1}$$

Since any hypothesis in $\mathcal{R}_\theta$ is task-valid and outperforms all unacceptable hypotheses, hitting $\mathcal{R}_\theta$ is sufficient for correct finite-set selection. Therefore, $P_{\mathrm{hit}}(B)$ provides a lower bound on the task-success probability, $\Pr(\hat{y}_B \in \mathcal{G}) \geq P_{\mathrm{hit}}(B)$, and, for fixed B, this lower bound increases monotonically with W_θ^π. Figure 1 illustrates this relationship geometrically.

We empirically examine this relation in a one-dimensional visual localization setting. For each query, we measure the winning-region extent, the corresponding hitting probability, and Recall@1, and group the queries into four data-driven regimes according to winning-region extent. The resulting statistics are reported in Table 1.

4 Geometric Variation of Competitive Tolerance

4.1 First-Order Variation of Competitive Tolerance

For an existing winning region, the first-order variation of its measure is governed by the motion of its interior boundary. Classical shape calculus expresses such regional variation through the normal displacement of the moving boundary (Delfour & Zolésio, 2011; Sokołowski & Zolésio, 1992). In our setting, this boundary is the zero level set of the competitive margin,

$$\Gamma_\theta = \{y \in \operatorname{int} \mathcal{G} : m_\theta(y) = 0\}, \tag{2}$$

which we refer to as the *competitive boundary*. Under the standard regularity condition $\nabla_y m_\theta(y) \neq 0$ on Γ_θ, its motion can be related to parameter-induced changes of m_θ through standard level-set kinematics (Subasic, 2003; Osher & Sethian, 1988). The precise regularity conditions are stated in Appendix A.

Lemma 1 (Level-set boundary displacement)**.** *For $y \in \Gamma_\theta$, let $\delta m_\theta(y)$ denote the parameter-induced first-order change of the competitive margin at the fixed hypothesis y, and let $\delta r(y)$ denote the resulting signed outward normal displacement of the competitive boundary. The former describes the local change of the competitive landscape, while the latter describes the corresponding geometric motion of its zero level set. Standard level-set kinematics relates the two quantities by*

$$\delta r(y) = \frac{\delta m_\theta(y)}{\|\nabla_y m_\theta(y)\|}. \tag{3}$$

Proof. Using $n_{\mathrm{out}} = -\nabla_y m_\theta / \|\nabla_y m_\theta\|$, the first-order variation of the zero-level condition directly gives $\delta r(y) = \delta m_\theta(y) / \|\nabla_y m_\theta(y)\|$. A complete derivation is given in Appendix B. □

Equation (3) shows that a margin variation $\delta m_\theta(y)$ induces signed boundary motion, scaled by the inverse local spatial margin slope $1/\|\nabla_y m_\theta(y)\|$. The task quantity W_θ, however, depends on the aggregate effect of these local displacements over the entire competitive boundary Γ_θ.

Proposition 1 (First-order variation of the winning-region measure)**.** *Let the reference measure μ admit a local density $\rho(y)$. For a signed normal displacement $\delta r(y)$ of the competitive boundary, the Hadamard variation formula gives $\delta W_\theta = \int_{\Gamma_\theta} \rho(y)\, \delta r(y)\, dS(y)$. Combining this relation with Lemma 1 yields*

$$\delta W_\theta = \int_{\Gamma_\theta} \frac{\delta m_\theta(y)}{\|\nabla_y m_\theta(y)\|}\, \rho(y)\, dS(y). \tag{4}$$

Proof. The first identity is the standard Hadamard variation of a domain integral (Delfour & Zolésio, 2011; Sokołowski & Zolésio, 1992). Substituting the level-set displacement in Eq. (3) gives Eq. (4). □

Equation (4) weights local margin changes by $\rho(y)/\|\nabla_y m_\theta(y)\|$. Thus, equal margin improvements over boundary patches of equal surface measure can contribute differently to winning-region expansion, depending on the reference density and local spatial margin slope. Learning acts through a shared parameter perturbation $\delta\theta$, which couples these local margin changes through $\delta m_\theta(y) = \langle \nabla_\theta m_\theta(y), \delta\theta \rangle$. Substituting this relation into Eq. (4) yields the corresponding parameter-space gradient.

Corollary 1 (Parameter-space gradient of the winning-region measure)**.** *The gradient of the winning-region measure is*

$$\nabla_\theta W_\theta = \int_{\Gamma_\theta} \frac{\nabla_\theta m_\theta(y)}{\|\nabla_y m_\theta(y)\|}\, \rho(y)\, dS(y). \tag{5}$$

When nonzero, $\nabla_\theta W_\theta$ defines the direction of steepest first-order expansion under a fixed Euclidean update norm. It therefore provides a task-optimal reference for evaluating comparison-induced learning directions.

4.2 Task Utility of Training Comparisons

The previous section established $\nabla_\theta W_\theta$ as the task-optimal direction for expanding the winning region. Metric learning, however, constructs parameter updates from sampled positive–negative comparisons. This raises a natural question: *how strongly does the update induced by a particular training comparison expand or contract the winning region?*

Comparison-induced direction. In the score-landscape framework, a training comparison pairs an acceptable hypothesis $y^+ \in \mathcal{G}$ with a sampled competitor $y_i^- \in \mathcal{Y} \setminus \mathcal{G}$ for the same input x. Metric learning encourages the acceptable hypothesis to receive a higher score than its competitor. Writing their scores as $s^+ = F_\theta(x, y^+)$ and $s_i^- = F_\theta(x, y_i^-)$, the ranking violation of comparison i is commonly expressed through the score difference $d_i = s_i^- - s^+$, which we refer to as the comparison deficit. A larger deficit indicates a harder comparison, with $d_i > 0$ corresponding to a ranking violation. Reducing this deficit induces the parameter-space descent direction $-\nabla_\theta d_i = \nabla_\theta s^+ - \nabla_\theta s_i^-$.

Geometrically, d_i specifies a sampled relative-height correction, while $-\nabla_\theta d_i$ specifies the parameter-space direction through which this correction deforms the score landscape. Its task effect therefore should be evaluated against the boundary-derived gradient $\nabla_\theta W_\theta$.

Definition 1 (Boundary-derived task utility)**.** *Under the preceding regularity conditions, the task utility of comparison i for winning-region expansion is its first-order contribution to the winning-region measure along the induced direction $-\nabla_\theta d_i$:*

$$u_i^{\exp} := \langle \nabla_\theta W_\theta,\, -\nabla_\theta d_i \rangle = \int_{\Gamma_\theta} \frac{\langle \nabla_\theta m_\theta(y),\, -\nabla_\theta d_i \rangle}{\|\nabla_y m_\theta(y)\|}\, \rho(y)\, dS(y). \tag{6}$$

The boundary form explains how improving a sampled score difference translates into regional gain. At each boundary point, the comparison-induced margin response $\langle \nabla_\theta m_\theta(y), -\nabla_\theta d_i \rangle$ drives a normal displacement; weighting by $\rho(y)/\|\nabla_y m_\theta(y)\|$ and integrating gives the net contribution to winning-region expansion. For a step $\delta\theta_i = -\eta \nabla_\theta d_i$ with $\eta > 0$, this contribution determines $W_{\theta+\delta\theta_i} - W_\theta = \eta u_i^{\exp} + o(\eta)$. Positive utility therefore indicates net expansion to first order, while negative utility indicates contraction.

Loss as comparison allocation. A differentiable metric loss L acting on the sampled deficits assigns each comparison a gradient weight $w_i = \partial L / \partial d_i$, producing the descent direction $-\nabla_\theta L = \sum_i w_i(-\nabla_\theta d_i)$. The loss thus constructs its update by allocating weight across comparison-induced directions, whose task effects are measured by the boundary-derived utilities.

To distinguish the effect of the total comparison weight from that of its allocation, suppose $w_i \geq 0$ and let $\lambda = \sum_i w_i > 0$. The normalized weights $q_i = w_i/\lambda$, satisfying $\sum_i q_i = 1$, describe how this total weight is distributed. For a gradient step $\delta\theta = -\eta \nabla_\theta L$ with step size $\eta > 0$, the first-order task response is

$$W_{\theta+\delta\theta} - W_\theta = \eta\lambda \sum_i q_i\, u_i^{\exp} + o(\eta). \tag{7}$$

A loss can therefore be decomposed into how much comparison weight it supplies and how that weight is allocated across the available comparison directions: λ specifies the total comparison weight, while q_i specifies its normalized allocation; the resulting $\sum_i q_i u_i$ is the allocation-weighted average task utility. Task utility thus provides the task-side reference for evaluating how effectively a loss distributes its training signal.

