GEOMETRY, PREFERENCE, AND SHAPING: A UNIFIED VIEW OF PRECISION–RECALL OBJECTIVES

Anonymous authors
Paper under double-blind review

ABSTRACT

Region-overlap objectives are usually presented as separate formulas, obscuring which differences change their optimization behavior. We organize a broad class of precision–recall scores through one construction: a weighted quasi-arithmetic mean followed by an outer transformation. Its generator, weight, and outer map specify normalized geometry, preference, and shaping. Dice, IoU, F_β, and Tversky share a harmonic core; arithmetic and geometric aggregation belong to the same power-mean family. The score's derivative ratio determines how it mixes precision and recall gradients, and these coefficients together with the model's metric-gradient Gram matrix determine angular discrepancies. This separates shared geometry from directional equivalence and explains why shaping can preserve each item's direction yet rotate an aggregate update. We characterize compatible ratio fields and recover the generator, preference, and outer shaping of representable scores. Controlled numerical diagnostics validate the identities and constructions. The resulting framework connects objective classification, parameter-gradient interpretation, and inverse design without equating a common representation with identical training dynamics.

1 INTRODUCTION

Dice, IoU, F-measures, and Tversky are commonly treated as distinct region-overlap objectives, each with its own formula and motivation (Dice, 1945; Tversky, 1977). Yet a new formula may change a score's scale, its preference for precision over recall, or its state-dependent gradient direction. These changes have different consequences for learning. A unified account should identify the structure shared by existing objectives and explain which choices actually change a parameter update. Our organizing construction is

$$\boxed{S_{\phi,w,\psi}(P,R) = \psi\left[\phi^{-1}(w\phi(P) + (1-w)\phi(R))\right].} \tag{1}$$

The generator ϕ defines a shared aggregation rule, w sets preference, and ψ shapes the resulting score. Dice uses the harmonic core; IoU transforms that same core; F_β changes its preference; and Tversky combines preference with outer shaping. Arithmetic and geometric aggregation are other members of the power-mean family, with min and max as limiting endpoints. Thus many named objectives become explicit choices within one construction, rather than unrelated formulas (Table 1; Figure 6). This representation also exposes their gradient structure. For a score with positive partial derivatives,

$$\nabla_\theta S = S_R(\Gamma_S g_P + g_R), \qquad \Gamma_S = S_P/S_R, \quad g_P = \nabla_\theta P, \quad g_R = \nabla_\theta R. \tag{2}$$

Within Eq. (1), $\Gamma_S = [w/(1-w)]\mathcal{G}_\phi$, where $\mathcal{G}_\phi(P,R) = \phi'(P)/\phi'(R)$. This gives the anatomy of *geometry, preference, and shaping*: ϕ controls normalized state dependence, w changes the relative metric coefficients, and ψ rescales the core gradient. The model maps these coefficients into parameter space through g_P, g_R. A shared normalized geometry therefore need not imply a shared direction; a shared direction need not imply the same gradient magnitude or training trajectory.

The optimization unit and reduction complete this account. Dice and IoU have the same within-item gradient direction, but transforming each item before averaging can rotate their aggregate update. Other overlap objectives require alternative statistics, and some require information beyond any

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.

fixed ordinary (P, R) pair. The framework makes these boundaries explicit rather than assigning every segmentation loss the same representation.

We develop three connected results:

- **Classify objectives.** We separate normalized geometry, preference, and shaping, characterize directional equivalence, and establish the quadratic restriction of constant linear error weighting.
- **Explain updates.** We derive parameter-direction discrepancy through the model's metric-gradient Gram matrix and identify when shaping and reduction preserve or break directional equivalence.
- **Construct objectives.** We construct cores from compatible ratio fields and recover the full representation of existing scores, connecting inverse design to classical power means.

Quasi-arithmetic means, the chain rule, and monotone-transform equivalence are established ingredients. Their combined parameter-space and reduction analysis is the focus here. We first establish the parameter-update interpretation, then classify objectives and develop inverse design. Controlled diagnostics verify these results without claiming improved segmentation accuracy.

2 Related Work

Overlap losses and gradient analysis. V-Net uses a squared-denominator Dice loss (Milletari et al., 2016), whereas Generalised Dice pools weighted class statistics (Sudre et al., 2017). Tversky changes error coefficients (Salehi et al., 2017); focal Tversky changes the outer loss (Abraham & Khan, 2019); Unified Focal also includes cross-entropy terms (Yeung et al., 2022). Exponential-logarithmic losses also combine shaping and heterogeneous terms (Wong et al., 2018). These distinctions matter because the same loss name need not specify the statistics or reduction. Eelbode et al. (2020) study approximation relationships between Dice, Jaccard, and related losses, with segmentation experiments. Gradient modifications of Dice have also been investigated for segmentation (Ming & Xiao, 2024). We instead organize local comparisons by a metric-coefficient ratio, its realization through a common model Jacobian, and the effect of reduction. The Dice–IoU transform itself is not a new contribution.

Non-decomposable performance measures. Algorithms for non-decomposable and multiclass metrics (Narasimhan et al., 2015a;b) and calibrated linear-fractional surrogates (Bao & Sugiyama, 2020) address optimization procedures and statistical consistency. Lovász-Softmax constructs a surrogate for IoU (Berman et al., 2018), while boundary loss uses spatial distance information (Kervadec et al., 2021). Neither is automatically a function of ordinary aggregate (P, R). Here the objects are differentiable soft scores at fixed model states. Directional equivalence does not imply calibration of hard predictions, convergence, or identical finite-step optimizer trajectories.

Quasi-arithmetic and power means. Shared-generator means and their characterizations are classical (Nagumo, 1930; Aczél, 1948; Bullen, 2003); their modern extensions include quasi-arithmetic mixtures (Nielsen, 2023). The weighted harmonic interpretation and the power-mean construction use this established theory (Bullen, 2003). We use these constructions to express compatibility directly in terms of a score's derivative ratio and recover its outer shaping. The resulting scale-free family is an application of classical mean theory and multiplicative functional equations, not a new family of means.

3 Precision–Recall Gradient Geometry

When does a difference between metric-space gradients become a difference between parameter updates? The model-side gradients determine the answer.

3.1 From Objectives to Parameter Updates

An optimization unit (or item) is a specified image, class, or image–class pair whose counts are combined before forming a score. For predictions $p_j(\theta) \in (0, 1)$ and fixed binary labels y_j containing

both foreground and background, define

$$T = \sum_j p_j y_j, \quad A = \sum_j p_j, \quad B = \sum_j y_j, \qquad P = T/A, \quad R = T/B. \tag{3}$$

Write $S_P = \partial S/\partial P$, $S_R = \partial S/\partial R$, $g_P = \nabla_\theta P$, and $g_R = \nabla_\theta R$. We assume $T, A, B > 0$ and work in the interior $P, R \in I = (0, 1)$. No smoothing is used in exact identities. Empty targets, squared-denominator Dice, and additions of other losses require separate analysis. Abstract scores are defined on I^2 even when a particular model reaches only a subset of it.

Definition 1 (Gradient-ratio field)**.** *For a C^1 score $S : I^2 \to \mathbb{R}$ with $S_P, S_R > 0$, its gradient-ratio field is $\Gamma_S(P, R) = S_P/S_R$. Equal fields specify equal positive coefficient rays in the fixed (P, R) coordinates. Normalized state geometry, introduced in Section 4, is a separate quantity.*

The score-contour slope is $dR/dP = -\Gamma_S$: this field specifies a gradient ray in fixed metric coordinates, rather than a coordinate-invariant Riemannian metric. For a decreasing loss $\ell(S)$, gradient descent follows $-\nabla_\theta \ell = |\ell'(S)|\nabla_\theta S$ wherever $\ell' < 0$.

3.2 Objective Geometry Meets Model Geometry

Let $J_{\mathbf{m}} = [g_P \;\; g_R]$ and define the model geometry $K_\theta = J_{\mathbf{m}}^\top J_{\mathbf{m}}$. Under score gradient flow, $\dot{\mathbf{m}} = S_R K_\theta (\Gamma_S, 1)^\top$ for $\mathbf{m} = (P, R)^\top$, with

$$K_\theta = \begin{pmatrix} \|g_P\|^2 & g_P^\top g_R \\ g_P^\top g_R & \|g_R\|^2 \end{pmatrix}. \tag{4}$$

For $v = (\Gamma_S, 1)^\top$ and $u = (\Gamma_{\widetilde{S}}, 1)^\top$, nonzero gradients satisfy

$$\cos(\nabla_\theta S, \nabla_\theta \widetilde{S}) = \frac{v^\top K_\theta u}{\sqrt{v^\top K_\theta v}\sqrt{u^\top K_\theta u}}. \tag{5}$$

Rank-two metric gradients identify distinct coefficient rays; rank-one gradients may obscure them. In particular, Γ_S is a coefficient ratio, not the ratio of achieved precision and recall improvements. Under infinitesimal gradient ascent,

$$\dot{P} = S_R\big(\Gamma_S \|g_P\|^2 + g_P^\top g_R\big), \quad \dot{R} = S_R\big(\Gamma_S g_P^\top g_R + \|g_R\|^2\big). \tag{6}$$

Even positive score derivatives do not ensure both quantities increase when their parameter gradients conflict.

