ITM: INDUCED TCHEBYCHEFF MEASURES FOR WEIGHT DESIGN

Anonymous authors
Paper under double-blind review

ABSTRACT

Uniform Tchebycheff weights need not produce uniform coverage of a Pareto front. We analyse the distribution induced by the weight-to-optimum map and its relation to front coverage. We introduce ITM (Induced Tchebycheff Measures), which estimates the solutions selected by candidate weights from a non-dominated archive. Three alternative constructions use isotonic inversion, pullback k-center, or density flattening to propose new weights, with updates screened by an archive-based covering score. Across 13 analytically specified fronts, covering gains from isotonic inversion and density flattening correlate with departure from uniform arclength, with Spearman coefficients of 0.846 and 0.739, respectively. For isotonic weights, a similar association appears across seven front settings under preference-conditioned network training. Oracle comparisons in MOEA/D show that coverage at subproblem optima and performance after search can respond differently to the same weight set.

1 INTRODUCTION

Tchebycheff scalarisation covers a Pareto front by solving one subproblem per weight, both in decomposition-based optimization (Zhang & Li, 2007) and in preference-conditioned learning (Sener & Koltun, 2018; Lin et al., 2019; Navon et al., 2021). The default lattice is uniform on

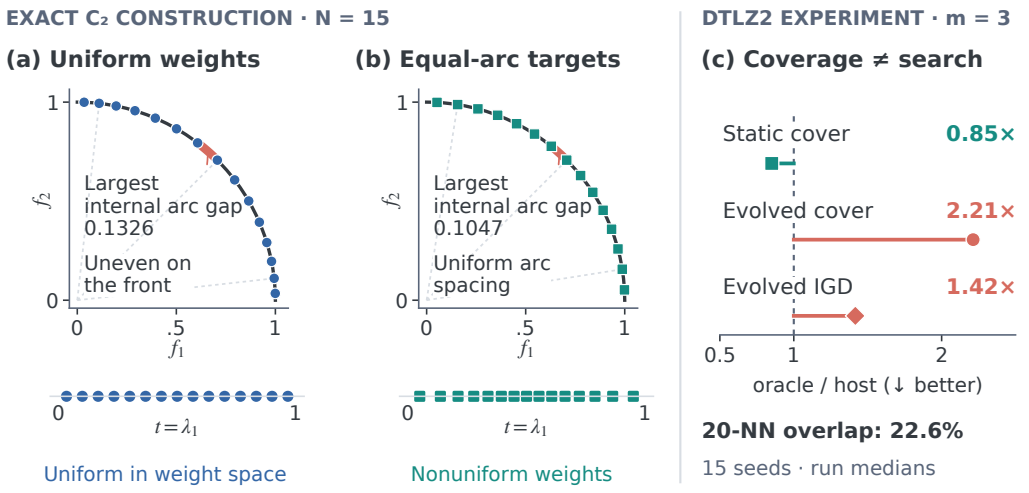

Figure 1: **Geometric coverage and evolutionary performance can move in opposite directions.** (a,b) Exact two-objective C_2 constructions with 15 midpoint weights or equal-arclength targets. Dashed rays follow $1/\lambda$; highlighted arcs show internal gaps between consecutive points, measured in arclength. (c) A separate three-objective DTLZ2 experiment using a true-front oracle lattice. Ratios are oracle/host; evolved values are ratios of medians over 15 seeds. The 22.6% overlap compares fixed-index 20-neighbour sets. Static coverage improves while evolved covering and IGD worsen (Tables 5–6).

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

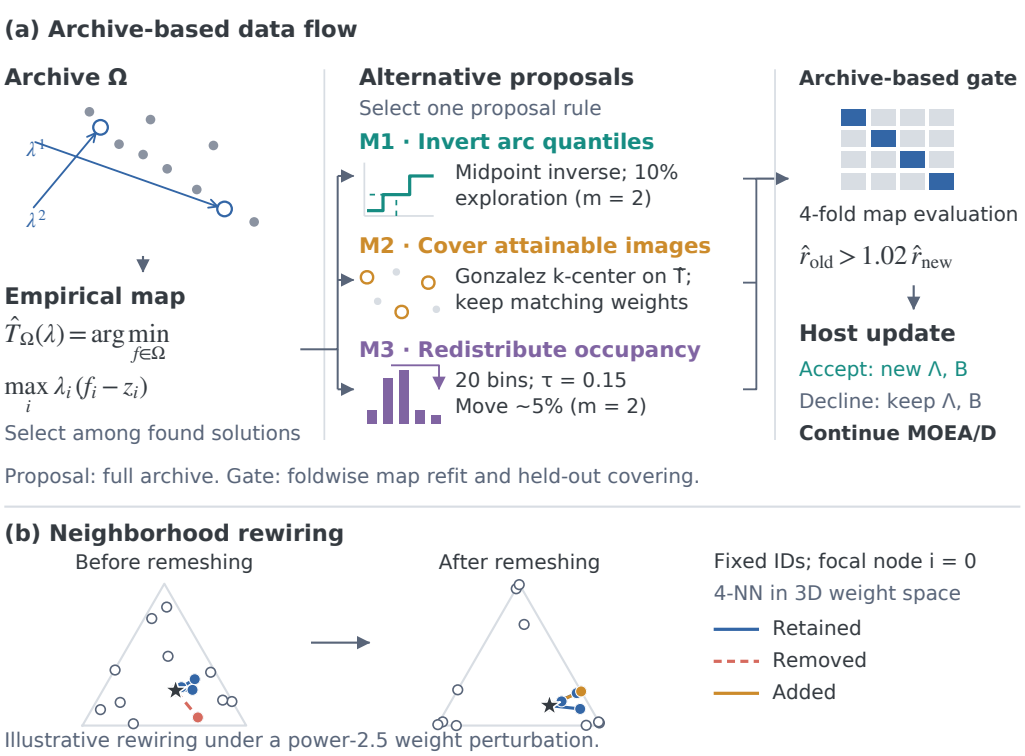

Figure 2: **Archive-based proposals and the search structure changed by remeshing.** (a) Each weight selects an attained archive point. M1, M2, and M3 are alternative proposal rules. Proposals use the full archive; the gate recomputes the empirical map on each training fold and scores held-out distances. A declined proposal retains the current state. (b) Illustrative neighbourhood rewiring: 15 Dirichlet weights (seed 7), a coordinate-wise power-2.5 perturbation, and four nearest neighbours computed in three-dimensional weight space. Node identities and the focal node are fixed.

the simplex (Das & Dennis, 1998), but the weight-to-optimum map T can produce a non-uniform distribution on the front. This distortion persists even with exact subproblem solutions. Figure 1 illustrates the resulting coverage pattern and its relationship to evolutionary search.