4.3 Boundary Contraction and the Formation Limit

Height dominance in the vanishing-region limit. The preceding analysis provides a task direction and utility for an existing winning region with $W_\theta > 0$. When the region is empty, its measure is locally constant away from the formation threshold and cannot distinguish updates that promote its emergence. The limit $W_\theta \to 0^+$ provides a geometric reference for extending the analysis to this missing regime.

To characterize this limit, let $m_\theta^\star = \sup_{y \in \mathcal{G}} m_\theta(y)$ denote the maximum competitive margin, or *competitive height*. Under the isolated-birth conditions below, $m_\theta^\star > 0$ corresponds to a positive winning-region measure, while $m_\theta^\star \leq 0$ corresponds to an empty region; crossing zero marks *region formation*. Suppose the region approaches an isolated birth at a unique, nondegenerate interior maximizer p, with $H_\theta := -\nabla_y^2 m_\theta(p) \succ 0$. Under the regularity conditions detailed in the appendix, the two-dimensional local geometry gives

$$W_\theta \approx \mathcal{S}_\theta m_\theta^\star, \qquad \mathcal{S}_\theta = \frac{2\pi\rho(p)}{\sqrt{\det H_\theta}}, \tag{8}$$

where $\mathcal{S}_\theta$ describes how local curvature and reference density convert competitive height into regional extent. The corresponding parameter response is

$$\nabla_\theta W_\theta \approx \mathcal{S}_\theta \nabla_\theta m_\theta^\star + m_\theta^\star \nabla_\theta \mathcal{S}_\theta. \tag{9}$$

The first term is the *height channel*, describing growth through competitive-height variation under the current shape; the second is the *shape channel*, describing growth through changes in the geometry supporting that height. For example, reduced curvature can support a broader winning region at the same competitive height. As $m_\theta^\star \to 0^+$, and hence $W_\theta \to 0^+$, the shape contribution becomes negligible relative to the height contribution, provided the shape factor remains positive and regular and the competitive-height gradient does not vanish. Consequently, $\nabla_\theta W_\theta \approx \mathcal{S}_\theta \nabla_\theta m_\theta^\star$: near birth, the task-optimal expansion direction is asymptotically aligned with the competitive-height direction. The local asymptotic derivation and its extension to general dimension are given in Appendix D.

Height dominance identifies the limiting comparison. The dominant height channel has a direct comparison interpretation. Let $h = \arg\max_{z \in \mathcal{Y} \setminus \mathcal{G}} F_\theta(z)$ be the strongest unacceptable competitor. Since p is the best acceptable hypothesis, $m_\theta^\star = F_\theta(p) - F_\theta(h)$. When these optimizers are locally unique and the optimal values are differentiable,

$$\nabla_\theta m_\theta^\star = \nabla_\theta F_\theta(p) - \nabla_\theta F_\theta(h) = -\nabla_\theta d_h, \qquad d_h := F_\theta(h) - F_\theta(p). \tag{10}$$

Here, the score gradients are evaluated at fixed hypotheses. Thus, the direction governing regional growth near birth is exactly the descent direction of the comparison between the best acceptable hypothesis and its strongest unacceptable competitor. For the fixed acceptable endpoint p, h is the task-side hardest negative. Its relevance follows from height dominance: improving this pair increases the competitive height that controls leading-order regional growth.

Extension to the empty-region regime. Under continuous differentiability and a nonvanishing competitive-height gradient near the threshold, the limiting direction extends continuously to the empty-region side. When $m_\theta^\star < 0$, W_θ is locally zero, and Equation (9) shows that the task-side hardest comparison provides the steepest first-order increase in competitive height, thereby driving the landscape toward region formation. We therefore identify $\nabla_\theta m_\theta^\star$ as the task-optimal formation direction throughout the empty-region regime.

By the birth-limit relation established above, this formation direction is also the height-dominated limit of the task-optimal expansion direction. Accordingly, for bounded comparison gradients, $u_i^{\mathrm{exp}} \approx \mathcal{S}_\theta u_i^{\mathrm{form}}$ as $m_\theta^\star \to 0^+$. Beyond this limit, expansion additionally depends on the shape channel in Equation (10), so expansion utility reflects both competitive-height and shape variation of an existing region. The envelope identity and the formation–expansion birth-limit analysis, including its general-dimensional extension, are given in Appendices C and D, respectively.

4.4 Coarea Representation of Task Utility

The remaining structural difference lies in the geometric factor $\|\nabla_y m_\theta(y)\|^{-1}$, which has no explicit counterpart in the pointwise formation utility. Can this factor be absorbed into a representation that

places formation and expansion on the same first-order footing? The Coarea formula provides such a representation.

Lemma 2 (Coarea representation of expansion utility)**.** *Under the regularity conditions of the zero-margin level set, the boundary-defined expansion utility admits the equivalent narrow-band representation*

$$u_i^{\exp} = \lim_{b\to 0^+} \frac{1}{2b} \int_{|m_\theta(y)|<b} \rho(y)\, \langle \nabla_\theta m_\theta(y), -\nabla_\theta d_i \rangle \, dy. \tag{11}$$

The result follows directly from the Coarea formula; the derivation and regularity conditions are given in Appendix E. The factor $\|\nabla_y m_\theta(y)\|^{-1}$ in the boundary form is not discarded by this transformation. Instead, it is absorbed into the amount of hypothesis-space volume contained in the shrinking margin band: flatter portions of the margin field occupy more state-space volume and therefore contribute proportionally more to the integral.

Formal unification of formation and expansion. Crucially, the Coarea representation separates the two ingredients of expansion utility in the same way as formation utility:

$$u_i^{(r)} = \left\langle g_\theta^{(r)}, -\nabla_\theta d_i \right\rangle, \qquad r \in \{\text{form}, \exp\}, \tag{12}$$

with

$$g_\theta^{\text{form}} = \nabla_\theta m_\theta^\star, \qquad g_\theta^{\exp} = \lim_{b\to 0^+} \frac{1}{2b} \int_{|m_\theta(y)|<b} \rho(y) \nabla_\theta m_\theta(y)\, dy. \tag{13}$$

Here, $g_\theta^{(r)}$ is the task-side reference direction in regime r. Formation uses the pointwise competitive-height sensitivity at the maximal-margin hypothesis, whereas expansion uses its distributed counterpart over the zero-margin neighborhood. In both regimes, the same metric-learning comparison direction, $-\nabla_\theta d_i$, is evaluated by its alignment with the corresponding task-side direction.

Discrete computation. For a discretely sampled hypothesis space, the Coarea representation yields the finite-band estimator

$$\widehat{g}_\theta^{\exp} = \frac{\Delta V}{2b} \sum_{|m_\theta(y_k)|<b} \rho(y_k) \nabla_\theta m_\theta(y_k), \qquad \widehat{u}_i^{\exp} = \langle \widehat{g}_\theta^{\exp}, -\nabla_\theta d_i \rangle, \tag{14}$$

where b is the finite margin bandwidth and ΔV is the state-space volume represented by each hypothesis under uniform discretization. For nonuniform sampling, ΔV is replaced by the corresponding quadrature weight. This estimator avoids explicit reconstruction of Γ_θ and direct evaluation of $\|\nabla_y m_\theta\|^{-1}$. The latter is represented implicitly by the occupancy of the margin band: flatter portions of the margin field occupy more state-space volume and therefore contribute more samples to the aggregation.

5 EXPERIMENTS

In this section, we empirically evaluate the theoretical analysis and its implications for task performance. Details of the datasets and experimental settings are provided in Appendix F.1.

Validation of the task gradient. We first test whether the expansion direction $\nabla_\theta W_\theta$ predicts finite changes in the winning-region measure. For updates with step size η along this direction and its opposite, let W^+ and W^- denote the recomputed measures and W^0 the measure before the update. Table 2 reports the percentage of queries satisfying the weak ordering $W^+ \geq W^0 \geq W^-$ and its strict counterpart. As the step size η increases, strict ordering becomes more pronounced, while weak ordering remains high despite occasional deviations, consistent with the local nature of the first-order prediction.