3.3 When Do Two Objectives Produce Different Updates?

At a fixed model state, let $u_j = \Gamma_{S_j} g_P + g_R$ for scores S_1, S_2 with positive partials ($j = 1, 2$). Their nonzero parameter gradients have the directions of u_j.

Proposition 1 (Pairwise directional discrepancy)**.** *If $u_1, u_2 \neq 0$, the angle $\omega \in [0, \pi]$ between them satisfies*

$$|\sin\omega| = \frac{|\Gamma_{S_1} - \Gamma_{S_2}|\sqrt{\|g_P\|^2\|g_R\|^2 - (g_P^\top g_R)^2}}{\|u_1\|\|u_2\|}. \tag{7}$$

For nonzero component gradients, the square root equals $\|g_P\|\|g_R\||\sin\vartheta|$, where ϑ is their angle.

The Gram determinant of (u_1, u_2) is $(\Gamma_{S_1} - \Gamma_{S_2})^2$ times that of (g_P, g_R), proving the identity. Objective differences become directional differences through model geometry and the mixture norms. Equal fields align; positively collinear components cannot rotate the mixture. Antiparallel components can instead cause a direction reversal across cancellation. Since $\sin\omega$ cannot distinguish alignment from opposition, Equation (5) supplies the corresponding cosine. Conflict alone does not monotonically determine the discrepancy.

Corollary 1 in Appendix A quantifies local sensitivity to the gradient ratio. Sensitivity vanishes for collinear components away from cancellation, but can be large near a zero mixture. The antiparallel direction can flip across that undefined point.

Figure 1 connects objective coefficients, model geometry, and reduction. We next identify which parts of an objective control these effects.

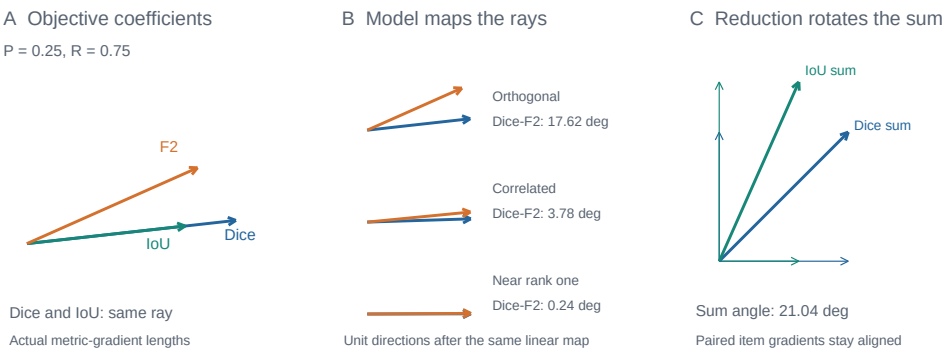


Figure 1: From objective to update. **A:** Dice and IoU have collinear metric gradients; F_2 changes the coefficient ratio. **B:** the same Dice/F_2 coefficients produce different angles under orthogonal, correlated, and nearly rank-one model maps (arrows normalized). **C:** corresponding item gradients align, but their sums rotate by $21.04°$ in the exact example of Section 4.4. Colors identify Dice (blue), IoU (green), and F_2 (orange).

4 Anatomy and Equivalence of Precision–Recall Objectives

We first construct a family in which the full field separates into constant preference and normalized state dependence, then characterize equivalence and its reduction boundary.

4.1 Geometry, Preference, and Shaping

We now analyze the unified score in Equation (1). Here $\phi, \psi \in C^1(I)$, $\phi', \psi' > 0$, and $0 < w < 1$. We abbreviate the full score by S and its inner mean by $M = M_{\phi,w}$, so $S = \psi(M)$. The parameters are generator ϕ, preference w, and outer shaping ψ. A decreasing outer map defines a loss $\ell(M)$ whose descent follows core ascent. A generator with negative derivative can be negated without changing M. Differentiation gives

$$\nabla_{P,R} S_{\phi,w,\psi} = \frac{\psi'(M)}{\phi'(M)} \begin{pmatrix} w\phi'(P) \\ (1-w)\phi'(R) \end{pmatrix}. \tag{8}$$

Thus ψ remains essential to the full gradient, even though its derivative cancels from the ratio:

$$\Gamma_S(P,R) = \frac{w}{1-w} \frac{\phi'(P)}{\phi'(R)}. \tag{9}$$

Use w as the sole preference parameter and define the normalized geometry $\mathcal{G}_\phi(P,R) = \phi'(P)/\phi'(R)$. For $S = \psi(M_{\phi,w})$, the full gradient ratio is

$$\Gamma_S = \frac{w}{1-w}\mathcal{G}_\phi, \qquad \mathcal{G}_\phi(t,t) = 1, \qquad \nabla_\theta S = \psi'(M_{\phi,w}) M_R(\Gamma_S g_P + g_R). \tag{10}$$

Thus Γ_S includes preference, whereas $\mathcal{G}_\phi$ does not; $\Gamma_S(t,t) = w/(1-w)$ fixes w. Changing w changes preference, ϕ controls normalized geometry, and ψ rescales the core gradient without changing Γ_S. These are functional roles, not independent controls of gradient magnitude: M_R also depends on the generator and preference. Section 5 gives necessary and sufficient conditions for this representation.

Figure 2 visualizes this separation: the outer map changes gradient strength without rotating the metric-plane field. Parameter-space directions additionally depend on $J_{\mathbf{m}}$.

4.2 Directional Equivalence

Definition 2. *Scores are directionally equivalent, $S \sim_\Gamma \widetilde{S}$, when their full fields agree throughout their common domain. Equal normalized geometries need not give this equivalence.*

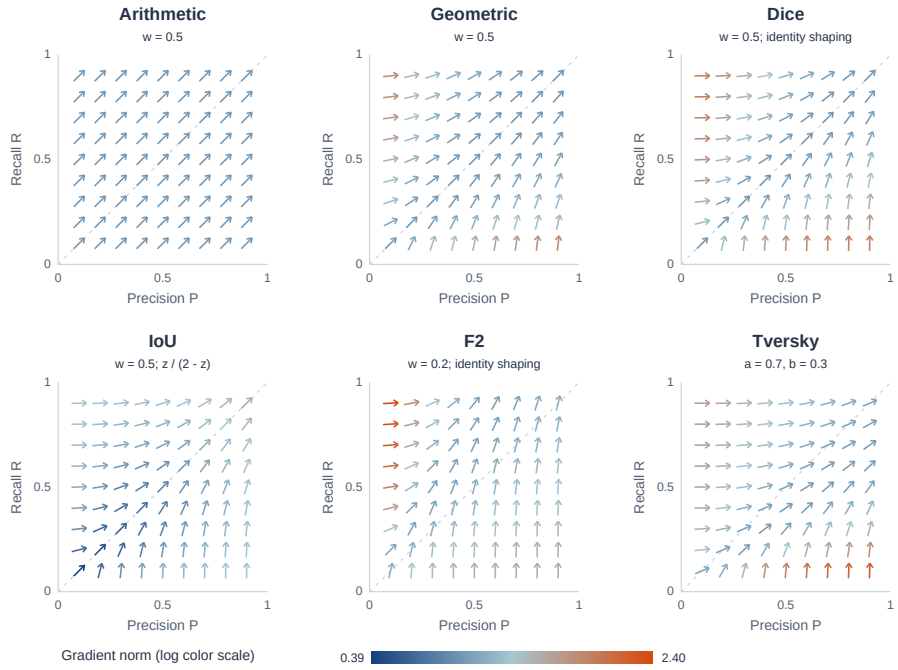


Figure 2: Score-gradient fields $\nabla_{P,R} S$ in fixed precision–recall coordinates. Equal-length arrows show direction; a shared logarithmic color scale shows the actual gradient norm. Arithmetic, geometric, and Dice change state dependence; Dice and IoU share directions but differ in strength through ψ. F_2 and Tversky change preference within the harmonic family. The dashed line is $P = R$. These are metric-plane fields, not parameter gradients.

Proposition 2 (Equivalence and parameter realization)**.** *For C^1 scores with positive partials on an open domain, equality of their fields is equivalent to a local representation $\widetilde{S} = h(S)$ with $h' > 0$ about every point. For every differentiable model map, it implies $\nabla_\theta \widetilde{S} = (\widetilde{S}_R / S_R) \nabla_\theta S$. Conversely, positive collinearity at a state with linearly independent g_P, g_R implies equal ratios at that state.*

Use (P, S) as local coordinates to prove the first claim; use the injectivity of $J_{\mathbf{m}}$ for the converse in parameter space. A global scalar transform needs further domain conditions. Smooth equivalent gradient flows have the same nonstationary paths up to positive time reparameterization (Corollary 2); discrete optimizers need not.