Existing adaptations use crowding, estimated geometry, or stagnation to relocate weights (Qi et al., 2014; Li & Yao, 2020; Panichella, 2019; Han et al., 2025). Weight design depends on how the scalariser maps weights to solutions. ITM estimates this correspondence from a non-dominated archive and constructs weights to improve coverage of the estimated image. The same update changes the neighbourhoods used by MOEA/D for mating and replacement, so we evaluate geometric coverage and the subsequent search separately.

Contributions.

- **Density and coverage analysis.** We give the weight-to-front map and its density on C_p, then analyse the covering cost of non-uniform sampling (Section 4).
- **Archive-based weight design.** ITM constructs weights by inverting an estimated arclength map, selecting separated solution images, or moving weights toward coverage holes. An archive-based score determines whether to accept each update (Figure 2; Section 5).
- **Geometric and search outcomes.** On analytically specified fronts, covering gains from isotonic inversion and density flattening correlate with induced-distribution mismatch. Oracle and online comparisons measure coverage, IGD and neighbourhood changes in MOEA/D (Section 6).

2 Related Work

Adaptive weights and front geometry. Weight adaptation uses population distribution, objective scales, or front geometry (Ma et al., 2020; Li, 2024). MOEA/D-AWA and AdaW relocate, add or delete weights using the distribution of solutions (Qi et al., 2014; Li & Yao, 2020). Reference-vector approaches adapt to objective scales (RVEA) or redistribute directions (M2M) (Cheng et al., 2016; Liu et al., 2014); SSWA and ATM regulate update timing (Han et al., 2024; 2025). Geometry estimates include the front exponent in AGE-MOEA (Panichella, 2019; 2022) and the Kriging model in SRVA (Namura, 2021). ITM bases its proposals on the archive solutions selected by the scalariser. We also compare with a fitted-shell method (Appendix E).

Decomposition and sampling. Das–Dennis lattices specify uniform spacing in weight space (Das & Dennis, 1998). Generalised decomposition constructs weights for target points (Giagkiozis et al., 2013), and direction constraints provide explicit subproblem geometry (Ma et al., 2018). The C_p correspondence is implicit in adaptive-weight analysis (Qi et al., 2014); we make its push-forward density and covering consequences explicit. Normalization handles objective scales (Ishibuchi et al., 2017). Runtime analysis studies how MOEA/D recovers optima (Doerr et al., 2024).

Preference-conditioned learning. Multi-objective learning studies Pareto-stationary solutions and preference-conditioned solution maps (Sener & Koltun, 2018; Lin et al., 2019; Navon et al., 2021). EPO optimizes user-specified trade-offs (Mahapatra & Rajan, 2020); Pareto set learning models a continuous solution set for Bayesian optimization (Lin et al., 2022). Smooth Tchebycheff scalarisation supports gradient-based optimization (Lin et al., 2024). EPS evolves preference vectors (Ye et al., 2024b), while GPSL learns a distribution transformation by maximizing hypervolume (Ye et al., 2024a). ITM constructs weights for scalarised training; Section 6.4 tests these weights with fixed losses and a shared network.

3 Preliminaries

Consider minimizing $f : \mathcal{X} \to \mathbb{R}^m$. Let $\mathcal{P}$ denote the Pareto front, $\mathbf{z}$ the ideal point, and $\Delta_m = \{\boldsymbol{\lambda} \geq 0 : \sum_i \lambda_i = 1\}$ the weight simplex. MOEA/D assigns a weight $\boldsymbol{\lambda}$ to each Tchebycheff subproblem (Ehrgott, 2005; Zhang & Li, 2007):

$$g_{\mathrm{tch}}(\mathbf{x} \mid \boldsymbol{\lambda}, \mathbf{z}) = \max_i \lambda_i \big(f_i(\mathbf{x}) - z_i\big). \tag{1}$$

Write $T(\boldsymbol{\lambda})$ for its ideal objective-space solution when unique, Λ for a lattice, and Ω for an attained non-dominated archive. The host uses Euclidean nearest-weight neighbourhoods $B(i)$ in Δ_m. Coverage measures the largest gap, $r(S, \mathcal{P}) = \sup_{f \in \mathcal{P}} \min_{s \in S} \|s - f\|_2$; For a reference set R, inverted generational distance (IGD) is $\mathrm{IGD}(S, R) = |R|^{-1} \sum_{f \in R} \min_{s \in S} \|s - f\|_2$ (Coello Coello & Lecante, 2004; Ishibuchi et al., 2019). We evaluate ideal images $T(\Lambda)$ and the solutions reached by a solver with these indicators.

4 Geometric Mismatch in Decomposition

We first study the canonical fronts

$$C_p = \{f \in \mathbb{R}^m_{\geq 0} : \textstyle\sum_i f_i^p = 1\}, \qquad p > 0, \tag{2}$$

linear at $p = 1$, sphere-like for $p > 1$, and convex for $p < 1$. Write $C_p^\circ = C_p \cap \mathbb{R}^m_{>0}$ for its positive-coordinate part. Complete proofs are in Appendix D.

4.1 The weight map and its induced density

Lemma 1 (Weight-to-front map)**.** *Let* $\mathbf{z} = \mathbf{0}$ *and* $\boldsymbol{\lambda} \in \operatorname{relint} \Delta_m$. *The unique minimizer of* $g_{\mathrm{tch}}(\cdot \mid \boldsymbol{\lambda}, \mathbf{0})$ *on* C_p *is* $T_p(\boldsymbol{\lambda})$ *with*

$$T_p(\boldsymbol{\lambda})_i = \frac{\lambda_i^{-1}}{\big(\sum_{j=1}^m \lambda_j^{-p}\big)^{1/p}}, \qquad i = 1, \ldots, m. \tag{3}$$

In particular T_p is a diffeomorphism from $\operatorname{relint}\Delta_m$ *onto* C_p°*, and for* $m = 2$ *with* $\boldsymbol{\lambda} = (t, 1-t)$ *and* $p = 2$ *it reduces to* $T_p(t) = \big((1-t),\, t\big)/\sqrt{t^2 + (1-t)^2}$.