Table 2: Directional consistency under finite-step updates. Values report the percentage of queries satisfying each ordering criterion.

		Step size η		
Dataset	Order	10^{-4}	10^{-3}	10^{-2}
Zurich	Weak	100.0	100.0	100.0
	Strict	0.0	13.0	87.0
Oxford	Weak	100.0	99.7	95.0
	Strict	0.2	15.2	80.6

Table 3: Median relative change in winning-region measure (%) after multi-step updates.

Dataset	Task gradient	Utility aligned	Hardness aligned	Utility reversed	Opposite task gradient
Zurich	**+18.5**	+12.2	+7.5	−0.6	−18.6
Oxford	**+38.7**	+25.9	+16.4	−1.8	−51.6

Validation of comparison allocation. Equation (7) predicts that, for fixed comparisons and total weight, the task effect depends on how the normalized allocation q_i is distributed across comparison utilities. We test this by reassigning a fixed InfoNCE (van den Oord et al., 2018) weight spectrum according to task utility u_i, hardness d_i, or reversed utility, with the task gradient and its opposite as references. Extending Table 2, we apply five updates with $\eta = 10^{-3}$. Table 3 shows that utility-aligned allocation consistently produces larger increases in the winning-region measure than hardness-aligned allocation, whereas reversing the utility ordering largely removes or reverses the gain. This directly verifies the allocation term $\sum_i q_i u_i^{\exp}$ in Eq. (7). Figure 2 further shows that this ordering persists over a longer 25-step trajectory.

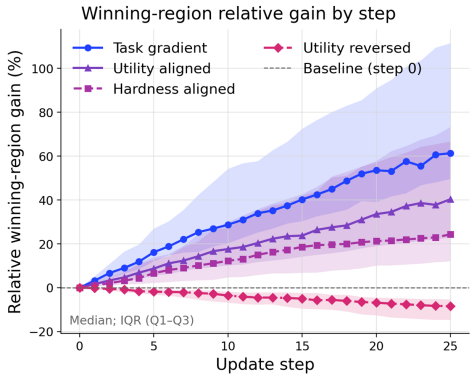


Figure 2: Long-horizon winning-region evolution under different weight allocations.

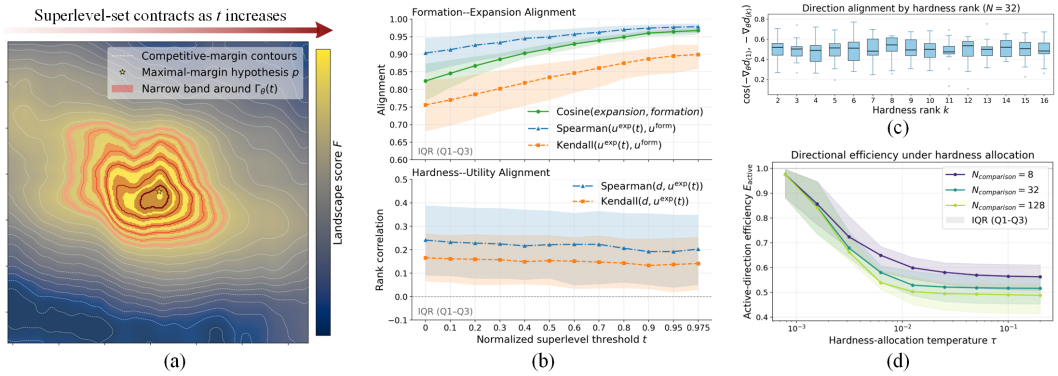


Figure 3: Formation–expansion continuity and the directional limitation of hardness. Superlevel contraction recovers the formation regime, while hardness remains insufficient to characterize comparison utility and its aggregation. Shaded bands denote the IQR (Q_1–Q_3). In (c), k denotes hardness rank, with $k = 1$ the strongest comparison. In (d), N denotes the comparison-set size.

Boundary contraction and the formation limit. Section 4.3 predicts that, as a winning region approaches formation, its expansion direction becomes height-dominated and aligns with $\nabla_\theta m_\theta^\star$. We probe this limiting geometry at fixed θ by evaluating the boundary aggregation on normalized superlevel sets $\Gamma_\theta(t) = \{y \in \operatorname{int} \mathcal{G} : m_\theta(y) = t\, m_\theta^\star\}$. Increasing t contracts the boundary toward the maximal-margin hypothesis p (Fig. 3(a)), thereby progressively localizing the expansion response around the competitive-height maximum. Consistent with the height-dominance prediction, its alignment with the formation direction $\nabla_\theta m_\theta^\star$ increases under contraction (Fig. 3(b), top). Equation (10) further identifies the limiting formation direction with the strongest comparison, $\nabla_\theta m_\theta^\star = -\nabla_\theta d_h$. However, this task relevance does not extend through the hardness ranking: hardness–utility rank agreement remains low even as expansion approaches formation (Fig. 3(b), bottom). Figure 3(c) explains why: top-hard comparisons remain directionally heterogeneous with respect to $-\nabla_\theta d_h$. Thus, proximity in scalar hardness does not imply proximity in the task-relevant parameter direction. This directional heterogeneity becomes relevant once a loss

distributes weight over multiple comparisons. To quantify how much of the allocated directional mass remains aligned with the formation reference $v_h = -\nabla_\theta d_h$, we define the active-direction efficiency $E_{\text{active}} := (\sum_i q_i \|v_i\|)^{-1} \sum_i q_i \frac{\langle v_h, v_i \rangle}{\|v_h\|}$, $v_i = -\nabla_\theta d_i$. It is the $q_i\|v_i\|$-weighted average directional alignment with the hardest comparison, and equals one for hardest-only allocation. For InfoNCE, increasing the temperature τ broadens the allocation beyond the strongest comparison. Because these additional directions are heterogeneous with respect to v_h, E_{active} decreases as the allocation broadens (Fig. 3(d)), indicating that temperature controls allocation concentration relative to both the available set size and its hardness distribution.

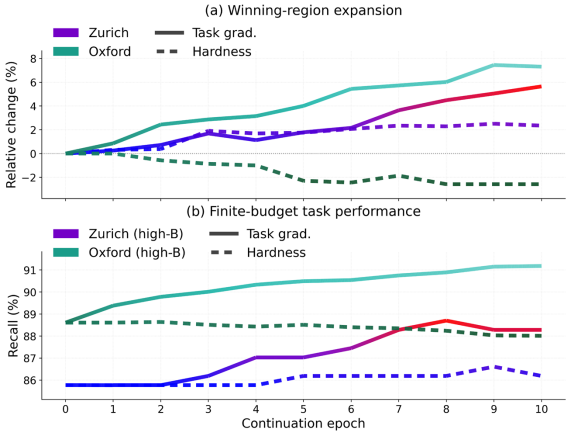
Realizing task tolerance under finite candidate budgets. Finally, we test whether task-directed winning-region expansion transfers to finite-budget task performance. Both branches resume from the same checkpoint, selected after training performance has saturated. The hardness-aligned branch uses InfoNCE, with update magnitudes normalized to remove gradient-scale effects; the task-gradient branch is normalized in the same way. Train/test regions are disjoint to control for regional generalization. As shown in Table 4 and Fig. 4, ΔW and ΔR@1 are consistent in sign and broadly track each other in magnitude. Hardness-aligned InfoNCE yields only small gains in both quantities on Zurich and decreases in both on Oxford, whereas the task-gradient branch produces substantially larger improvements, especially on Oxford and under the lower candidate budget.

Table 4: Winning-region growth and finite-budget recall after continuation training.

Metric (%)	Baseline	Hardness	Task
Zurich			
Mean ΔW	—	+2.35	**+5.65**
R@1, higher B	85.77	86.19	**88.28**
R@1, lower B	64.12	64.44	**67.78**
Oxford			
Mean ΔW	—	−2.65	**+7.22**
R@1, higher B	88.61	88.01	**91.18**
R@1, lower B	83.54	83.17	**88.01**

ΔW and ΔR@1 agree in sign and broadly in magnitude: task-gradient gains are larger, whereas hardness-aligned InfoNCE is small on Zurich and negative on Oxford.

Figure 4: Winning-region and higher-budget recall trajectories.