4.3 Common Objectives and the Quadratic Restriction

Table 1 distinguishes ordinary (P, R) objectives from alternative statistics and compositions. For ordinary soft counts, $\mathrm{FP} = A - T$ and $\mathrm{FN} = B - T$. Dice is $D = 2PR/(P + R)$, and

$$J = \frac{D}{2 - D}, \qquad \nabla_\theta J = \frac{2}{(2 - D)^2} \nabla_\theta D. \tag{11}$$

Thus Dice and IoU have equal full fields within a unit. For Tversky, fix a as the FP coefficient and b as the FN coefficient. With $k = a + b$ and $w = a/k$,

$$T_{a,b} = \psi_k(M_{-1,w}), \qquad \psi_k(z) = \frac{z}{k + (1 - k)z}, \qquad \psi_k'(z) = \frac{k}{[k + (1 - k)z]^2} > 0. \tag{12}$$

Here $M_{-1,w} = [w/P + (1 - w)/R]^{-1}$ is the harmonic mean. Changing w changes preference; changing k at fixed w changes shaping.

Proposition 3 (Quadratic restriction)**.** *For constants $a_0 \geq 0$ and $b_0, d_0, c_0 > 0$, the score*

$$Q = \frac{c_0 T}{a_0 T + b_0 \mathrm{FP} + d_0 \mathrm{FN}} = \frac{c_0}{a_0 - b_0 - d_0 + b_0/P + d_0/R} \quad \textit{has} \quad \Gamma_Q = \frac{b_0}{d_0} (R/P)^2. \tag{13}$$

Any outer transform with positive derivative preserves this field.

Differentiation proves the result. Constant coefficients in this TP-over-linear-count family cannot change quadratic state dependence. State-dependent coefficients, smoothing, alternative statistics, and different numerators fall outside this restriction.

4.4 Shaping Before and After Reduction

Let $\alpha_i > 0$ be fixed item weights summing to one, and let $\ell' < 0$ on the attained score range. Shaping before reduction gives

$$L_{\text{item}} = \sum_i \alpha_i \ell(S_i), \qquad -\nabla_\theta L_{\text{item}} = \sum_i \alpha_i |\ell'(S_i)| S_{i,R}(\Gamma_{S_i} g_{P,i} + g_{R,i}). \tag{14}$$

The multiplier changes relative item contributions; their actual norms also depend on the core and model gradients. By contrast,

$$L_{\text{after}} = \ell\left(\sum_i \alpha_i S_i\right), \qquad -\nabla_\theta L_{\text{after}} = \left|\ell'\left(\sum_i \alpha_i S_i\right)\right| \sum_i \alpha_i \nabla_\theta S_i \tag{15}$$

has one common multiplier and preserves the current nonzero aggregate direction. More generally, with $\mathbf{B} = [\alpha_1 \nabla_\theta S_1 \ \cdots \ \alpha_N \nabla_\theta S_N]$, positive multipliers $\mathbf{c}$ preserve it exactly when $\mathbf{c} - \lambda \mathbf{1} \in \ker \mathbf{B}$ for some $\lambda > 0$ (Appendix A.4).

Exact counterexample. Two items with labels $(1, 0)$ and predictions $(\theta_i, 1 - \theta_i)$ have $D_i = \theta_i$. At $(\theta_1, \theta_2) = (0.2, 0.8)$, summed Dice has gradient $(1, 1)$ and summed IoU has gradient $(50/81, 25/18)$. Their angle is $\arctan(9/4) - \pi/4 \approx 21.04°$, although each paired item gradient aligns. IoU gives the higher-Dice item a larger relative multiplier, not necessarily a larger absolute contribution in other models.

For focal powers, $|\ell'(s)| = \gamma(1 - s)^{\gamma - 1}$: lower scores have larger multipliers only for $\gamma > 1$; for $0 < \gamma < 1$ the multiplier grows near a perfect score. This concerns the actual exponent and differs from focal cross-entropy (Lin et al., 2017). Pooled-count objectives are a third reduction: their apparent averaging weights depend on predictions and must also be differentiated (Appendix B.2).

5 From Gradient Geometry to Objective Design

We now construct the full family in Equation (1), including its outer map, and characterize its scope: which positive ratio fields admit a shared generator, and when does an existing score have this representation?

5.1 Compatibility and Canonical Recovery

Theorem 1 (Shared-generator recovery)**.** *A continuous prescribed field $\Gamma : I^2 \to (0, \infty)$ is the gradient ratio of a weighted quasi-arithmetic mean with $\phi \in C^1(I)$, $\phi' > 0$, if and only if some $w \in (0, 1)$ satisfies*

$$\Gamma(t, t) = \frac{w}{1 - w}, \qquad \Gamma(x, y)\Gamma(y, z) = \frac{w}{1 - w}\Gamma(x, z) \quad (x, y, z, t \in I). \tag{16}$$

Here Γ has no score subscript until a score is constructed. For any anchor $t_0 \in I$, recover

$$w = \frac{\Gamma(t_0, t_0)}{1 + \Gamma(t_0, t_0)}, \qquad \phi(x) = \frac{1 - w}{w} \int_{t_0}^{x} \Gamma(u, t_0)\, du. \tag{17}$$

The weight is unique, and the generator is unique up to a positive affine transformation. If $S \in C^1(I^2)$ has positive partials and ratio Γ, then globally

$$S(P, R) = \psi(M_{\phi,w}(P, R)), \qquad \psi(t) = S(t, t), \qquad \psi' > 0. \tag{18}$$

Proof sketch. Necessity follows from Equation (9). Setting $z = t_0$ in the cocycle gives $\Gamma(P, R) = [w/(1 - w)]\phi'(P)/\phi'(R)$ for the recovered generator, which constructs the mean. In $(\phi(P), \phi(R))$

Table 1: Formula-based taxonomy of the unified family $S_{\phi,w,\psi} = \psi(M_{\phi,w})$. Put $z = M_{\phi,w}(U, V)$, $\mathcal{G}_\phi = \phi'(U)/\phi'(V)$. The outer column lists increasing score maps ψ or explicitly marked decreasing losses ℓ. All formulas are unsmoothed.

Objective	Score / statistics	$\phi(x)$	w	$\psi(z)$ or loss $\ell(z)$	$\mathcal{G}_\phi(U, V)$
A. Ordinary Precision–Recall: $(U, V) = (P, R)$					
Arithmetic	$wU + (1-w)V$	x	w	z	$(V/U)^0$
Geometric	$U^w V^{1-w}$	$\log x$	w	z	$(V/U)^1$
Power mean	$[wU^q + (1-w)V^q]^{1/q}$	x^q/q	w	z; $q \neq 0$	$(V/U)^{1-q}$
Dice / F_1	$2UV/(U+V)$	$-1/x$	$1/2$	z; $\ell = 1 - z$	$(V/U)^2$
F_β	$\frac{(1+\beta^2)UV}{\beta^2 U+V}$	$-1/x$	$\frac{1}{1+\beta^2}$	z; $\ell = 1 - z$	$(V/U)^2$
Tversky	$\frac{T}{T+a\mathrm{FP}+b\mathrm{FN}}$	$-1/x$	$\frac{a}{a+b}$	$\psi_k(z) = \frac{z}{k+(1-k)z}$	$(V/U)^2$
IoU / Jaccard	$\frac{UV}{U+V-UV}$	$-1/x$	$1/2$	$z/(2-z)$	$(V/U)^2$
Log / log-cosh / focal Dice	$z = 2UV/(U+V)$	$-1/x$	$1/2$	$\ell = -\log z$; $\ell = \log\cosh(1-z)$; $\ell = (1-z)^\gamma$	$(V/U)^2$
Focal Tversky	$z = [w/U + (1-w)/V]^{-1}$	$-1/x$	$\frac{a}{a+b}$	$\ell = [1 - \psi_k(z)]^\gamma$	$(V/U)^2$
Min / max[a]	$\min(U, V)$; $\max(U, V)$	—	$(0, 1)$	z	—
B. Alternative metric pairs					
clDice; boundary F_1	$2UV/(U+V)$	$-1/x$	$1/2$	z	$(V/U)^2$
Boundary IoU	$UV/(U+V-UV)$	$-1/x$	$1/2$	$z/(2-z)$	$(V/U)^2$
Balanced accuracy	$(U+V)/2$	x	$1/2$	z	$(V/U)^0$
G-mean	$\sqrt{UV}$	$\log x$	$1/2$	z	$(V/U)^1$
C. Lifted or pooled statistics					
Generalised Dice	$\frac{2T_*}{A_*+B_*}$; $(U, V) = (\frac{T_*}{A_*}, \frac{T_*}{B_*})$	$-1/x$	$1/2$	z	$(V/U)^2$
Squared Dice / V-Net	$\frac{2T}{A_2+B_2}$; $(U, V) = (\frac{T}{A_2}, \frac{T}{B_2})$	$-1/x$	$1/2$	z	$(V/U)^2$
L^2 Jaccard / Tanimoto	$T/(A_2 + B_2 - T)$	$-1/x$	$1/2$	$z/(2-z)$	$(V/U)^2$
Fractal Tanimoto[b]	$\frac{T}{2^{nf}(A_2+B_2)-(2^{nf+1}-1)T}$	$-1/x$	$1/2$	$\frac{z}{2^{nf+1}-(2^{nf+1}-1)z}$	$(V/U)^2$
D. Compositions: entries describe each constituent, not their sum					
Macro / weighted / focal	$\sum_i \alpha_i \psi_i(M_i)$	ϕ_i	w_i	$\psi_i(z)$ or $\ell_i(z)$	$\mathcal{G}_{\phi_i}$
Tanimoto + complement	$\frac{\mathcal{T}(p,y)+\mathcal{T}(1-p,1-y)}{2}$	$-1/x$	$1/2$	$z/(2-z)$	$(V/U)^2$
E. Conditional representations and scope boundaries					
Surface Dice	$\frac{o_{pred}+o_{GT}}{A_{pred}+A_{GT}}$	x	w_A	z; $\nabla w_A = 0$	$(V/U)^0$
Soft-label Dice semimetric	$D_{\mathrm{semimetric}}(p, y)$	—	—	—[c]	—
Overlap + CE	$\lambda_o \ell_{overlap} + \lambda_c \ell_{CE}$	—	—	—[c]	—
Recognition quality	$RQ = 2UV/(U+V)$	$-1/x$	$1/2$	z	$(V/U)^2$
Panoptic quality	$PQ = SQ\,RQ$	$-1/x$	$1/2$	$SQ\,z$; $\nabla SQ = 0$	$(V/U)^2$
Wasserstein Dice	$D_W(p, y)$	—	—	—[c]	—