The pushforward $\mu = (T_p)_{\#}\mathrm{Unif}(\Delta_m)$ is the distribution of $T_p(\boldsymbol{\lambda})$ when $\boldsymbol{\lambda}$ is uniform on Δ_m. Write ρ for its density with respect to arclength $d\sigma$ at $m = 2$ or $(m-1)$-dimensional Hausdorff measure at $m > 2$. Define the density distortion

$$\Delta_p = \sup_{C_p} \rho / \inf_{C_p} \rho \in [1, \infty]. \tag{4}$$

Proposition 1 (Density mismatch)**.** *For* $m = 2$ *and* $\boldsymbol{\lambda} = (t, 1-t)$ *with* $t \sim \mathrm{Unif}(0,1)$*, the pushforward density with respect to arclength satisfies* $\rho(t) \propto \|T_p'(t)\|_2^{-1}$*, hence* $\Delta_p = \sup_t \|T_p'(t)\|_2 \,/\, \inf_t \|T_p'(t)\|_2$*. Moreover:*

(i) $\Delta_1 = 1$*: uniform weights are exactly measure-preserving on the linear front.*

(ii) $\Delta_2 = 2$ exactly. *Parametrising* C_2 *by polar angle* θ*, the induced density is* $\rho(\theta) \propto t^2 + (1-t)^2$ *with* $\theta = \arctan\big(t/(1-t)\big)$*, whose range on* $t \in [0,1]$ *is* $[\frac{1}{2}, 1]$.

(iii) $\Delta_p > 1$ *for every* $p \neq 1$*. For* $0 < p < 1$*,* $\Delta_p = \infty$ *because the induced density tends to zero at the endpoints.*

Uniform weights crowd the axes of C_2 and undersample the endpoints of $C_{0.5}$; Figure 6 shows both patterns.

4.2 Covering bounds and the cost of non-uniformity

For a front of arclength L, abbreviate $r(S, \mathcal{P})$ as $r(S)$. Here $\mathrm{IGD}(S, \mathcal{P})$ averages distance over normalized arclength.

Theorem 1 (Near-optimality of equal arclength)**.** *Let* $\mathcal{P} \subset \mathbb{R}^2$ *be a connected, rectifiable, coordinatewise-monotone Pareto front of arclength* L*, and let* $S \subset \mathcal{P}$*. Then for* every S *with* $|S| = N$*,*

$$r(S) \;\geq\; \frac{L}{2\sqrt{2}\,N}, \qquad \textit{and} \qquad \mathrm{IGD}(S, \mathcal{P}) \;\geq\; \frac{L}{4\sqrt{2}\,N}, \tag{5}$$

while the equal-arclength set $S^\star$ *placing points at arclength coordinates* $(2k-1)L/(2N)$*,* $k = 1, \ldots, N$*, satisfies*

$$r(S^\star) \;\leq\; \frac{L}{2N}, \qquad \textit{and} \qquad \mathrm{IGD}(S^\star, \mathcal{P}) \;\leq\; \frac{L}{4N}. \tag{6}$$

Consequently $S^\star$ *is within a factor* $\sqrt{2}$ *of optimal for both indicators, uniformly in* N*,* L*, and front shape.*

The factor $\sqrt{2}$ follows from the chord–arc inequality for monotone curves. Branchwise quotas extend the construction to disconnected fronts (Corollary 1); Theorem 4 gives a finite-net bound for $m \geq 3$.

Theorem 2 (Mismatch penalty)**.** *Let* ρ *be an arclength density on* $\mathcal{P}$ *with* $\int_{\mathcal{P}} \rho\, d\sigma = 1$*, and let* S_N^ρ *place* N *points at the quantiles of* ρ*. Then as* $N \to \infty$*,*

$$\frac{r(S_N^\rho)}{L/(2N)} \;\longrightarrow\; \sup_{\mathcal{P}} \frac{1}{L\rho}, \qquad \frac{\mathrm{IGD}(S_N^\rho, \mathcal{P})}{L/(4N)} \;\longrightarrow\; \frac{1}{L^2} \int_{\mathcal{P}} \frac{d\sigma}{\rho}. \tag{7}$$

Both limits are ≥ 1*, with equality if and only if* $\rho \equiv 1/L$ *almost everywhere.*

On C_2, the asymptotic penalties are $4/\pi \approx 1.27$ for covering and $(1 + \pi/2)/(\pi/2)^2 \approx 1.04$ for IGD. Finite-point measurements are 1.28 and 1.06 (Table 7). Thus coverage and average distance respond differently even before search begins. The density distortion Δ_p compares the largest and smallest densities, whereas the penalties compare sampling to uniform arclength.

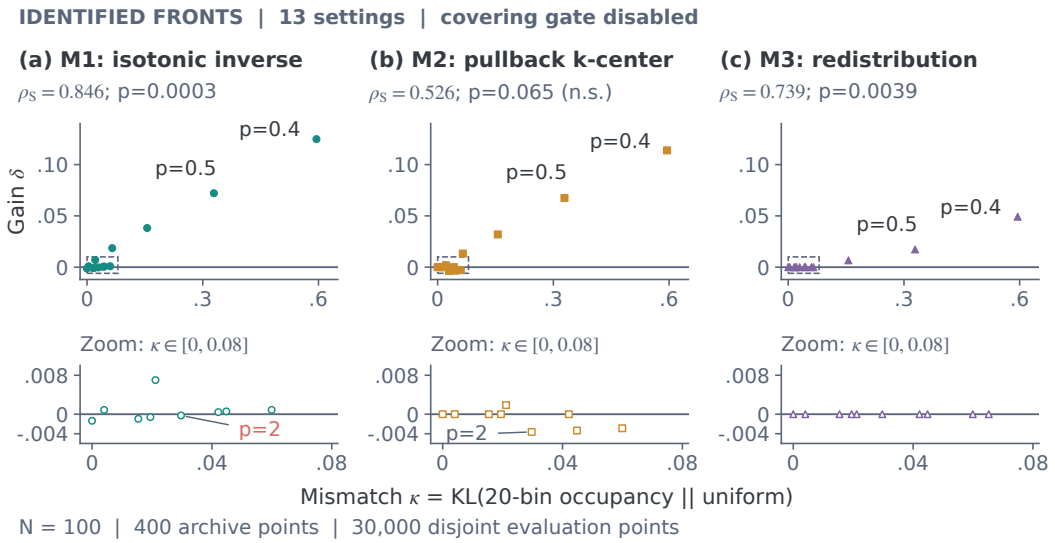

Figure 3: **Mismatch and covering gain on identified fronts.** Covering gains of M1, M2 and M3 across 13 C_p settings. Positive δ denotes improved covering; lower panels enlarge the dashed regions. Panel annotations report Spearman correlations. Each proposal uses $N = 100$ weights and a 400-point equal-arclength archive, and is evaluated on a disjoint 30,000-point reference with the covering gate disabled.