6 Conclusion

This work develops a geometric analysis of metric learning under task-defined tolerance. We characterize task-valid behavior through the winning-region measure and derive task-optimal reference directions for region expansion and formation. The Coarea representation further places the two regimes on a common first-order footing, providing a task-side reference for evaluating metric-learning comparison directions. Our experiments validate the main predictions of the analysis: task-utility alignment consistently predicts the direction of winning-region change, and the resulting geometric responses agree with the theoretical characterization of region formation and expansion. They further confirm that the geometry of winning-region is meaningfully related to how well the learned score landscape realizes task-defined tolerance under practical inference approximation. Overall, our analysis establishes a geometric connection between task-defined tolerance and metric learning, linking task-level objectives to the parameter-space directions through which metric losses shape the score landscape.

AI use statement

The drafting and linguistic refinement of portions of this paper, as well as the formatting of LaTeX and the references, were assisted by generative artificial intelligence tools. All content generated with

the assistance of artificial intelligence has been reviewed and revised by the authors. The authors have verified the mathematical proofs, experimental data, cited references, and scientific conclusions by cross-referencing their derivation processes, code, experimental results, and cited literature, and assume full responsibility for the final content of this paper.

REFERENCES

Jeremy Cohen, Elan Rosenfeld, and Zico Kolter. Certified adversarial robustness via randomized smoothing. In *international conference on machine learning*, pp. 1310–1320. PMLR, 2019.

Michel C Delfour and J-P Zolésio. *Shapes and geometries: metrics, analysis, differential calculus, and optimization*. SIAM, 2011.

Fabian Deuser, Konrad Habel, and Norbert Oswald. Sample4Geo: Hard negative sampling for cross-view geo-localisation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 16847–16856, 2023.

Pete Florence, Corey Lynch, Andy Zeng, Oscar A Ramirez, Ayzaan Wahid, Laura Downs, Adrian Wong, Johnny Lee, Igor Mordatch, and Jonathan Tompson. Implicit behavioral cloning. In *Conference on robot learning*, pp. 158–168. PMLR, 2022.

Raia Hadsell, Sumit Chopra, and Yann LeCun. Dimensionality reduction by learning an invariant mapping. *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*, 2:1735–1742, 2006. URL https://api.semanticscholar.org/CorpusID:8281592.

Chuying Han, Yasong Feng, and Tianyu Wang. From random search to bandit learning in metric measure spaces. *arXiv preprint arXiv:2305.11509*, 2023. doi: 10.48550/arXiv.2305.11509.

Yann LeCun, Sumit Chopra, Raia Hadsell, Marc'Aurelio Ranzato, and Fu-Jie Huang. A tutorial on energy-based learning. In Gökhan H. Bakir, Thomas Hofmann, Bernhard Schölkopf, Alexander J. Smola, and Ben Taskar (eds.), *Predicting Structured Data*. MIT Press, 2006.

William P. Maddern, Geoffrey Pascoe, Chris Linegar, and Paul Newman. 1 year, 1000 km: The oxford robotcar dataset. *The International Journal of Robotics Research*, 36:15 – 3, 2017. URL https://api.semanticscholar.org/CorpusID:22556995.

Aleksander Madry, Aleksandar Makelov, Ludwig Schmidt, Dimitris Tsipras, and Adrian Vladu. Towards deep learning models resistant to adversarial attacks. *arXiv preprint arXiv:1706.06083*, 2017.

R. Manmatha, Chaoxia Wu, Alex Smola, and Philipp Krähenbühl. Sampling matters in deep embedding learning. *2017 IEEE International Conference on Computer Vision (ICCV)*, pp. 2859–2867, 2017. URL https://api.semanticscholar.org/CorpusID:24718057.

Kieran Murphy, Carlos Esteves, Varun Jampani, Srikumar Ramalingam, and Ameesh Makadia. Implicit-pdf: Non-parametric representation of probability distributions on the rotation manifold. *arXiv preprint arXiv:2106.05965*, 2021.

Stanley Osher and James A Sethian. Fronts propagating with curvature-dependent speed: Algorithms based on hamilton-jacobi formulations. *Journal of computational physics*, 79(1):12–49, 1988.

David E. Rumelhart, Geoffrey E. Hinton, and Ronald J. Williams. Learning representations by back-propagating errors. *Nature*, 323:533–536, 1986. URL https://api.semanticscholar.org/CorpusID:205001834.

Torsten Sattler, William P. Maddern, Carl Toft, Akihiko Torii, Lars Hammarstrand, Erik Stenborg, Daniel Safari, M. Okutomi, Marc Pollefeys, Josef Sivic, Fredrik Kahl, and Tomás Pajdla. Benchmarking 6dof outdoor visual localization in changing conditions. *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 8601–8610, 2018. URL https://api.semanticscholar.org/CorpusID:4644084.

Florian Schroff, Dmitry Kalenichenko, and James Philbin. Facenet: A unified embedding for face recognition and clustering. *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 815–823, 2015. URL https://api.semanticscholar.org/CorpusID:206592766.

Kihyuk Sohn. Improved deep metric learning with multi-class n-pair loss objective. In *Neural Information Processing Systems*, 2016. URL https://api.semanticscholar.org/CorpusID:911406.

Jan Sokołowski and Jean-Paul Zolésio. Introduction to shape optimization : shape sensitivity analysis. 1992. URL https://api.semanticscholar.org/CorpusID:118509233.

Hyun Oh Song, Yu Xiang, Stefanie Jegelka, and Silvio Savarese. Deep metric learning via lifted structured feature embedding. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 4004–4012, 2015. URL https://api.semanticscholar.org/CorpusID:5726681.

Marko Subasic. Level set methods and fast marching methods. 2003. URL https://api.semanticscholar.org/CorpusID:62769056.

Yifan Sun, Changmao Cheng, Yuhan Zhang, Chi Zhang, Liang Zheng, Zhongdao Wang, and Yichen Wei. Circle loss: A unified perspective of pair similarity optimization. *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 6397–6406, 2020. URL https://api.semanticscholar.org/CorpusID:211296865.

Ben Taskar, Carlos Guestrin, and Daphne Koller. Max-margin markov networks. In *Neural Information Processing Systems*, 2003. URL https://api.semanticscholar.org/CorpusID:201720.

Dimitris Tsipras, Shibani Santurkar, Logan Engstrom, Alexander Turner, and Aleksander Madry. Robustness may be at odds with accuracy. *arXiv preprint arXiv:1805.12152*, 2018.

Ioannis Tsochantaridis, Thorsten Joachims, Thomas Hofmann, Yasemin Altun, and Yoram Singer. Large margin methods for structured and interdependent output variables. *Journal of machine learning research*, 6(9), 2005.

Aäron van den Oord, Yazhe Li, and Oriol Vinyals. Representation learning with contrastive predictive coding. *ArXiv*, abs/1807.03748, 2018. URL https://api.semanticscholar.org/CorpusID:49670925.

Xun Wang, Xintong Han, Weilin Huang, Dengke Dong, and Matthew R. Scott. Multi-similarity loss with general pair weighting for deep metric learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 5022–5030, 2019.

SUPPLEMENTARY MATERIAL

A NOTATION AND REGULARITY CONDITIONS

A.1 MAIN NOTATION

For completeness, Table 5 summarizes the principal task-side and loss-side notation used throughout the paper. Proof-local auxiliary quantities are defined separately at their first occurrence.

A.2 REGULARITY CONDITIONS AND SCOPE

We state explicitly the local conditions under which the ordinary gradients in the main text are well defined. In a neighborhood of the competitive boundary, the hypothesis space is represented by a finite-dimensional Euclidean chart of dimension D, the acceptable set $\mathcal{G}$ is fixed with respect to θ, and the reference measure has density ρ with respect to the local volume element. The density is

Table 5: Notation for the main task- and loss-side objects.