Conventions. $n_f \geq 0$ is the fractal depth; $k = a + b$, $w = a/k$ for Tversky; $q = 0$ uses $\phi = \log x$. $T = \sum_j p_j y_j$, $A_2 = \sum_j p_j^2$, $B_2 = \sum_j y_j^2$; starred counts share pooling weights. The last two rows in C use the squared-Dice pair above. In B, (U, V) denotes topology precision/sensitivity, boundary precision/recall under the stated matching or boundary-set rule, or sensitivity/specificity, respectively. $\mathcal{T}$ is the L^2 Tanimoto score. All denominators are positive; $0 < w < 1$, $a, b, \beta, \gamma > 0$.

Conditional rows. Surface Dice uses $(U, V) = (o_{pred}/A_{pred}, o_{GT}/A_{GT})$ and $w_A = A_{pred}/(A_{pred} + A_{GT})$; variable weights add $(U - V)\nabla w_A$. Panoptic quality uses instance precision/recall; variable SQ adds $RQ\nabla SQ$. Their displayed fields hold only when these additional quantities are fixed.

Scope. $\psi' > 0$ preserves score-ascent rays; $\ell' < 0$ gives the corresponding loss-descent rays. Different statistics change model-side gradients; sums need not preserve direction. Hard matching needs a differentiable surrogate. [a]Min/max are the lower/upper limiting members of the power-mean family, for any fixed $w \in (0, 1)$; they lie outside the strictly positive-partial theory. [b]The exact displayed fractal formula is classified. [c]No universal ordinary (P, R) representation is asserted; see Appendix B.3.

coordinates, its level sets are connected line segments. A score with the same ratio is constant on each segment, which intersects the diagonal; its value is therefore $S(M, M)$. This proves the global shaping formula. The anchor fixes $\phi(t_0) = 0$ and $\phi'(t_0) = 1$; the field alone does not fix ψ. Full details are in Appendix A.

For a C^3 score, an equivalent differential test is

$$\partial_P \partial_R \log \Gamma_S = 0, \qquad \Gamma_S(t, t) = w/(1 - w). \tag{19}$$

Both conditions matter. The score $(P + R^2)/2$ has a separable ratio $1/(2R)$ but a nonconstant diagonal, so it needs different metric generators. The positive-partial score $(P + R + P^2R)/3$ has a nonzero mixed log-ratio derivative and fails separability itself (Appendix A.9). Thus incompatibility rules out this shared-generator anatomy, not the existence of a differentiable score.

5.2 Constructing a Desired Field

For a prescribed field Γ, Equation (16) is the exact compatibility test. When it holds, Equation (17) constructs the core and preference; choose a score transform with positive derivative or a loss transform with negative derivative:

$$\Gamma \longrightarrow (w, \phi) \longrightarrow \psi(M_{\phi,w}). \tag{20}$$

A desired positive loss multiplier $m(s)$ can be implemented by $\ell(s) = \int_s^1 m(u)\, du$ when finite. This controls an additional multiplier of the mean gradient; it cannot assign different multipliers to two items with the same score. Approximate recovery from a noisy estimated field is a separate problem not solved by the exact test.

Worked construction. Suppose the desired ratio is $\Gamma(P, R) = 2R/P$. Its diagonal is 2, and it satisfies Equation (16). Recovery gives $w = 2/3$ and a generator equivalent to $\phi(x) = \log x$, hence $M = P^{2/3}R^{1/3}$. Choosing $\psi(z) = z^2$ yields $S = P^{4/3}R^{2/3}$ with $S_P/S_R = 2R/P$. The prescribed field fixes the core and preference; the outer map is an additional design choice.

5.3 Scale-Free Fields Recover Power Means

Call a normalized field scale-free when $\mathcal{G}_\phi(P, R) = F(P/R)$. The cocycle then gives $F(xy) = F(x)F(y)$ for all positive x, y; continuity implies $F(x) = x^{q-1}$. Consequently, all scale-free compatible means are

$$M_{q,w}(P, R) = \begin{cases} [wP^q + (1 - w)R^q]^{1/q}, & q \neq 0, \\ P^w R^{1-w}, & q = 0, \end{cases} \qquad \Gamma_{M_{q,w}} = \frac{w}{1 - w}(P/R)^{q-1}. \tag{21}$$

The increasing generator is $\phi(t) = t^q/q$ for $q \neq 0$, and $\log t$ for $q = 0$. Arithmetic, geometric, and harmonic means correspond to $q = 1, 0, -1$. Equivalently, a target $[w/(1 - w)](R/P)^\lambda$ gives $q = 1 - \lambda$. Changing q changes state dependence, which constant linear error weights cannot do. When $P < R$ and $q < 1$, the precision coefficient exceeds its diagonal value; achieved metric progress still depends on Equation (6).

A compatible field need not be scale-free. For $\phi'(t) = e^{-\tau t}$, $\tau > 0$,

$$M = -\tau^{-1} \log\left(we^{-\tau P} + (1 - w)e^{-\tau R}\right), \qquad \Gamma_M = \frac{w}{1 - w} e^{-\tau(P-R)}. \tag{22}$$

This separates shared-generator compatibility from dependence on the metric ratio alone. Appendix A proves the scale-free characterization; Figure 5 illustrates the classical cases.

6 Empirical Validation of Gradient Geometry

We distinguish numerical verification of identities from measurements of their consequences. All experiments use synthetic shared-parameter models at common states; no segmentation training is reported.

6.1 Setup and Implementation Checks

We sample 200 shared-parameter models, each with eight binary-labeled items, 64 pixels per item, and 12 parameters. Both labels occur in every item. Seeds, predictions, and finite-difference protocols are specified in Appendix C.

Parameter finite differences check the analytic gradients; Tables 2 and 3 report implementation residuals in Appendix D. The following comparisons concern direction, magnitude, and aggregation.

6.2 Equivalence and Objective–Model Interaction

Within-item Dice–IoU cosines agree with one to floating-point precision. Their IoU-to-Dice norm ratio is $2/(2-D)^2$, spanning 0.504–1.824 in these trials. Independent parameter differences also check log-Dice, squared-residual Dice loss $(1-D)^2$, and squared-residual Tversky loss.

Across 1,600 states, the three neutral power means give 4,800 pairwise comparisons. Measured direction discrepancies match Equation (7); Figure 4 relates them to component alignment. This verifies the identity rather than predictive performance; its acute angle does not distinguish alignment from opposition.

6.3 Construction and Reduction Effects

For targets $[w/(1-w)](R/P)^\lambda$ with $\lambda \in \{0, 0.5, 1, 1.5, 2\}$, power means with $q = 1-\lambda$ recover the prescribed ratios at three preferences. Independent parameter differences on ten trials also recover the parameter gradients of all five neutral designs. Exponential means supply a non-scale-free control. These are construction checks; they do not show that one geometry trains better.

Post-reduction shaping preserves direction, whereas itemwise Dice–IoU shaping produces a median batch rotation of 11.637° and a maximum of 31.067° (Figure 3). Together with the exact 21.04° counterexample, this demonstrates possible aggregate nonequivalence, not its prevalence in trained networks.

7 Discussion

Equal $\mathcal{G}_\phi$ means equal normalized geometry; equal Γ_S additionally fixes preference; equal gradients also require matching scale. Directional equivalence holds within each item for every differentiable model, but may fail after shaping and reduction. Comparisons must specify the statistics, optimization unit, and reduction. The compatibility test defines the shared-generator family, not all differentiable scores. Constant FP/FN coefficients retain harmonic geometry; compatible nonquadratic fields require a different core. Neither compatibility nor a larger metric coefficient implies better accuracy, calibration, or optimization stability.