4.3 Geometric coverage and host dynamics

Proposition 2 (Host IGD under remeshing)**.** *Let Φ_Λ be one generation of MOEA/D with lattice Λ (neighbourhoods $B_\Lambda(i)$ and Tchebycheff replacement on $B_\Lambda(i)$). The covering radius $r(T(\Lambda), \mathcal{P})$ does not determine* $\mathrm{IGD}(\Phi_\Lambda(X), \mathcal{P})$*. In particular, if $X \neq T(\Lambda)$, a remesh to Λ^* with $r(T(\Lambda^*), \mathcal{P}) < r(T(\Lambda), \mathcal{P})$ can increase the evolutionary covering radius and the IGD of the subsequent trajectory.*

Figure 4 and Section 6.3 compare static and evolved outcomes for the same oracle lattice.

5 ITM: Archive-Based Weight Design

ITM estimates which archive solution each weight selects. One of three rules, M1–M3, proposes new weights, and an archive-based covering score determines acceptance (Figure 2).

5.1 Estimating the weight-to-solution map

At a checkpoint, ITM forms the empirical map on the current archive:

$$\hat{T}(\boldsymbol{\lambda}) = \arg\min_{f \in \Omega} \max_i \lambda_i (f_i - z_i). \tag{8}$$

The inverse $\lambda_i \propto 1/(f_i - z_i)$ generates candidates (Lemma 2). Candidates are evaluated through the estimated map $\hat{T}$. Objectives are normalized by the current ideal point and archive extent before proposal construction and scoring. Appendix C gives additional estimator properties.

5.2 Three lattice constructions

M1: isotonic reparametrisation. For $m = 2$, let s be arclength on Ω and $\psi = s \circ \hat{T}$. An oriented pool-adjacent-violators fit (de Leeuw et al., 2009) is inverted at equal-arclength quantiles; plateaus use midpoint inverses. A long archive chord is treated as a disconnected gap only when its midpoint is dominated. M1 retains a fixed 10% explorer fraction and keeps the lattice unchanged at $m \geq 3$.

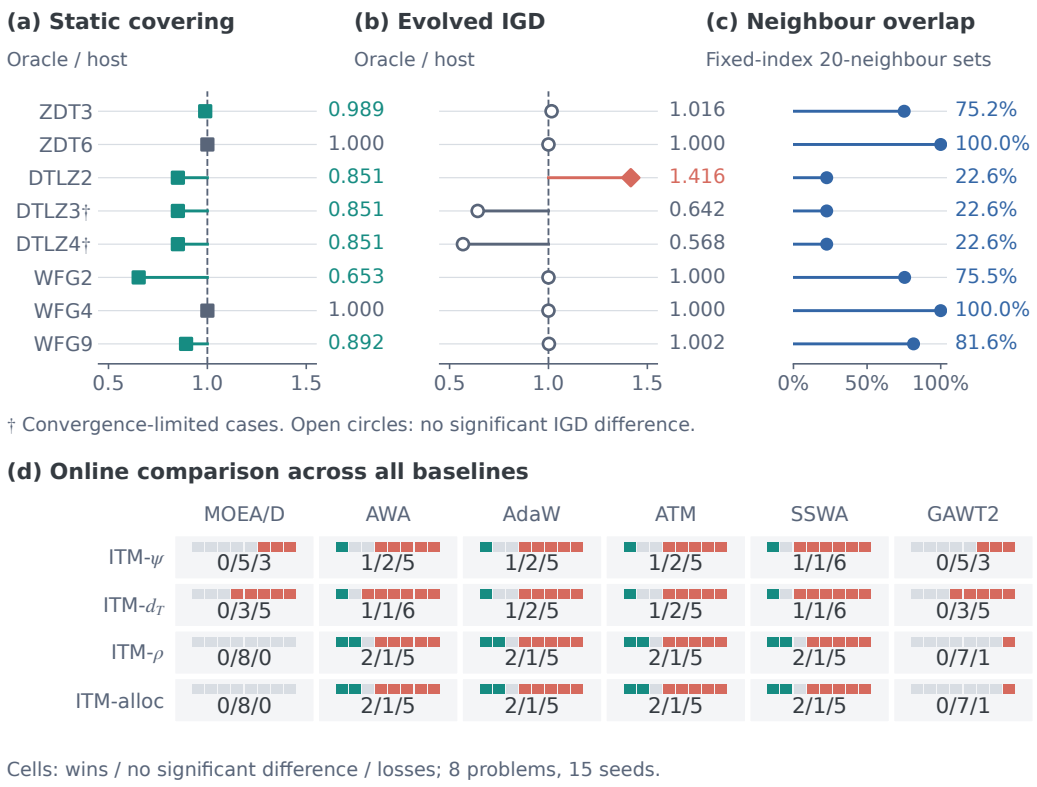

Figure 4: **Static coverage does not determine finite-budget search performance.** (a) Covering of ideal images $T(\Lambda)$ and (b) evolved IGD, both as oracle/host ratios on the same scale. Evolved ratios divide run medians over 15 seeds. The filled diamond marks significant DTLZ2 degradation; open circles indicate no detected IGD difference. Daggers mark convergence-limited DTLZ3/4 cases. (c) Fixed-index overlap of 20-neighbour sets. (d) Online ITM against all six baselines: each cell gives wins/no significant differences/losses on eight problems; green, grey, and coral blocks encode those outcomes. Each evolutionary comparison uses 15 seeds. All tests use uncorrected Wilcoxon rank-sum at $\alpha = 0.05$.

Proposition 3 (Quantile inversion)**.** *Let $\mathcal{P} \subset \mathbb{R}^2$ be a connected, rectifiable monotone front of arclength L, and let $T : \operatorname{relint} \Delta_2 \to \mathcal{P}^\circ$ be a homeomorphism, where $\mathcal{P}^\circ$ excludes the two endpoints. Write s for arclength and $\psi = s \circ T$. The lattice $\lambda_k = \psi^{-1}((k - \frac{1}{2})L/N)$, $k = 1, \ldots, N$, uniquely minimizes arclength covering radius, up to labelling, at $L/(2N)$. Its image $S_N = \{T(\lambda_k)\}_{k=1}^N$ satisfies*

$$r(S_N, \mathcal{P}) \leq \sqrt{2} \inf_{S \subseteq \mathcal{P},\, |S|=N} r(S, \mathcal{P})$$

in the Euclidean metric. These are the targets of the unmixed M1 quantile construction with exact access to ψ^{-1}. On C_p, they form the equal-arclength set of Theorem 1.