Symbol	Term	Meaning / definition
$\mathcal{Y}$	Hypothesis space	Space of all candidate hypotheses.
$\mathcal{G} \subseteq \mathcal{Y}$	Acceptable set	Task-defined subset of the hypothesis space satisfying the prescribed task tolerances, independent of the current model state.
$F_\theta(y)$	Score landscape	Score assigned by the current model to hypothesis y.
$m_\theta(y)$	Competitive margin	Score advantage of $y \in \mathcal{G}$ over the strongest unacceptable alternative: $m_\theta(y) = F_\theta(y) - \sup_{z \in \mathcal{Y} \setminus \mathcal{G}} F_\theta(z)$.
$\mathcal{R}_\theta$	Winning region	Subset of the acceptable set with positive competitive margin: $\mathcal{R}_\theta = \{y \in \mathcal{G} : m_\theta(y) > 0\}$.
W_θ	Winning-region measure	Measure of the winning region under the reference measure μ: $W_\theta = \mu(\mathcal{R}_\theta)$.
Γ_θ	Competitive boundary	Zero-margin boundary of the winning region: $\Gamma_\theta = \{y \in \operatorname{int} \mathcal{G} : m_\theta(y) = 0\}$.
$m_\theta^\star$	Maximum competitive margin	Maximum competitive margin within the acceptable set: $m_\theta^\star = \sup_{y \in \mathcal{G}} m_\theta(y)$.
$\mathcal{S}_\theta$	Local shape factor	Local geometric factor that converts competitive height into winning-region extent near region birth, as in $W_\theta \approx \mathcal{S}_\theta m_\theta^\star$ for a two-dimensional hypothesis landscape.
d_i	Comparison deficit / hardness	Violation associated with comparison i; for example, $d_i = s_i^- - s^+$.
$-\nabla_\theta d_i$	Comparison-induced direction	Parameter direction induced by reducing the deficit of comparison i.
u_i	Task utility	First-order effect of the comparison-induced direction on the task-relevant landscape functional: $u_i = \langle \nabla_\theta J_\theta, -\nabla_\theta d_i \rangle$.
w_i	Comparison gradient weight	Gradient weight assigned by the loss to comparison i: $w_i = \partial L / \partial d_i$.
λ	Total gradient weight	Sum of comparison-gradient weights: $\lambda = \sum_i w_i$.
q_i	Normalized gradient allocation	Normalized allocation of gradient weight: $q_i = w_i / \lambda$.

continuous and nonnegative. The competitive margin is continuously differentiable in (θ, y) in the first-variation and coarea arguments below.

The supremum over unacceptable hypotheses is assumed to be attained by a locally unique, stable branch h_θ. Consequently,

$$m_\theta(y) = F_\theta(y) - F_\theta(h_\theta)$$

is ordinarily differentiable with respect to θ along the branch. If the strongest unacceptable hypothesis is tied or switches under an infinitesimal perturbation, the corresponding statements hold only directionally, with the active-set version of Danskin's theorem; a single classical gradient need not exist at such a switching point.

For expansion, the free competitive boundary

$$\Gamma_\theta = \{y \in \operatorname{int} \mathcal{G} : m_\theta(y) = 0\}$$

is assumed compact and regular: $\nabla_y m_\theta(y) \neq 0$ on Γ_θ. Any portion of $\partial\mathcal{G}$ is task-fixed and therefore has zero normal velocity. The formulas below concern the moving free boundary; possible corners where it meets $\partial\mathcal{G}$ are excluded from the local argument (or, equivalently, are assumed to have zero surface measure).

To keep the correspondence with the main text explicit, p_θ and h_θ below are the same objects denoted there by p and h; the subscript only exposes their parameter dependence inside the proofs. Symbols such as A_θ, e_θ, ϕ_i, and G_i are proof-local auxiliary quantities and are defined at their first occurrence.

B First Variation of Winning-Region Measure

Proposition 2 (Boundary displacement and mass variation)**.** *Let $\delta\theta$ be a small parameter perturbation. At a fixed hypothesis y, denote the induced first-order margin change by*

$$\delta m_\theta(y) := \langle \nabla_\theta m_\theta(y), \delta\theta \rangle.$$

For $y \in \Gamma_\theta$, let $\delta r(y)$ be the signed first-order displacement of the free boundary along $n_{\mathrm{out}}(y)$. Under the conditions in Appendix A,

$$\delta r(y) = \frac{\delta m_\theta(y)}{\|\nabla_y m_\theta(y)\|}, \tag{15}$$

$$\delta W_\theta := \int_{\Gamma_\theta} \frac{\delta m_\theta(y)}{\|\nabla_y m_\theta(y)\|} \rho(y)\, dS(y). \tag{16}$$

Here δW_θ denotes the first-order part of $W_{\theta+\delta\theta} - W_\theta$. In particular,

$$\nabla_\theta W_\theta = \int_{\Gamma_\theta} \frac{\nabla_\theta m_\theta(y)}{\|\nabla_y m_\theta(y)\|} \rho(y)\, dS(y). \tag{17}$$

Proof. The winning side of the level set is $m_\theta > 0$, so its outward unit normal is

$$n_{\mathrm{out}}(y) = -\frac{\nabla_y m_\theta(y)}{\|\nabla_y m_\theta(y)\|}.$$

Regularity of the zero level set permits a local normal parametrization of the perturbed boundary. Thus a point $y \in \Gamma_\theta$ has a corresponding point

$$y' = y + \delta r(y) n_{\mathrm{out}}(y) + o(\|\delta\theta\|)$$

on the perturbed boundary, where the implicit-function theorem gives $\delta r(y) = O(\|\delta\theta\|)$. The identity $m_{\theta+\delta\theta}(y') = 0$ and a joint first-order expansion give

$$\begin{aligned} 0 &= m_\theta(y) + \delta m_\theta(y) + \delta r(y) \nabla_y m_\theta(y)^\top n_{\mathrm{out}}(y) + o(\|\delta\theta\|) \\ &= \delta m_\theta(y) - \delta r(y) \|\nabla_y m_\theta(y)\| + o(\|\delta\theta\|), \end{aligned}$$

because $m_\theta(y) = 0$ and $\nabla_y m_\theta(y)^\top n_{\mathrm{out}}(y) = -\|\nabla_y m_\theta(y)\|$. Equating the first-order terms proves Eq. (15). Its sign is consistent with the convention: a positive margin perturbation moves the zero set outward.

The Hadamard first-variation formula for a domain with fixed density gives

$$\delta W_\theta = \int_{\Gamma_\theta} \rho(y)\delta r(y)\, dS(y).$$

The task boundary $\partial\mathcal{G}$ contributes nothing because it is fixed. Substitution of Eq. (15) proves Eq. (16). Finally, substituting $\delta m_\theta(y) = \langle \nabla_\theta m_\theta(y), \delta\theta\rangle$ and moving the common $\delta\theta$ outside the integral yields Eq. (17). □

Equivalently, including the higher-order remainder,

$$W_{\theta+\delta\theta} - W_\theta = \langle \nabla_\theta W_\theta, \delta\theta\rangle + o(\|\delta\theta\|),$$

which is the precise differential interpretation of Eqs. (3)–(5).

Corollary 2 (Comparison and loss responses)**.** *For the comparison-induced step $\delta\theta_i = -\eta\nabla_\theta d_i$,*

$$W_{\theta+\delta\theta_i} - W_\theta = \eta u_i^{\exp} + o(\eta), \qquad u_i^{\exp} = \langle \nabla_\theta W_\theta, -\nabla_\theta d_i\rangle .$$

If $L = L(d_1, \ldots, d_N)$ is differentiable and $w_i = \partial L/\partial d_i$, then the gradient step $\delta\theta = -\eta\nabla_\theta L$ satisfies

$$W_{\theta+\delta\theta} - W_\theta = \eta \sum_{i=1}^{N} w_i u_i^{\exp} + o(\eta).$$

Proof. The first identity follows by substituting $\delta\theta_i = -\eta\nabla_\theta d_i$ into the first-order expansion above. The chain rule gives $-\nabla_\theta L = \sum_i w_i(-\nabla_\theta d_i)$; linearity of the differential then proves the second identity. □