The evidence is limited to synthetic shared-parameter models. Practical validation requires common-checkpoint measurements and controlled training comparisons separating geometry, preference, shaping, and reduction. Smoothing, empty targets, and hard matching also require extensions beyond the interior assumptions. Extensions beyond two metrics and recovery from noisy fields remain open; Appendix B.3 gives the basic multi-metric identity.

8 Conclusion

The unified score $S = \psi(M_{\phi,w})$ separates normalized geometry, preference, and shaping. It classifies familiar overlap objectives, explains their parameter-gradient relationships and reduction effects, and supports construction from compatible ratio fields. The resulting distinction is operational: both the gradient-ratio field and the multipliers under the chosen reduction determine an update. A shared core alone does not establish identical training behavior.

GENERATIVE AI DISCLOSURE

Generative AI tools were used to assist with manuscript preparation and refinement. The authors retain full responsibility for the accuracy, originality, and integrity of the final manuscript, including its claims, interpretations, and conclusions.

REFERENCES

Nabila Abraham and Naimul Mefraz Khan. A Novel Focal Tversky Loss Function With Improved Attention U-Net for Lesion Segmentation. In *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*, pp. 683–687, 2019.

J. Aczél. On mean values. *Bulletin of the American Mathematical Society*, 54(4):392–400, 1948.

Han Bao and Masashi Sugiyama. Calibrated Surrogate Maximization of Linear-fractional Utility in Binary Classification. In *Proceedings of the Twenty Third International Conference on Artificial Intelligence and Statistics*, volume 108 of *Proceedings of Machine Learning Research*, pp. 2337–2347, 2020.

Maxim Berman, Amal Rannen Triki, and Matthew B. Blaschko. The Lovász-Softmax Loss: A Tractable Surrogate for the Optimization of the Intersection-Over-Union Measure in Neural Networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 4413–4421, 2018.

P. S. Bullen. *Handbook of Means and Their Inequalities*. Kluwer Academic Publishers, 2003.

Bowen Cheng, Ross Girshick, Piotr Dollár, Alexander C. Berg, and Alexander Kirillov. Boundary IoU: Improving Object-Centric Image Segmentation Evaluation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 15334–15342, 2021.

Foivos I. Diakogiannis, François Waldner, and Peter Caccetta. Looking for Change? Roll the Dice and Demand Attention. *Remote Sensing*, 13(18):3707, 2021.

Lee R. Dice. Measures of the Amount of Ecologic Association Between Species. *Ecology*, 26(3): 297–302, 1945.

Tom Eelbode, Jeroen Bertels, Maxim Berman, Dirk Vandermeulen, Frederik Maes, Raf Bisschops, and Matthew B. Blaschko. Optimization for Medical Image Segmentation: Theory and Practice When Evaluating With Dice Score or Jaccard Index. *IEEE Transactions on Medical Imaging*, 39 (11):3679–3690, 2020.

Lucas Fidon, Wenqi Li, Luis C. Garcia-Peraza-Herrera, Jinendra Ekanayake, Neil Kitchen, Sébastien Ourselin, and Tom Vercauteren. Generalised Wasserstein Dice Score for Imbalanced Multi-class Segmentation Using Holistic Convolutional Networks. In *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*, pp. 64–76, 2018.

Hoel Kervadec, Jihene Bouchtiba, Christian Desrosiers, Eric Granger, Jose Dolz, and Ismail Ben Ayed. Boundary loss for highly unbalanced segmentation. *Medical Image Analysis*, 67: 101851, 2021.

Alexander Kirillov, Kaiming He, Ross Girshick, Carsten Rother, and Piotr Dollár. Panoptic Segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 9404–9413, 2019.

Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, and Piotr Dollár. Focal Loss for Dense Object Detection. In *2017 IEEE International Conference on Computer Vision (ICCV)*, pp. 2999–3007, 2017.

Fausto Milletari, Nassir Navab, and Seyed-Ahmad Ahmadi. V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation. In *2016 Fourth International Conference on 3D Vision (3DV)*, pp. 565–571, 2016.

Qi Ming and Xiaowu Xiao. Towards Accurate Medical Image Segmentation With Gradient-Optimized Dice Loss. *IEEE Signal Processing Letters*, 31:191–195, 2024.

Mitio Nagumo. Über eine Klasse der Mittelwerte. *Japanese Journal of Mathematics*, 7:71–79, 1930.

Harikrishna Narasimhan, Purushottam Kar, and Prateek Jain. Optimizing Non-decomposable Performance Measures: A Tale of Two Classes. In *Proceedings of the 32nd International Conference on Machine Learning*, volume 37 of *Proceedings of Machine Learning Research*, pp. 199–208, 2015a.

Harikrishna Narasimhan, Harish Ramaswamy, Aadirupa Saha, and Shivani Agarwal. Consistent Multiclass Algorithms for Complex Performance Measures. In *Proceedings of the 32nd International Conference on Machine Learning*, volume 37 of *Proceedings of Machine Learning Research*, pp. 2398–2407, 2015b.

Frank Nielsen. Quasi-arithmetic Centers, Quasi-arithmetic Mixtures, and the Jensen-Shannon ∇-Divergences. In *Geometric Science of Information*, pp. 147–156, 2023.

Stanislav Nikolov, Sam Blackwell, Alexei Zverovitch, Ruheena Mendes, Michelle Livne, Jeffrey De Fauw, Yojan Patel, Clemens Meyer, Harry Askham, Bernardino Romera-Paredes, Christopher Kelly, Alan Karthikesalingam, Carlton Chu, Dawn Carnell, Cheng Boon, Derek D'Souza, Syed Ali Moinuddin, Bethany Garie, Yasmin McQuinlan, Sarah Ireland, Kiarna Hampton, Krystle Fuller, Hugh Montgomery, Geraint Rees, Mustafa Suleyman, Trevor Back, Cían Owen Hughes, Joseph R Ledsam, and Olaf Ronneberger. Clinically Applicable Segmentation of Head and Neck Anatomy for Radiotherapy: Deep Learning Algorithm Development and Validation Study. *Journal of Medical Internet Research*, 23(7):e26151, 2021.

Seyed Sadegh Mohseni Salehi, Deniz Erdogmus, and Ali Gholipour. Tversky Loss Function for Image Segmentation Using 3D Fully Convolutional Deep Networks. In *Machine Learning in Medical Imaging*, pp. 379–387, 2017.

Suprosanna Shit, Johannes C. Paetzold, Anjany Sekuboyina, Ivan Ezhov, Alexander Unger, Andrey Zhylka, Josien P. W. Pluim, Ulrich Bauer, and Bjoern H. Menze. clDice – A Novel Topology-Preserving Loss Function for Tubular Structure Segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 16560–16569, 2021.

Carole H. Sudre, Wenqi Li, Tom Vercauteren, Sebastien Ourselin, and M. Jorge Cardoso. Generalised Dice Overlap as a Deep Learning Loss Function for Highly Unbalanced Segmentations. In *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*, pp. 240–248, 2017.

Amos Tversky. Features of similarity. *Psychological Review*, 84(4):327–352, 1977.

Zifu Wang, Teodora Popordanoska, Jeroen Bertels, Robin Lemmens, and Matthew B. Blaschko. Dice Semimetric Losses: Optimizing the Dice Score with Soft Labels. In *Medical Image Computing and Computer Assisted Intervention – MICCAI 2023*, pp. 475–485, 2023.

Ken C. L. Wong, Mehdi Moradi, Hui Tang, and Tanveer Syeda-Mahmood. 3D Segmentation with Exponential Logarithmic Loss for Highly Unbalanced Object Sizes. In *Medical Image Computing and Computer Assisted Intervention – MICCAI 2018*, pp. 612–619, 2018.

Michael Yeung, Evis Sala, Carola-Bibiane Schönlieb, and Leonardo Rundo. Unified Focal loss: Generalising Dice and cross entropy-based losses to handle class imbalanced medical image segmentation. *Computerized Medical Imaging and Graphics*, 95:102026, 2022.

A PROOFS AND CONSTRUCTIONS

A.1 EQUIVALENCE AND PARAMETER REALIZATION

Proof of Proposition 2. Equality of ratios gives $\nabla_{P,R}\widetilde{S} = \rho\nabla_{P,R}S$, where $\rho = \widetilde{S}_R/S_R > 0$. Since $S_R > 0$, the inverse function theorem makes (P, S) valid local coordinates. At fixed S, the derivative of $\widetilde{S}$ in P is $\widetilde{S}_P - \widetilde{S}_R S_P/S_R = 0$. After restricting to a coordinate rectangle, $\widetilde{S} = h(S)$ with $h' = \rho > 0$. Conversely, the chain rule cancels h' in the ratio.