M2: pullback k-center. Use $d_T(\lambda, \lambda') = \|\hat{T}(\lambda) - \hat{T}(\lambda')\|_2$. Candidates combine the current lattice, a Dirichlet sample, and inverse rays through archive points. Gonzalez selection (Gonzalez, 1985) on their images returns the Tchebycheff owner of each selected point; 10% explorers are retained.

Proposition 4 (Pullback Gonzalez)**.** *Let $C \subset \Delta_m$ be finite, $A = \hat{T}(C)$, and $1 \leq N \leq |A|$. Gonzalez selection of N images in A, with one owning weight in C for each image, returns Λ satisfying*

$$r(\hat{T}(\Lambda), A) \leq 2 \min_{S \subseteq A,\, |S|=N} r(S, A).$$

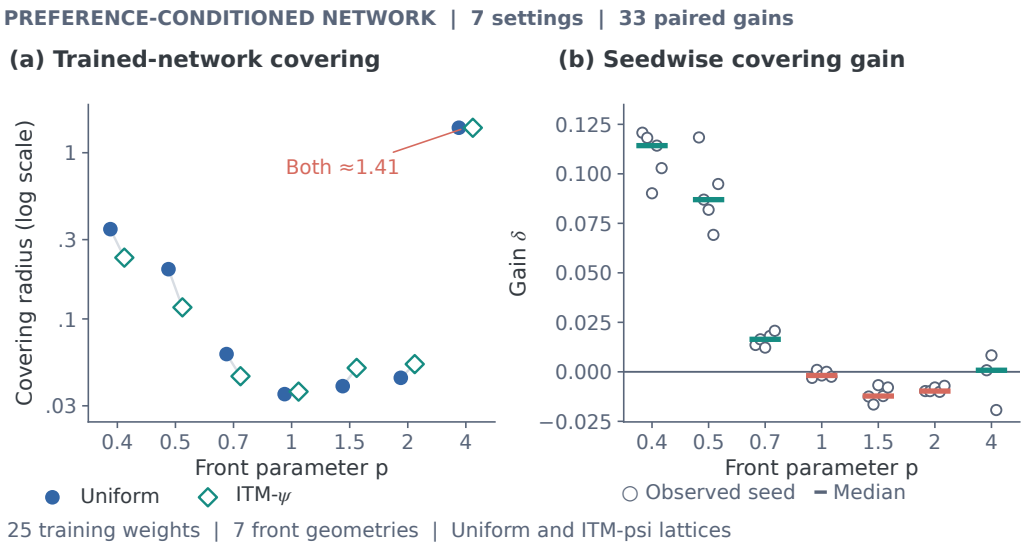

Figure 5: **Preference-conditioned learning on controlled two-loss fronts.** (a) Median covering radius across seven C_p settings, using 25 training weights; the vertical axis is logarithmic. Each median uses five values, except ψ at $p = 4$, which uses three. (b) The 33 available paired gains δ (open circles, horizontally offset by seed); bars show $\mathrm{median}_j(r_{\mathrm{uniform},j} - r_{\psi,j})$ over available pairs ($n = 5$ per setting, except $n = 3$ at $p = 4$). Positive gains improve covering. At $p = 4$, both covering radii are approximately 1.41.

M3: density flattening. For $m = 2$, let $\hat{q}$ be the 20-bin probability histogram of $s(\hat{T}(\lambda))$, with uniform bin masses $u_b = 1/20$. When $\mathrm{KL}(\hat{q}\|u) \geq \tau_{\mathrm{KL}} = 0.15$, move 5% of non-corner weights from redundant images toward maximal coverage holes. Otherwise, and at $m \geq 3$, retain the lattice. Proposition 5 bounds the penalty when the density moves toward uniformity in the majorization order (Marshall et al., 2011).

5.3 Foldwise gate and host integration

For each of four interleaved folds, recompute $\hat{T}_K$ on training points K and score lattice W on held-out points S:

$$\hat{r}_K(W, S) = \max_{f \in S} \min_{\lambda \in W} \|f - \hat{T}_K(\lambda)\|_2.$$

Write $\hat{r}(W, S_{\mathrm{ho}})$ for the maximum score across valid held-out folds. A weight update is accepted if

$$\hat{r}(W_{\mathrm{cur}}, S_{\mathrm{ho}}) > \gamma \hat{r}(W_{\mathrm{new}}, S_{\mathrm{ho}}), \qquad \gamma = 1.02, \tag{9}$$

Folds unable to form $\hat{T}_K$ are excluded. The proposal uses the full archive before splitting; each scoring fold refits only the empirical map on its training points.

Theorem 3 (Host equivalence without accepted proposals)**.** *Let $\mathcal{E}$ be the event that* (9) *never triggers. On $\mathcal{E}$, the populations and evaluations of* ITM *are identical generation-for-generation to MOEA/D with the same seed and initial lattice. On the complementary event, every commit satisfies* (9) *on the held-out archive available at that instant.*

After an accepted update, ideal-point rollback restores weights but retains the population. The archive contains at most $3N$ points. When it exceeds this cap, axis-extreme points are pinned and remaining entries are selected by greedy k-center on normalized objective vectors. Each step adds the point farthest from the selected set in minimum Euclidean distance. Checkpoints occur every 20 generations after a 30-generation warm-up and stop after 85% of the evaluation budget. M1 and M2 accept at most three updates; M3 accepts at most eight, starting after half the budget. With $\delta_{\mathrm{stab}} = 0.05$, rollback occurs when ideal-point drift exceeds $3\delta_{\mathrm{stab}}$ of the objective extent. After acceptance, the host rebuilds weight-space neighbourhoods and reassigns incumbents from the population–archive pool. A rejected proposal leaves the current weights and population in place. These state transitions are summarized in Algorithm 1, where nd selects non-dominated points.