C Formation Direction Before Region Birth

Proposition 3 (Envelope identity and the critical comparison)**.** *Suppose the maximum*

$$m_\theta^\star = \sup_{y\in\mathcal{G}} m_\theta(y)$$

is attained, throughout a parameter neighborhood, by a unique active maximizer $p_\theta \in \operatorname{int}\mathcal{G}$. Assume that p_θ is a C^1 parameter branch and that m_θ is continuously differentiable near (θ, p_θ). (For example, the branch condition follows locally from the implicit-function theorem when the interior maximizer is nondegenerate and remains the unique global maximizer.) Then

$$\nabla_\theta m_\theta^\star = \nabla_\theta m_\theta(p_\theta). \tag{18}$$

If h_θ is the locally unique, stable strongest unacceptable hypothesis whose optimal-value branch is differentiable as in Appendix A, and

$$d_h := F_\theta(h_\theta) - F_\theta(p_\theta),$$

then

$$-\nabla_\theta d_h = \nabla_\theta m_\theta^\star. \tag{19}$$

Proof. The interior first-order condition is $\nabla_y m_\theta(p_\theta) = 0$. Differentiating $m_\theta^\star = m_\theta(p_\theta)$ gives

$$\nabla_\theta m_\theta^\star = \partial_\theta m_\theta(p_\theta) + (D_\theta p_\theta)^\top \nabla_y m_\theta(p_\theta) = \partial_\theta m_\theta(p_\theta),$$

which proves Eq. (18). This is the local envelope identity. Moreover, stability of the two active branches gives the exact scalar identity

$$d_h = F_\theta(h_\theta) - F_\theta(p_\theta) = -m_\theta(p_\theta) = -m_\theta^\star$$

throughout the local branch. Differentiating it proves Eq. (19). □

When $m_\theta^\star < 0$, the strict superlevel set defining $\mathcal{R}_\theta$ is empty; whenever this strict negative gap persists under small parameter perturbations, W_θ is locally constant at zero. At the threshold $m_\theta^\star = 0$, the strict superlevel set is still empty, but a favorable infinitesimal perturbation may create a positive-mass component, so an ordinary gradient of W_θ need not exist there. The formation quantity $m_\theta^\star$ therefore supplies the relevant smooth first-order signal on the pre-birth branch and up to the zero crossing. Increasing it moves the best acceptable hypothesis toward region birth. Equation (19) also clarifies that the relevant comparison is the task-side pair (p_θ, h_θ); it need not coincide with the hardest member of a particular sampled training batch.

D FORMATION–EXPANSION CONNECTION NEAR REGION BIRTH

We now consider the one-sided regime $m_\theta^\star \to 0^+$. In addition to the conditions above, assume that m_θ is C^3 in y near p_θ, that

$$H_\theta := -\nabla_y^2 m_\theta(p_\theta) \succ 0,$$

and that these bounds are uniform along the local parameter branch. More precisely, assume that there are constants $r_0, \kappa > 0$, independent of θ sufficiently near birth, such that the ball $B(p_\theta, r_0)$ lies inside the common Taylor neighborhood and

$$\sup_{y \in \mathcal{G} \setminus B(p_\theta, r_0)} m_\theta(y) \le -\kappa.$$

Thus the winning region and its zero-level boundary near birth are exactly the isolated component born at p_θ; no remote zero-level component contributes to W_θ. Finally, assume $\rho(p_\theta) > 0$ and that $y \mapsto \nabla_\theta m_\theta(y)$ is locally uniformly continuous in y, uniformly with respect to θ along the local parameter branch near birth.

Lemma 3 (Collapse of the emerging boundary)**.** *Under these assumptions,*

$$\sup_{y \in \Gamma_\theta} \|y - p_\theta\| = O\Big(\sqrt{m_\theta^\star}\Big) \longrightarrow 0. \tag{20}$$

Proof. Writing $y = p_\theta + \Delta y$, Taylor's theorem gives

$$m_\theta(p_\theta + \Delta y) = m_\theta^\star - \frac{1}{2}\Delta y^\top H_\theta \Delta y + R_\theta(\Delta y), \qquad |R_\theta(\Delta y)| \le C\|\Delta y\|^3.$$

Uniform positive definiteness supplies a constant $\lambda_0 > 0$ with $\lambda_{\min}(H_\theta) \ge \lambda_0$. In a sufficiently small uniform neighborhood, the remainder is bounded by $\lambda_0\|\Delta y\|^2/4$, and hence

$$m_\theta(p_\theta + \Delta y) \le m_\theta^\star - \frac{\lambda_0}{4}\|\Delta y\|^2.$$

The uniform gap outside $B(p_\theta, r_0)$ ensures that, for sufficiently small $m_\theta^\star > 0$, the entire zero-level boundary lies in this neighborhood. Setting $m_\theta(y) = 0$ therefore yields $\|y - p_\theta\|^2 \le 4m_\theta^\star/\lambda_0$, proving the claim. □

Proposition 4 (Directional connection at emergence)**.** *For $m_\theta^\star > 0$, define*

$$A_\theta := \int_{\Gamma_\theta} \frac{\rho(y)}{\|\nabla_y m_\theta(y)\|}\, dS(y) > 0.$$

Then

$$\nabla_\theta W_\theta = A_\theta\big(\nabla_\theta m_\theta^\star + e_\theta\big), \qquad \|e_\theta\| \le \sup_{y \in \Gamma_\theta} \|\nabla_\theta m_\theta(y) - \nabla_\theta m_\theta(p_\theta)\| \longrightarrow 0. \tag{21}$$

If, in addition,

$$\frac{\|e_\theta\|}{\|\nabla_\theta m_\theta^\star\|} \longrightarrow 0, \tag{22}$$

then

$$\frac{\nabla_\theta W_\theta}{\|\nabla_\theta W_\theta\|} \longrightarrow \frac{\nabla_\theta m_\theta^\star}{\|\nabla_\theta m_\theta^\star\|}, \qquad m_\theta^\star \to 0^+. \tag{23}$$

A sufficient condition for Eq. (22) *is $\|\nabla_\theta m_\theta^\star\| \ge c > 0$ along the branch.*

Proof. By Proposition 2,

$$\nabla_\theta W_\theta = \int_{\Gamma_\theta} a_\theta(y) \nabla_\theta m_\theta(y)\, dS(y), \qquad a_\theta(y) := \frac{\rho(y)}{\|\nabla_y m_\theta(y)\|} > 0.$$

Using Eq. (18), add and subtract $\nabla_\theta m_\theta(p_\theta) = \nabla_\theta m_\theta^\star$ inside the integral. This gives Eq. (21) with

$$e_\theta = \frac{1}{A_\theta} \int_{\Gamma_\theta} a_\theta(y) \big[\nabla_\theta m_\theta(y) - \nabla_\theta m_\theta(p_\theta)\big]\, dS(y).$$

Because a_θ is positive, the norm of this weighted average is bounded by the displayed supremum. Lemma 3 and local uniform continuity make that supremum vanish. Under Eq. (22), normalization is continuous at the nonzero vector $\nabla_\theta m_\theta^\star$, while the positive scalar A_θ has no effect on direction. Equation (23) follows. □

Combining Eqs. (19) and (23) yields

$$\cos(\nabla_\theta W_\theta, -\nabla_\theta d_h) \longrightarrow 1.$$

The result is independent of the local dimension D: dimension changes the positive prefactor A_θ, and hence gradient magnitude, but not the limiting normalized direction.

Local height–shape relation. The direction theorem above does not differentiate an uncontrolled asymptotic remainder. A separate quadratic surrogate supplies the geometric interpretation used in the main text. Freeze the density at $\rho(p_\theta)$ and set

$$m_\theta^{\mathrm{quad}}(p_\theta + \Delta y) = m_\theta^\star - \frac{1}{2}\Delta y^\top H_\theta \Delta y.$$

Its positive region is the ellipsoid $\Delta y^\top H_\theta \Delta y < 2m_\theta^\star$. Under $z = H_\theta^{1/2}\Delta y$ and with $V_D = \pi^{D/2}/\Gamma(D/2+1)$, its mass is exactly

$$W_\theta^{\mathrm{quad}} = \mathcal{S}_\theta^{(D)}(m_\theta^\star)^{D/2}, \qquad \mathcal{S}_\theta^{(D)} := 2^{D/2} V_D \frac{\rho(p_\theta)}{\sqrt{\det H_\theta}}. \tag{24}$$

For completeness, the same coefficient gives the leading-order mass of the smooth landscape. Indeed, under the rescaling

$$\Delta y = \sqrt{2m_\theta^\star}\, H_\theta^{-1/2} z,$$

the uniform third-order Taylor remainder is $O((m_\theta^\star)^{3/2}\|z\|^3)$, so the rescaled positive-margin domain converges to the unit ball $\{z : \|z\| < 1\}$. Continuity of ρ gives $\rho(p_\theta + \Delta y) = \rho(p_\theta) + o(1)$ uniformly on the shrinking component. The change of variables and dominated convergence therefore yield

$$W_\theta = \mathcal{S}_\theta^{(D)}(m_\theta^\star)^{D/2} + o\Big((m_\theta^\star)^{D/2}\Big). \tag{25}$$