Composing with a differentiable parameter map gives $\nabla_\theta\widetilde{S} = \rho\nabla_\theta S$. If $J_{\mathbf{m}} = [g_P\ g_R]$ has rank two, positive collinearity implies $J_{\mathbf{m}}(\Gamma_{\widetilde{S}}, 1)^\top = \lambda J_{\mathbf{m}}(\Gamma_S, 1)^\top$ after removing the positive score scales. Injectivity and the second coordinate imply $\lambda = 1$ and equal ratios. Zero gradients satisfy the vector equality but have no direction. Rank one need not identify ratios: for the abstract map $P = R = \theta$, all positive-partial scores have positive scalar derivatives, regardless of their ratios. □

A.2 PAIRWISE DISCREPANCY AND SENSITIVITY

Proof of Proposition 1. Set $W = [u_1\ u_2] = J_{\mathbf{m}}C$ with $C = \left(\begin{smallmatrix}\Gamma_{S_1} & \Gamma_{S_2}\\ 1 & 1\end{smallmatrix}\right)$. Then $\det(W^\top W) = \det(C)^2\det(J_{\mathbf{m}}^\top J_{\mathbf{m}})$. The left side is $\|u_1\|^2\|u_2\|^2\sin^2\omega$ and $\det C = \Gamma_{S_1} - \Gamma_{S_2}$. Taking square roots proves the formula, even if a component gradient vanishes, provided both mixtures are nonzero. Zero sine identifies collinearity, not its sign. □

Corollary 1 (Local directional sensitivity)**.** *At a fixed model state, let $n(r) = (rg_P + g_R)/\|rg_P + g_R\|$, with scalar ratio $r > 0$, wherever the denominator is nonzero. Then*

$$\left\|\frac{\partial n}{\partial r}\right\| = \frac{\sqrt{\|g_P\|^2\|g_R\|^2 - (g_P^\top g_R)^2}}{\|rg_P + g_R\|^2}. \tag{23}$$

Proof of Corollary 1. Put $v = rg_P + g_R$ and $n = v/\|v\|$. At fixed component gradients, $n' = (\mathrm{Id} - nn^\top)g_P/\|v\|$. Thus

$$\|n'\|^2 = \frac{\|g_P\|^2\|v\|^2 - (g_P^\top v)^2}{\|v\|^4} = \frac{\|g_P\|^2\|g_R\|^2 - (g_P^\top g_R)^2}{\|v\|^4}.$$

Taking square roots proves the result wherever $v \neq 0$. □

A.3 QUADRATIC RESTRICTION

Proof of Proposition 3. Since $\mathrm{FP}/T = P^{-1} - 1$ and $\mathrm{FN}/T = R^{-1} - 1$, division by T gives $Z = a_0 + b_0(P^{-1} - 1) + d_0(R^{-1} - 1) > 0$ on I^2. Differentiating $Q = c_0/Z$ gives $Q_P = c_0 b_0/(P^2Z^2)$ and $Q_R = c_0 d_0/(R^2Z^2)$. Their ratio is $(b_0/d_0)(R/P)^2$, unchanged by a positive outer derivative. The excluded boundary point $P = R = 1$ would give $Z = 0$ if $a_0 = 0$. □

With $k = b_0 + d_0$ and $w = b_0/k$, the score is $Q = c_0 M_{-1,w}/[k + (a_0 - k)M_{-1,w}]$. Its outer derivative $c_0 k/[k + (a_0 - k)M_{-1,w}]^2$ is positive. Thus the family has a harmonic core even when c_0 does not normalize its range to $(0, 1)$.

A.4 BATCH ALIGNMENT

Proposition 4 (Exact batch-alignment criterion)**.** *Let $\mathbf{B} = [\alpha_1\nabla_\theta S_1\ \cdots\ \alpha_N\nabla_\theta S_N]$, let $c_i > 0$, and suppose $\mathbf{B1}$ and $\mathbf{Bc}$ are nonzero. They are positively collinear if and only if some $\lambda > 0$ satisfies*

$$\mathbf{c} - \lambda\mathbf{1} \in \ker\mathbf{B}. \tag{24}$$

If the item gradients are linearly independent, all multipliers must be equal. With dependent gradients, unequal multipliers can also align.

Proof. Positive collinearity is precisely $\mathbf{Bc} = \lambda\mathbf{B1}$ with $\lambda > 0$. Rearrangement gives the kernel criterion; full column rank makes the kernel trivial. □

For the soft-count counterexample, $A_i = B_i = 1$ and $T_i = \theta_i$, so $D_i = \theta_i$ and $J_i = \theta_i/(2-\theta_i)$. At $(0.2, 0.8)$ the two aggregate directions are $(1, 1)$ and $(50/81, 25/18)$; their angle is $\arctan(9/4) - \pi/4$. Replacing sums by means leaves it unchanged.

A.5 Shared-generator recovery and uniqueness

Proof of Theorem 1. Equation (9) implies the diagonal and cocycle identities. Conversely, define $\phi'(x) = [(1-w)/w]\Gamma(x, t_0)$. The cocycle with $z = t_0$ gives $\Gamma(x, y) = [w/(1-w)]\phi'(x)/\phi'(y)$. Integrating this positive continuous derivative defines a C^1 generator. Its inverse and mean are C^1 and have the requested ratio.

For an existing C^1 score, define $\overline{S}(x, y) = S(\phi^{-1}(x), \phi^{-1}(y))$ on the rectangle $\phi(I)^2$. Its partial-derivative ratio is $w/(1-w)$. It is therefore constant along each segment $wx + (1-w)y = \zeta$, since the directional derivative along $(1-w, -w)$ vanishes. The segment intersects the diagonal at (ζ, ζ). Writing $m = \phi^{-1}(\zeta)$ gives $S(P, R) = S(m, m)$, hence $S = \psi(M)$ with $\psi(t) = S(t, t)$ and $\psi'(t) = S_P(t, t) + S_R(t, t) > 0$.

Any other representation has the same diagonal ratio and thus the same w. Equality of generator ratios implies $\widetilde{\phi}'(x)/\phi'(x) = \widetilde{\phi}'(y)/\phi'(y)$ for every x, y, a positive constant. Thus $\widetilde{\phi} = a\phi + b$, $a > 0$. Conversely, this affine change leaves the mean unchanged. The anchor convention fixes both affine freedoms. □

A.6 Differential characterization

Proposition 5 (Smooth compatibility test)**.** *If $S \in C^3(I^2)$ has positive partials, shared-generator representability with positive generator and shaping derivatives is equivalent to*

$$f_{PR} = 0, \qquad f(t, t) \text{ constant}, \qquad f = \log(S_P/S_R). \tag{25}$$

The generator and shaping may be chosen in C^3.

Proof. Necessity follows from $f(P, R) = \log[w/(1-w)] + \log\phi'(P) - \log\phi'(R)$. Conversely, on the rectangle, $f_{PR} = 0$ gives $f(P, R) = u(P) + v(R)$ with $u, v \in C^2$. The diagonal condition gives $v(t) = \log[w/(1-w)] - u(t)$, so $\phi' = e^u > 0$ has class C^2. Integrating $\phi' = e^u$ gives $\phi \in C^3$; Theorem 1 supplies $\psi(t) = S(t, t) \in C^3$. □

A.7 Scale-free recovery

Suppose $[(1-w)/w]\Gamma(P, R) = F(P/R)$ with continuous positive F on $(0, \infty)$. For any $s, t > 0$, choose $z > 0$ small enough that $z, tz, stz \in I$. The cocycle gives $F(s)F(t) = F(st)$. Hence $H(u) = \log F(e^u)$ is continuous and additive on $\mathbb{R}$. It is linear first on rationals and then, by continuity, on reals. Thus $F(s) = s^{q-1}$. Integrating $\phi'(x) = x^{q-1}$ yields x^q/q for $q \neq 0$ and $\log x$ for $q = 0$, giving Equation (21). Negative q still gives a positive derivative.

A.8 Gradient flows and optimizers

Corollary 2 (Continuous-flow equivalence)**.** *Suppose two parameter-gradient fields are related by $V_2(\theta) = \rho(\theta)V_1(\theta)$ with $\rho > 0$ and sufficient smoothness for locally unique flows. Their nonstationary paths agree up to a positive time reparameterization on intervals where that change exists.*

Proof. If $\theta_1(s)$ solves the first flow, define $ds/dt = \rho(\theta_1(s))$. Then $\theta_2(t) = \theta_1(s(t))$ solves the second by the chain rule. The two vector fields also have the same stationary points. The time change need not exist over identical infinite horizons. □

Fixed discrete step sizes do not implement this exact time change. Momentum and adaptive optimizer states can evolve differently. Global norm clipping preserves direction at a common nonzero state but changes step magnitude; coordinatewise clipping can also change direction. Itemwise equivalence is insufficient for this corollary if reduction produces noncollinear aggregate fields.

A.9 Design scope and score-dependent weights

For $S = (P + R^2)/2$, $\Gamma_S = 1/(2R)$ is separable but has nonconstant diagonal. For $S = (P + R + P^2R)/3$, both partials are positive, but

$$\partial_P \partial_R \log \Gamma_S = \frac{2}{(1 + 2PR)^2} \neq 0. \tag{26}$$

Neither score admits a shared-generator representation. More generally, constructing a score from a prescribed field requires a positive integrating factor $\mu = S_R$ with $\partial_R(\mu\Gamma) = \partial_P \mu$ in the smooth setting. Failure of the shared-generator test is not a general nonexistence result for that problem.