Algorithm 1 ITM-MOEA/D

Require: problem f, weights W_0, neighbours, budget B; estimator $E \in \{\psi, d_T, \rho\}$, γ, τ_{KL}
Ensure: approximate Pareto front

```
P ← initial population, A ← ∅, W ← W_0
for generation until B evaluations do
    One MOEA/D generation on (P, W); update archive A by k-center
    if checkpoint and ideal settled then
        Ω ← nd(P ∪ A); T̂(λ) ← arg min_{f∈Ω} max_i λ_i(f_i − z_i)
        W̃ ← E(T̂, W, Ω)
        if W̃ exists and r̂(W, S_ho) > γ · r̂(W̃, S_ho) then
            W ← W̃; rebuild neighbours; reassign P from P ∪ A
        end if
    end if
    if installed W and ideal drifted beyond 3δ_stab then
        W ← W_0
    end if
end for
return nd(P)
```

Allocation with fixed neighbourhoods. To isolate evaluation allocation, keep Λ_0 and $B(i)$ fixed. After half the budget, the same bi-objective KL trigger redirects 5% of visits from zero-deficit cells to the largest coverage deficits. At $m \geq 3$ the visit schedule is unchanged. Without a trigger, the trajectory equals the host.

6 Experiments

6.1 Experimental setup

We evaluate true-front covering and attained-solution IGD. Evolutionary runs use 15 seeds, $25{,}000$ evaluations, $N = 100$ for $m = 2$ or 105 for $m = 3$ (Das–Dennis $H = 13$), and matched DE/rand/1 operators within MOEA/D-DE (Li & Zhang, 2009). Baselines are MOEA/D, AWA, AdaW, ATM, SSWA and the shell comparator GAWT2. We report medians and uncorrected Wilcoxon rank-sum tests at $\alpha = 0.05$. In win/tie/loss counts, a tie denotes no statistically detected difference. Network runs use five seeds. Appendices F and B give implementation details and full numerical tables.

6.2 Geometry on analytically specified fronts

We test whether greater distribution mismatch predicts larger covering gains. The pre-registered predictor κ is the Kullback–Leibler (KL) divergence of the 20-bin arclength occupancy of $T(\Lambda_0)$ from uniform, evaluated on the true front. We use the 13 values of p in Table 2, 100 weights, a 400-point equal-arclength archive and a disjoint $30{,}000$-point reference. These tests disable the held-out covering gate. The criterion is a positive Spearman correlation ρ_S at 0.05 between κ and covering gain $\delta = r_{\mathrm{uniform}} - r_{\mathrm{proposal}}$, with relative gain at most 0.01 on the flat segment.

Across the 13 settings in Figure 3, M1 gives $\rho_S = 0.846$ ($p = 3 \cdot 10^{-4}$) and M3 gives $\rho_S = 0.739$ ($p = 4 \cdot 10^{-3}$). M2 gives $\rho_S = 0.526$ ($p = 0.065$). On the flat segment, M3 keeps the lattice unchanged and M1's relative change is -0.13. Table 2 includes an explorer ablation (Section 6.5).

6.3 From static geometry to evolutionary search

Controlled C_p search. We next test evolutionary search on ten-variable `CPFront`, with front C_p and $p \in \{0.4, 0.5, 0.7, 1, 1.5, 2, 4\}$. With the covering gate enabled, M1 improves IGD at $p = 0.4, 0.5, 0.7$, has no detected difference at $p = 1, 2$, and has higher IGD at $p = 4$. M3 improves at $p = 0.4, 0.5$ and has no detected difference at $p = 1, 2$. Across these settings, mismatch/IGD-gain correlations are $\rho_S = 0.775$ ($p = 0.041$) for M1 and $\rho_S = 0.802$ ($p = 0.030$) for M3. Table 1 compares baselines at $p = 0.4, 0.5, 1, 2$; AWA, AdaW and SSWA improve IGD at $p = 2$.

Table 1: IGD medians on evolutionary C_p, 15 seeds. $^+/^-$ are Wilcoxon vs MOEA/D.

Front	MOEA/D	ITM-psi	ITM-rho	AWA	AdaW	ATM	SSWA
cp0.4	2.815e-2	1.943e-2$^+$	8.415e-3$^+$	7.586e-3$^+$	4.712e-3$^+$	5.553e-3$^+$	4.561e-3$^+$
cp0.5	1.199e-2	9.535e-3$^+$	6.437e-3$^+$	5.544e-3$^+$	4.547e-3$^+$	4.994e-3$^+$	4.371e-3$^+$
cp1.0	3.574e-3	3.575e-3	3.574e-3	3.570e-3	3.603e-3$^-$	3.573e-3	3.574e-3
cp2.0	4.201e-3	4.201e-3	4.201e-3	4.185e-3$^+$	4.167e-3$^+$	4.197e-3	4.165e-3$^+$

Oracle and online comparisons. We also test ZDT3/6 (Zitzler et al., 2000), DTLZ2/3/4 at $m = 3$ (Deb et al., 2005), and WFG2/4/9 at $m = 2$ (Huband et al., 2006). Oracle M1/M2 use the true front; online variants use the current non-dominated archive.

Figure 4(a–c) compares the same oracle lattice before and after search. On DTLZ2, static covering improves from 0.158 to 0.135, while evolved covering changes from 0.108 to 0.239 and IGD increases by a factor of 1.416. The remesh retains 22.6% of fixed-index neighbours. ZDT3 and WFG2 also show improved static coverage without a significant IGD gain. Across eight problems, oracle IGD is $0/7/1$ wins/no significant differences/losses; DTLZ3/4 are convergence-limited cases.

Panel (d) compares all four online actions against six baselines. Relative to MOEA/D, M1 is $0/5/3$, M2 is $0/3/5$, and M3 and allocation are both $0/8/0$. M1's losses occur on ZDT3, ZDT6 and WFG4; M2 also has higher IGD on DTLZ2. Tables 4–6 provide the underlying counts and values; Appendix F specifies the screening protocol.

6.4 Preference-conditioned training

For shared-network training, the losses have the same C_p fronts: $L_1(\theta) = \theta_0$ and $L_2(\theta) = (1 - \theta_0^p)^{1/p} + \|\theta_{1:}\|_2^2$. We train one network $\theta(\lambda)$ with the Tchebycheff scalariser on a uniform or a ψ lattice of 25 weights. Covering is evaluated at the training preferences against 8,000 reference points. Appendix F gives the network and training settings.

Figure 5 reports 33 available paired gains across seven settings; the correlation of (κ, δ) across setting medians is $\rho_S = 0.893$ ($p = 6.8 \cdot 10^{-3}$). At $p = 4$, median covering radii are 1.414 (uniform, $n = 5$) and 1.413 (ψ, $n = 3$); Table 3 gives all medians. A separate L-BFGS-B check, initialized at T_p with residuals below 10^{-8}, gives 13-setting covering-gain correlations of 0.846 for M1 and 0.739 for M3.