We do not differentiate Eq. (25), because a little-o remainder need not have a correspondingly controlled parameter derivative. For the following exact quadratic-surrogate differentiation, assume in addition that the maps $\theta \mapsto p_\theta$, $\theta \mapsto H_\theta$, and $\theta \mapsto \rho(p_\theta)$ are C^1 along the branch. These conditions hold, for example, when the maximizer branch is C^1, ρ is C^1 near that branch, and the mixed derivatives needed to differentiate $H_\theta = -\nabla_y^2 m_\theta(p_\theta)$ are continuous. Consequently $\mathcal{S}_\theta^{(D)}$ is C^1. Within the exact quadratic surrogate, differentiation is then legitimate and gives

$$\nabla_\theta W_\theta^{\mathrm{quad}} = \frac{D}{2}\mathcal{S}_\theta^{(D)}(m_\theta^\star)^{D/2-1}\nabla_\theta m_\theta^\star + (m_\theta^\star)^{D/2}\nabla_\theta \mathcal{S}_\theta^{(D)}. \tag{26}$$

For $D = 2$, $V_2 = \pi$ and

$$W_\theta^{\mathrm{quad}} = \mathcal{S}_\theta m_\theta^\star, \qquad \nabla_\theta W_\theta^{\mathrm{quad}} = \mathcal{S}_\theta \nabla_\theta m_\theta^\star + m_\theta^\star \nabla_\theta \mathcal{S}_\theta, \qquad \mathcal{S}_\theta = \frac{2\pi\rho(p_\theta)}{\sqrt{\det H_\theta}}. \tag{27}$$

Thus the second, shape-dependent channel is lower order near birth whenever $\|\nabla_\theta \mathcal{S}_\theta\|/(\mathcal{S}_\theta\|\nabla_\theta m_\theta^\star\|)$ remains bounded. This surrogate decomposition explains the transition from height-dominated to shape-aware growth; Proposition 4, rather than formal differentiation of a little-o term, supplies the rigorous directional claim.

For any fixed comparison direction $-\nabla_\theta d_i$, Eq. (21) also gives the precise utility connection

$$u_i^{\exp} = A_\theta\left(u_i^{\mathrm{form}} + \langle e_\theta, -\nabla_\theta d_i\rangle\right).$$

Hence formation and expansion use the same task direction in the birth limit, up to a positive geometric scale. For the frozen-density quadratic surrogate in two dimensions, this scale can be identified exactly. Define

$$A_\theta^{\mathrm{quad}} := \int_{\{m_\theta^{\mathrm{quad}}=0\}} \frac{\rho(p_\theta)}{\|\nabla_y m_\theta^{\mathrm{quad}}(y)\|}\, dS(y).$$

Applying the coarea derivative to the exact identity $W_\theta^{\mathrm{quad}} = \mathcal{S}_\theta m_\theta^\star$ while holding H_θ and $\rho(p_\theta)$ fixed gives

$$A_\theta^{\mathrm{quad}} = \frac{\partial W_\theta^{\mathrm{quad}}}{\partial m_\theta^\star} = \mathcal{S}_\theta = \frac{2\pi\rho(p_\theta)}{\sqrt{\det H_\theta}}.$$

Thus $\mathcal{S}_\theta$ is exactly, rather than only heuristically, the two-dimensional quadratic-surrogate version of the geometric scale A_θ.

E COAREA REPRESENTATION AND THE DISCRETE NARROW-BAND ESTIMATOR

Fix θ and define the proof-local scalar integrand

$$\phi_i(y) := \rho(y) \langle \nabla_\theta m_\theta(y), -\nabla_\theta d_i \rangle .$$

All integrals in this section are restricted to $\operatorname{int} \mathcal{G}$. Assume that, for some $b_0 > 0$, the portion $\{y \in \operatorname{int} \mathcal{G} : |m_\theta(y)| \le b_0\}$ is compact, m_θ is C^1 and has no critical point there, ϕ_i is continuous (hence integrable) on the compact narrow-band neighborhood, and the level-set functional below is continuous at zero. These conditions hold, for example, in a sufficiently small compact tubular neighborhood of a regular zero level set when the displayed data are continuous.

Proposition 5 (Coarea narrow-band limit)**.** *Under the preceding assumptions,*

$$u_i^{\exp} = \lim_{b \to 0^+} \frac{1}{2b} \int_{\substack{y \in \operatorname{int} \mathcal{G}: \\ |m_\theta(y)| < b}} \rho(y) \langle \nabla_\theta m_\theta(y), -\nabla_\theta d_i \rangle \, dy. \tag{28}$$

Proof. The coarea formula applied to m_θ gives, for $0 < b < b_0$,

$$\int_{\substack{y \in \operatorname{int} \mathcal{G}: \\ |m_\theta(y)| < b}} \phi_i(y) \, dy = \int_{-b}^{b} G_i(t) \, dt, \tag{29}$$

$$G_i(t) := \int_{m_\theta^{-1}(t) \cap \operatorname{int} \mathcal{G}} \frac{\phi_i(y)}{\|\nabla_y m_\theta(y)\|} \, dS(y). \tag{30}$$

The factor $\|\nabla_y m_\theta\|^{-1}$ is the Jacobian converting level-set spacing in margin coordinates to spatial thickness. Consequently it is not discarded in the volume representation; it is encoded by the thickness of the narrow band. Dividing Eq. (29) by $2b$ yields the symmetric local average of G_i. Continuity at zero implies

$$\lim_{b \to 0^+} \frac{1}{2b} \int_{-b}^{b} G_i(t) \, dt = G_i(0).$$

Because $m_\theta^{-1}(0) \cap \operatorname{int} \mathcal{G} = \Gamma_\theta$ and $G_i(0)$ is exactly the boundary expression in Eq. (6), the result follows. □

For a uniform grid with cell volume ΔV, the Riemann approximation at a fixed bandwidth b is

$$\widehat{u}_i^{\exp} = \frac{\Delta V}{2b} \sum_{\substack{y_k \in \operatorname{int} \mathcal{G}: \\ |m_\theta(y_k)| < b}} \rho(y_k) \langle \nabla_\theta m_\theta(y_k), -\nabla_\theta d_i \rangle . \tag{31}$$

For every fixed $b > 0$ whose band boundary has zero volume, this sum converges to the band integral as the mesh is refined. The theoretical boundary utility is then recovered by subsequently taking $b \to 0^+$. At finite bandwidth and finite resolution, Eq. (31) is therefore an estimator, not an exact identity; its dependence on b and ΔV is left implicit, as in Eq. (14). Nonuniform discretizations require replacing ΔV by the individual cell volumes.

F ADDITIONAL EXPERIMENTAL RESULTS AND ANALYSIS

F.1 EXPERIMENTAL SETTINGS AND DATA SOURCES

We conduct experiments on visual place recognition (VPR) and UAV visual localization, corresponding to one- and two-dimensional spatial hypothesis domains, respectively. For VPR, hypotheses are parameterized by position along a reference trajectory, using publicly available Oxford data (Maddern et al., 2017). For UAV visual localization, hypotheses are parameterized by planar position, with reference aerial imagery for the Swiss scenes obtained from swisstopo. We evaluate both the geometric response of the learned scores and the corresponding task performance through the winning-region measure and localization recall.

F.2 Representative Boundary Deformation

Figure 5 provides a qualitative example of how different parameter-space update policies deform the same local score landscape. Starting from a common initial state, we apply three updates using the task-gradient direction (O), utility-aligned allocation (A), hardness-aligned allocation (H), reversed-utility allocation (R), and the direction opposite to the task gradient (O_{neg}).

The resulting boundary motion follows the ordering predicted by the task-side geometry: in this example, the local winning fraction is largest under O, followed by A and H, while R contracts the winning region and O_{neg} produces the strongest contraction. This finite-update visualization is therefore consistent with the boundary-variation mechanism in Eq. (4). Because the figure shows a single query, it is intended only as a representative geometric illustration; aggregate evidence across queries is reported in Tables 2 and 3.