For a canonical mean, a continuous positive multiplier $m(s)$ gives the decreasing loss

$$\ell(s) = \int_s^1 m(u)\, du, \qquad -\ell'(s) = m(s), \tag{27}$$

provided the integral is finite. This adds a score-dependent multiplier when applied before reduction. It cannot realize arbitrary metadata or norm-based weights: equal scores necessarily receive equal multipliers.

B Scope, statistics, and boundary cases

B.1 Smoothing and empty labels

The smoothed score $D_\epsilon = (2T + \epsilon)/(A + B + \epsilon)$ equals

$$D_\epsilon = \frac{2 + \epsilon/T}{P^{-1} + R^{-1} + \epsilon/T}.$$

Scaling the counts preserves P, R but changes ϵ/T. Counts $(1, 2, 2)$ and $(2, 4, 4)$ both give $P = R = 1/2$; with $\epsilon = 1$, their scores are $3/5$ and $5/9$. Thus the score is not a universal function of unsmoothed P, R across scales. For fixed target mass B, substituting $T = BR$ gives a B-dependent two-variable score whose derivatives can be analyzed separately.

Recall is undefined for $B = 0$. Ignoring empty items, smoothing, and assigning a separate penalty define different objectives. At $T = 0$, the interior ratio representation also fails. All-foreground targets give $P = 1$ and fall outside the stated open domain. Boundary powers and logarithms can have unbounded derivatives. The synthetic protocol excludes these cases.

B.2 Pooled-count objectives

For fixed positive class or item weights α_i, consider

$$D_{\text{pool}} = \frac{2\sum_i \alpha_i T_i}{\sum_i \alpha_i (A_i + B_i)}. \tag{28}$$

This includes weighted pooling of the type used by Generalised Dice (Sudre et al., 2017). Put $Z_i = A_i + B_i$, $Z = \sum_i \alpha_i Z_i$, and $\eta_i = \alpha_i Z_i / Z$. Then $D_{\text{pool}} = \sum_i \eta_i D_i$ but

$$\nabla_\theta D_{\text{pool}} = \sum_i \eta_i \nabla_\theta D_i + \sum_i D_i \nabla_\theta \eta_i, \qquad \nabla_\theta \eta_i = \frac{\alpha_i}{Z} \nabla_\theta Z_i - \frac{\eta_i}{Z} \nabla_\theta Z. \tag{29}$$

Although $\sum_i \nabla_\theta \eta_i = 0$, the second term does not generally vanish. Prediction-dependent α_i require their derivatives too. In contrast, pooled statistics $U = T_*/A_*$ and $V = T_*/B_*$ still give the harmonic expression $2UV/(U + V)$; their parameter gradients include all pooling derivatives.

B.3 Interpreting the taxonomy

Table 1 classifies the stated formulas in their stated coordinates. The categories are not equivalence claims between losses in different statistics. The following definitions make the nonstandard rows explicit.

Alternative metric pairs. clDice combines topology precision and sensitivity harmonically (Shit et al., 2021). A boundary F_1 combines the precision and recall of its matching rule; Boundary IoU instead intersects and unions boundary-band sets (Cheng et al., 2021). These need not use the same pair. Balanced accuracy and geometric G-mean use sensitivity and specificity. Hard skeletons and matching do not automatically provide a differentiable parameter map.

Squared and fractal forms. Let $T = \sum_j p_j y_j$, $A_2 = \sum_j p_j^2$, and $B_2 = \sum_j y_j^2$. Squared Dice is $2T/(A_2 + B_2)$, as in V-Net (Milletari et al., 2016). It is harmonic in $(U, V) = (T/A_2, T/B_2)$, which need not lie in $(0, 1)^2$. These identities extend to positive pseudo-statistics, but are not ordinary precision–recall identities. The displayed fractal form (Diakogiannis et al., 2021) is an increasing transform of squared Dice. Averaging over depths or adding the complement produces a composition whose individual derivatives must be retained.

Conditional and additional-statistic objectives. Surface Dice (Nikolov et al., 2021) has $U = o_{pred}/A_{pred}$ and $V = o_{GT}/A_{GT}$, where o denotes the surface area within the chosen tolerance of the other surface. It equals $w_A U + (1 - w_A)V$, with $w_A = A_{pred}/(A_{pred} + A_{GT})$. If a differentiable surrogate varies w_A, its gradient includes $(U - V)\nabla_\theta w_A$. Hard-surface implementations require separate treatment. Soft-label Dice semimetrics (Wang et al., 2023) recover the standard soft Dice only in the specified hard-label restriction. Cross-entropy hybrids (Yeung et al., 2022) generally add information beyond two aggregate statistics. In panoptic evaluation (Kirillov et al., 2019), recognition quality is harmonic in instance precision and recall, whereas $PQ = SQ\,RQ$ adds $RQ\nabla_\theta SQ$ when a differentiable surrogate is used. Wasserstein Dice (Fidon et al., 2018) uses semantic transport information and is not assigned a universal ordinary precision–recall ratio here.

More than two statistics. For $\mathbf{m} = (m_1, \ldots, m_K)$, $\nabla_\theta S = \sum_k S_{m_k} \nabla_\theta m_k$. Positive partials define a coefficient ray with $K - 1$ relative degrees of freedom. A weighted quasi-arithmetic mean gives pairwise ratios $w_j \phi'(m_j)/(w_k \phi'(m_k))$. Parameter-space identification requires an injective metric-gradient map on the coefficient differences being compared. This chain-rule extension does not reduce arbitrary segmentation objectives to two statistics.

C Numerical protocol and validation limits

Figure 1 uses analytic gradients at $(P, R) = (1/4, 3/4)$ and maps $J_{\mathbf{m}}$ equal to the identity, $\left(\begin{smallmatrix} 1 & 0.95 \\ 0 & \sqrt{1-0.95^2} \end{smallmatrix}\right)$, and $\left(\begin{smallmatrix} 1 & 1 \\ 0 & 0.02 \end{smallmatrix}\right)$. Its reduction panel uses the exact two-item counterexample; the source is `experiments/update_bridge.py`. These are constructed illustrations, not training measurements.

The metric-plane figure is generated by `experiments/pr_vector_fields.py` using analytic score derivatives on a 9×9 grid in $[0.1, 0.9]^2$; central differences with step 10^{-6} check all six fields. Arrow lengths are normalized; colors share one scale for the Euclidean gradient norm. The source for the shared-parameter diagnostics is `experiments/validate.py`. Trial $t = 0, \ldots, 199$ uses seed $2027 + t$. NumPy's default generator draws $X \in \mathbb{R}^{8 \times 64 \times 12}$ from a standard normal distribution and adds one normally distributed offset per item and feature. It draws θ with independent normal entries of standard deviation 0.5. Item i has foreground probability $0.08 + 0.84i/7$ with independent labels, except that its first two labels are fixed to one and zero. This is a diagnostic distribution, not a model of real segmentation data.

With $\dot{p}_{ij} = p_{ij}(1 - p_{ij})x_{ij}$, compute

$$g_T = \sum_j y_j \dot{p}_j, \quad g_A = \sum_j \dot{p}_j, \quad g_P = (A g_T - T g_A)/A^2, \quad g_R = g_T/B.$$

Dice and IoU are also differentiated independently from counts. Central parameter differences use step 10^{-5} on the first ten trials. Relative vector residuals divide by the larger vector norm or 10^{-12}. Cosines are clipped to $[-1, 1]$. Additional item tests cover negative log-Dice, squared-residual Dice loss, and squared Tversky loss. Power-mean checks use seed 91027 and draw 100 points in $[0.15, 0.85]^2$ for each $q \in \{-1, -0.5, 0, 0.5, 1\}$ and $w \in \{0.2, 0.5, 0.8\}$ using partial-difference step 10^{-6}. Exponential means use $\tau = 3$, the same preferences and point range, and independent seed 20270926. The exact two-item predictions $(\theta_i, 1 - \theta_i)$ are checked by independent parameter differences at $(0.2, 0.8)$.

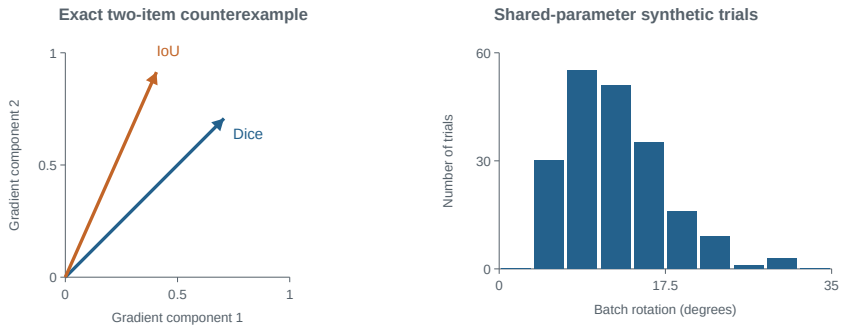


Figure 3: Left: exact two-item soft-count example. Right: batch rotation in 200 controlled shared-parameter trials. Paired item directions agree; unequal shaping multipliers can rotate their sum.