6.5 Component analysis

Exploration and gating. Removing M1's 10% explorer mixture increases its geometry-test correlation to 0.879. At $p = 2$, its gain changes from -2.5664×10^{-4} to $+0.00088$, compared with $+0.00224$ for true equal-arclength inversion. For evolutionary C_p at $p = 0.4, 0.5, 1, 2$, disabling M3's KL threshold or held-out gate leaves IGD unchanged.

Allocation without remeshing. The allocation variant leaves the visit schedule unchanged on DTLZ2, WFG4 and WFG9. It redirects visits in the second half of the budget on ZDT3, ZDT6 and WFG2, with no significant IGD change. On ZDT6, the archive-based KL trigger activates at a true-front mismatch of $\kappa = 0.012$.

7 Conclusion

ITM uses an archive estimate of the weight-to-solution map to construct weights. The analysis relates the induced front distribution to coverage; across the analytical fronts tested, covering gains from isotonic inversion and density flattening correlate with departure from uniform arclength. In MOEA/D, weight updates also change mating and replacement neighbourhoods, so the geometry of subproblem optima alone does not determine search performance.

AI USE STATEMENT

Generative AI tools were used for code implementation and language polishing of the manuscript. The authors are responsible for the final manuscript, claims, code and artifacts.

REFERENCES

Julian Blank and Kalyanmoy Deb. pymoo: Multi-objective optimization in Python. *IEEE Access*, 8: 89497–89509, 2020. doi: 10.1109/ACCESS.2020.2990567.

Ran Cheng, Yaochu Jin, Markus Olhofer, and Bernhard Sendhoff. A reference vector guided evolutionary algorithm for many-objective optimization. *IEEE Transactions on Evolutionary Computation*, 20(5):773–791, 2016. doi: 10.1109/TEVC.2016.2519378.

Carlos A Coello Coello and Gregorio Toscano Lecante. A study of the parallelization of a coevolutionary multi-objective evolutionary algorithm. In *Mexican International Conference on Artificial Intelligence*, pp. 688–697, 2004.

Indraneel Das and John E Dennis. Normal-boundary intersection: A new method for generating the Pareto surface in nonlinear multicriteria optimization problems. *SIAM Journal on Optimization*, 8(3):631–657, 1998.

Jan de Leeuw, Kurt Hornik, and Patrick Mair. Isotone optimization in R: Pool-adjacent-violators algorithm (PAVA) and active set methods. *Journal of Statistical Software*, 32(5):1–24, 2009. doi: 10.18637/jss.v032.i05.

Kalyanmoy Deb, Lothar Thiele, Marco Laumanns, and Eckart Zitzler. Scalable test problems for evolutionary multiobjective optimization. In *Evolutionary Multiobjective Optimization*, pp. 105–145. Springer, 2005.

Benjamin Doerr, Martin S Krejca, and Noé Weeks. Proven runtime guarantees for how the MOEA/D computes the Pareto front from the subproblem solutions. *arXiv preprint arXiv:2405.01014*, 2024.

Matthias Ehrgott. *Multicriteria Optimization*. Springer, 2 edition, 2005. doi: 10.1007/3-540-27659-9.

Ioannis Giagkiozis, Robin C. Purshouse, and Peter J. Fleming. Generalized decomposition. In *Evolutionary Multi-Criterion Optimization*, volume 7811 of *Lecture Notes in Computer Science*, pp. 428–442. Springer, 2013.

Teofilo F. Gonzalez. Clustering to minimize the maximum intercluster distance. *Theoretical Computer Science*, 38:293–306, 1985. doi: 10.1016/0304-3975(85)90224-5.

Xiaofeng Han, Tao Chao, Ming Yang, and Miqing Li. A steady-state weight adaptation method for decomposition-based evolutionary multi-objective optimisation. *Swarm and Evolutionary Computation*, 89:101641, 2024. doi: 10.1016/j.swevo.2024.101641.

Xiaofeng Han, Xiaochen Chu, Tao Chao, Ming Yang, and Miqing Li. A weight adaptation trigger mechanism in decomposition-based evolutionary multi-objective optimisation. arXiv:2502.16481, 2025.

Simon Huband, Philip Hingston, Luigi Barone, and Lyndon While. A review of multiobjective test problems and a scalable test problem toolkit. *IEEE Transactions on Evolutionary Computation*, 10(5):477–506, 2006.

Hisao Ishibuchi, Ken Doi, and Yusuke Nojima. On the effect of normalization in MOEA/D for multi-objective and many-objective optimization. *Complex & Intelligent Systems*, 3:279–294, 2017. doi: 10.1007/s40747-017-0061-9.

Hisao Ishibuchi, Ryo Imada, Yu Setoguchi, and Yusuke Nojima. Modified distance calculation in generational distance and inverted generational distance. *IEEE Transactions on Evolutionary Computation*, 23(5):796–813, 2019.

Hui Li and Qingfu Zhang. Multiobjective optimization problems with complicated Pareto sets, MOEA/D and NSGA-II. *IEEE Transactions on Evolutionary Computation*, 13(2):284–302, 2009. doi: 10.1109/TEVC.2008.925798.

Ke Li. A survey of decomposition-based evolutionary multi-objective optimization: Part I—past and future. arXiv:2404.14571, 2024.

Miqing Li and Xin Yao. What weights work for you? adapting weights for any Pareto front shape in decomposition-based evolutionary multiobjective optimisation. *Evolutionary Computation*, 28 (2):227–253, 2020. doi: 10.1162/evco_a_00269.

Xi Lin, Hui-Ling Zhen, Zhenhua Li, Qingfu Zhang, and Sam Kwong. Pareto multi-task learning. In *Advances in Neural Information Processing Systems*, volume 32, pp. 12037–12047, 2019.

Xi Lin, Zhiyuan Yang, Xiaoyuan Zhang, and Qingfu Zhang. Pareto set learning for expensive multi-objective optimization. In *Advances in Neural Information Processing Systems*, volume 35, pp. 19231–19247, 2022.

Xi Lin, Xiaoyuan Zhang, Zhiyuan Yang, Fei Liu, Zhenkun Wang, and Qingfu Zhang. Smooth Tchebycheff scalarization for multi-objective optimization. In *Proceedings of the International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 30479–30509, 2024.

Hai-Lin Liu, Fangqing Gu, and Qingfu Zhang. Decomposition of a multiobjective optimization problem into a number of simple multiobjective subproblems. *IEEE Transactions on Evolutionary Computation*, 18(3):450–455, 2014.

Xiaoliang Ma, Qingfu Zhang, Guangdong Tian, Junshan Yang, and Zexuan Zhu. On Tchebycheff decomposition approaches for multiobjective evolutionary optimization. *IEEE Transactions on Evolutionary Computation*, 22(2):226–244, 2018.