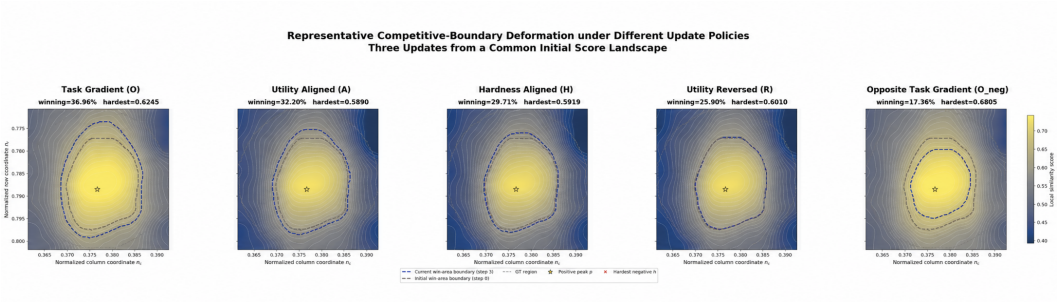


Figure 5: **Representative competitive-boundary deformation under different update policies.** All panels start from the same score landscape and show the result after three updates. O denotes the task-gradient direction; A and H assign a fixed comparison-weight spectrum according to task utility and hardness, respectively; R reverses the utility ordering; and O_{neg} uses the direction opposite to the task gradient. The gray and blue dashed contours denote the initial and updated competitive boundaries, respectively, while the contour labeled GT region denotes the fixed acceptable set $\mathcal{G}$. Thin light contours are level sets of the displayed similarity landscape. The yellow star denotes the maximal-margin acceptable hypothesis p, and the red cross denotes the strongest unacceptable hypothesis h when it lies inside the displayed crop. Gaussian smoothing ($\sigma = 2.5$, with $2\times$ upsampling) is used only for visualization. The displayed "winning" value is the percentage of cells belonging to the winning region within the shown uniform local grid, and is therefore a local discretized visualization statistic rather than the general measure W_θ. The displayed "hardest" value is the score $F_\theta(h)$ of the strongest unacceptable hypothesis and should not be confused with the comparison deficit d_h.

F.3 Numerical Formation–Expansion Alignment

The formation–expansion experiment contracts the superlevel boundary toward the maximal-margin hypothesis by increasing the normalized superlevel parameter t used in Section 4.3. Let $\widehat{g}_\theta^{\mathrm{exp}}(t)$ denote the numerically aggregated expansion direction on the contracted boundary $\Gamma_\theta(t)$, and let $g_\theta^{\mathrm{form}} := \nabla_\theta m_\theta^\star$. The explicit t-dependence is important: the experiment changes the boundary on which the expansion direction is evaluated while holding θ fixed; it does not recompute the gradient of the original zero-level winning measure W_θ. Table 6 gives numerical values underlying the trend shown in Figure 3. The directional cosine and the formation–expansion rank agreement both increase as the region contracts. In contrast, the rank agreement between hardness and expansion utility remains low. Thus, the observed approach to the formation-limit geometry is specific to the task direction and does not imply that scalar hardness recovers comparison utility.

F.4 Detailed Multi-Step Statistics on VPR

Table 7 reports paired-query statistics for the VPR experiment at InfoNCE temperature $\tau = 0.0125$. For each policy, the direction v is computed at the common initial state θ_0 and then held fixed. At step k, we evaluate the cumulative finite perturbation $\theta_k = \theta_0 + k\eta v$ and recompute the winning-region measure. Thus, "step" indexes displacement along a fixed initial direction rather than an

Table 6: Alignment as the superlevel region contracts toward formation. Values are means across valid queries. Form.–Exp. and Hard.–Exp. denote Kendall's τ_b between the corresponding comparison rankings.

t	$\cos\big(\widehat{g}_\theta^{\mathrm{exp}}(t), g_\theta^{\mathrm{form}}\big)$	Form.–Exp.	Hard.–Exp.
0.000	0.821	0.740	0.161
0.500	0.915	0.816	0.147
0.975	0.964	0.888	0.134

adaptive training trajectory that recomputes the direction after every step. The task-gradient, utility-aligned, and hardness-aligned directions all increase the median measure, but the task-gradient direction has the largest increase at both horizons. Reversing the utility ordering instead gives a negative median change. The sign pattern observed after five steps therefore remains visible after 25 steps, while the wider interquartile ranges show greater between-query variability at the longer finite displacement. These VPR statistics are distinct from the 25-step Zurich trajectory in Figure 2; the two results test the same policy ordering on different score landscapes and should not be read as numerical versions of one another.

Table 7: Relative change in the VPR winning-region measure along fixed initial directions at InfoNCE temperature $\tau = 0.0125$. Step k denotes the cumulative perturbation $\theta_k = \theta_0 + k\eta v$. Entries are the median percentage change across paired queries, with $[Q_1, Q_3]$ in brackets.

Update policy	5 steps	25 steps
Task gradient	$+17.3$ $[+8.4, +50.6]$	$+79.4$ $[+33.5, +164.5]$
Utility aligned	$+12.0$ $[+5.8, +40.0]$	$+62.3$ $[+28.1, +124.7]$
Hardness aligned	$+11.9$ $[+5.8, +38.8]$	$+63.6$ $[+29.4, +131.1]$
Utility reversed	-1.4 $[-3.9, +1.0]$	-4.1 $[-15.5, +3.8]$

F.5 From Local Geometry to Continuation Training

The fixed-direction experiment isolates the local geometric effect, but it does not by itself characterize an adaptive optimization trajectory. We therefore complement it with continuation training from a common checkpoint. Table 8 reports three representative checkpoints for the task-gradient branch and the Random64 InfoNCE baseline. The latter applies standard InfoNCE weighting to 64 randomly sampled comparisons and is therefore not identical to the controlled hardness-aligned policy H used above. Winning width is evaluated on the same 32 fixed probe queries, while Recall@1 is measured using gallery intervals of 5 m and 10 m.

For a probe query, D_W denotes the physical length of the connected winning interval containing its GT-aligned gallery position. Its threshold is the score of the strongest unacceptable gallery hypothesis, and the two boundary locations are obtained by linear interpolation between adjacent gallery positions. We set $D_W = 0$ when the GT-aligned position is not itself winning. Thus, D_W is a one-dimensional, discretized proxy for local winning-region extent rather than a second definition of the general measure W_θ.

The two branches start from the same mean width of 6.98 m. By epoch 10, the task-gradient branch increases the mean width by 7.22% and improves Recall@1 by 2.57 and 4.47 percentage points at the two gallery intervals. Over the same continuation budget, Random64 InfoNCE decreases the mean width by 2.65%, accompanied by Recall@1 changes of -0.60 and -0.37 percentage points. These checkpoint trajectories do not constitute an additional proof of the variational result. In this continuation run, their direction of change is consistent with the task-side geometry isolated by the local analysis.

Taken together, the supplemental evidence follows a single task-side chain. The predicted directions produce the expected competitive-boundary motion; formation and expansion become increasingly aligned as the superlevel region contracts; the sign pattern remains visible under finite displacement; and continuation training that increases the winning width also improves finite-budget retrieval in

Table 8: VPR continuation trajectory from a shared baseline (32 fixed probe queries, learning rate $\eta = 10^{-4}$). Width statistics are the checkpoint mean and sample standard deviation. Mean change is computed relative to the baseline checkpoint; Recall@1 is reported for 5 m and 10 m gallery intervals.

Training branch	Checkpoint	Mean D_W (m)	$\Delta\overline{D}_W$ (%)	R@1, 5 m (%)	R@1, 10 m (%)
Task gradient	Baseline	6.98 ± 6.47	0.00	88.61	83.54
	Epoch 5	7.26 ± 6.16	$+4.03$	90.49	86.87
	Epoch 10	7.49 ± 6.00	$+7.22$	91.18	88.01
Random64 InfoNCE	Baseline	6.98 ± 6.47	0.00	88.61	83.54
	Epoch 5	6.82 ± 6.24	-2.34	88.51	83.33
	Epoch 10	6.80 ± 6.23	-2.65	88.01	83.17

the reported run. The experiments therefore connect the proof-level geometry to observable VPR behavior without treating descriptive associations as substitutes for the analytical claims.