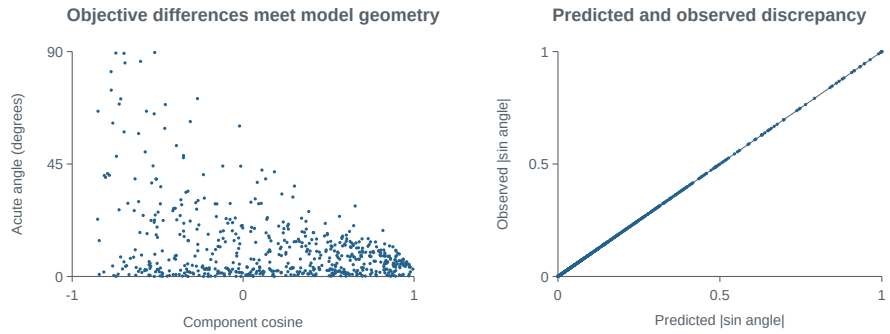


Figure 4: Synthetic common-state diagnostics. Left: $\arcsin|\sin\omega|$ versus the metric-gradient cosine; alignment and opposition are indistinguishable in this acute angle. Right: measured versus formula-predicted sine discrepancy. Every eighth comparison is plotted; all 4,800 enter the residual summary.

Further checks compare random linear-fractional score ratios with the no-go formula, and compare finite differences of normalized random vector mixtures with Equation (23). Their generator is reset to seed 91027 and advances past 900 point pairs to preserve the original diagnostic sample independently of power-grid extensions. The mixed difference of the counterexample's log-ratio tests Equation (26). Tolerances are asserted in code; summaries are saved as JSON and shared-parameter trial measurements as CSV. These summaries describe a fixed diagnostic sample without inferential confidence claims.

At all 1,600 item states, compare the three neutral arithmetic, geometric, and harmonic fields pairwise. The measured sine is the norm of the residual after projecting one normalized mixture onto the other; the predicted sine uses the Gram-determinant expression. These 4,800 rows are saved in `results/interaction_trials.csv`; the figure displays every eighth row. On the first ten trials, parameter differences for each neutral power mean are compared with $M_R(\Gamma_M g_P + g_R)$. The same differences for log-Dice and squared-residual Dice are negated to compare descent with score ascent. No additional random draws modify the original shared-parameter trials.

External validation remains unexecuted. Choose two accessible binary-segmentation datasets with different foreground distributions, fix an architecture and compute budget, and document access terms and source-level splits. Use at least three paired training seeds and validation-only tuning. At common checkpoints measure component and item gradients, batch rotations, and Jacobian conditioning. Compare power and exponential means with raw-gradient and separately labeled norm-matched controls. Norm matching is an intervention on the update and need not correspond to another scalar loss. Training trajectories are observations, not consequences of coefficient-ratio ordering alone. No benchmark results are supplied in this draft.

D ADDITIONAL DIAGNOSTIC RESULTS

For nonzero gradients we compare their cosine, norm ratio, and component cosine:

$$C_{\text{grad}} = \frac{\langle \nabla_\theta S_a, \nabla_\theta S_b \rangle}{\|\nabla_\theta S_a\| \|\nabla_\theta S_b\|}, \quad R_{\text{norm}} = \frac{\|\nabla_\theta S_a\|}{\|\nabla_\theta S_b\|}, \quad C_{PR} = \frac{\langle g_P, g_R \rangle}{\|g_P\| \|g_R\|}. \tag{30}$$

The code asserts nonzero denominators; none of the sampled comparisons is discarded. Loss descent is compared with score ascent. Table 2 summarizes residuals and a separate aggregation measurement.

Table 2: Executed synthetic checks. Maxima refer to the specified diagnostic sample.

Diagnostic	Observed value
Within-unit Dice–IoU: max. $\lvert 1 - \cos \rvert$	5.55×10^{-16}
Pairwise discrepancy: max. absolute sine error	1.80×10^{-15}
Inverse design: max. parameter FD relative error	1.08×10^{-10}
Power-mean field: max. relative error	1.30×10^{-9}
Shaped descent: max. $\lvert 1 - \cos \rvert$	3.33×10^{-16}
Item-wise Dice–IoU batch angle: median (degrees)	11.637

Table 3: Executed synthetic checks. Maxima refer to the specified diagnostic sample.

Diagnostic	Observed value
Chain-rule / Dice–IoU identity (max. relative error)	3.73×10^{-16}
Parameter finite differences (max. relative error)	8.33×10^{-11}
Shaped-loss finite differences (max. relative error)	6.30×10^{-11}
Single-item Dice–IoU (max. $\lvert 1 - \cos \rvert$)	5.55×10^{-16}
Post-reduction transform (max. $\lvert 1 - \cos \rvert$)	3.33×10^{-16}
Per-item IoU vs. Dice batch angle (median, degrees)	11.637
Per-item IoU vs. Dice batch angle (maximum, degrees)	31.067
Power-mean ratio (max. relative error)	1.30×10^{-9}
Exponential-mean ratio (max. relative error)	9.53×10^{-10}
Exact soft-count example (max. relative FD error)	5.87×10^{-11}
Linear-count ratio (max. relative error)	9.05×10^{-10}
Direction sensitivity (max. relative error)	4.46×10^{-10}

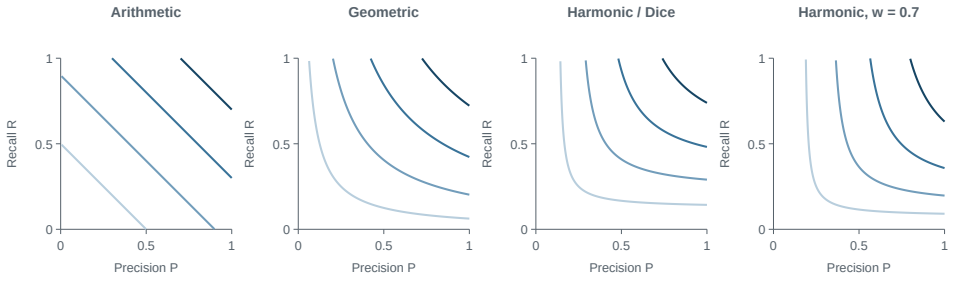


Figure 5: Formula-generated level sets at scores $0.25, 0.45, 0.65, 0.85$. Changing the power changes contour geometry; changing w changes diagonal preference. Dice and its IoU transform share contours after relabeling their levels. Axes use the same precision–recall coordinates.

E A VISUAL ATLAS OF THE UNIFIED FAMILY

The atlas is generated from the formulas by the accompanying Python script for Figure 6. It is a representation diagram, not an empirical performance comparison. Panel A covers the power-mean

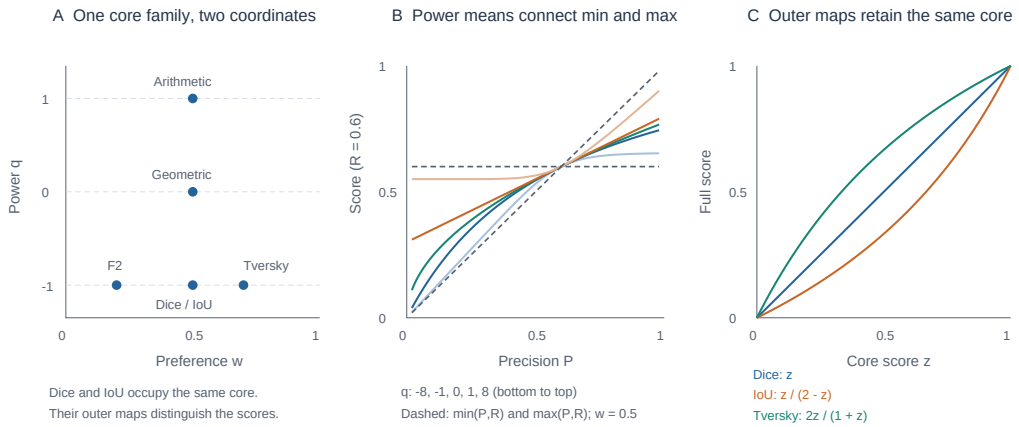


Figure 6: Three views of the scale-free subfamily. **A:** core coordinates (w, q) identify arithmetic, geometric, and harmonic aggregation. Dice and IoU occupy the same core point; F_2 has $w = 1/5$, and the marked Tversky core has $w = 0.7$. **B:** at $R = 0.6$ and $w = 1/2$, finite powers connect the harmonic, geometric, and arithmetic means between the min/max envelopes. **C:** a common harmonic core z yields Dice, IoU, and Tversky scores through different outer maps; the displayed Tversky map has $a = b = 1/4$. Core coordinates alone do not specify a full score.

subfamily; compatible non-scale-free generators, such as the exponential example in Equation (22), lie outside this two-coordinate map. Panel B uses powers $-8, -1, 0, 1, 8$, with geometric aggregation at zero. For every interior pair and fixed $0 < w < 1$, finite power means lie between $\min(P, R)$ and $\max(P, R)$; their limiting endpoints lose strict positive partials. Panel C fixes $w = 1/2$ and varies only ψ, making explicit a degree of freedom that the core map cannot display.