Xiaoliang Ma, Yanan Yu, Xiaodong Li, Yutao Qi, and Zexuan Zhu. A survey of weight vector adjustment methods for decomposition-based multiobjective evolutionary algorithms. *IEEE Transactions on Evolutionary Computation*, 24(4):634–649, 2020. doi: 10.1109/TEVC.2020.2978158.

Debabrata Mahapatra and Vaibhav Rajan. Multi-task learning with user preferences: Gradient descent with controlled ascent in Pareto optimization. In *Proceedings of the International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, pp. 6597–6607, 2020.

Albert W. Marshall, Ingram Olkin, and Barry C. Arnold. *Inequalities: Theory of Majorization and Its Applications*. Springer, 2 edition, 2011. doi: 10.1007/978-0-387-68276-1.

Nobuo Namura. Surrogate-assisted reference vector adaptation to various Pareto front shapes for many-objective Bayesian optimization. In *IEEE Congress on Evolutionary Computation (CEC)*, pp. 901–908. IEEE, 2021.

Aviv Navon, Aviv Shamsian, Gal Chechik, and Ethan Fetaya. Learning the Pareto front with hypernetworks. In *International Conference on Learning Representations*, 2021.

Annibale Panichella. An adaptive evolutionary algorithm based on non-Euclidean geometry for many-objective optimization. In *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO)*, pp. 595–603, 2019.

Annibale Panichella. An improved Pareto front modeling algorithm for large-scale many-objective optimization. In *GECCO '22: Genetic and Evolutionary Computation Conference*, pp. 565–573. ACM, 2022. doi: 10.1145/3512290.3528732.

Yutao Qi, Xiaoliang Ma, Fang Liu, Licheng Jiao, Jianyong Sun, and Jianshe Wu. MOEA/D with adaptive weight adjustment. *Evolutionary Computation*, 22(2):231–264, 2014.

Ozan Sener and Vladlen Koltun. Multi-task learning as multi-objective optimization. In *Advances in Neural Information Processing Systems*, volume 31, pp. 525–536, 2018.

Rongguang Ye, Longcan Chen, Wei-Bin Kou, Jinyuan Zhang, and Hisao Ishibuchi. Pareto front shape-agnostic Pareto set learning in multi-objective optimization. In *IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pp. 4730–4736. IEEE, 2024a. doi: 10.1109/SMC54092.2024.10832020.

Rongguang Ye, Longcan Chen, Jinyuan Zhang, and Hisao Ishibuchi. Evolutionary preference sampling for Pareto set learning. In *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO)*, pp. 630–638. ACM, 2024b. doi: 10.1145/3638529.3654024.

Qingfu Zhang and Hui Li. Moea/d: A multiobjective evolutionary algorithm based on decomposition. *IEEE Transactions on Evolutionary Computation*, 11(6):712–731, 2007.

Eckart Zitzler, Kalyanmoy Deb, and Lothar Thiele. Comparison of multiobjective evolutionary algorithms: Empirical results. *Evolutionary Computation*, 8(2):173–195, 2000.

APPENDIX

The appendix contains density illustrations, numerical tables, estimator properties, proofs, and implementation details.

A SUPPLEMENTARY VISUAL EVIDENCE

Figure 6 shows the exact maps and induced densities for $C_{0.5}$, C_1 and C_2, including endpoint sparsity and axis crowding.

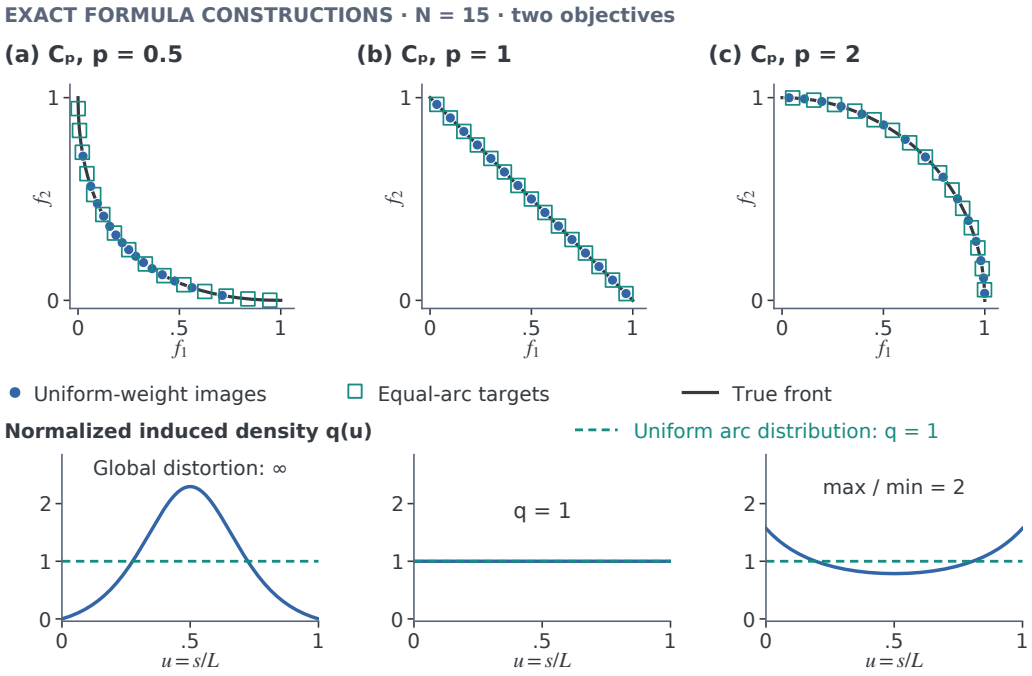

Figure 6: **The direction of sampling distortion depends on front geometry.** Analytical two-objective C_p maps with 15 midpoint weights. Upper panels compare uniform-weight images with equal-arclength targets obtained by high-resolution arc integration. Lower panels plot the normalized density $q(u) = L\rho(s)$ at $u = s/L$; uniform arclength has $q = 1$. For $p = 0.5$, the endpoint density tends to zero and the continuous global distortion is infinite. For the two-objective $p = 1$ segment, the point sets coincide; for $p = 2$, the density ratio is two.

B NUMERICAL RESULTS ACCOMPANYING THE FIGURES

Tables 2–6 report the geometry, network, online and oracle results using the protocols in Section 6.

For counts n_b in 20 equal-arclength bins, set $q_b = (n_b + 10^{-12})/\sum_j (n_j + 10^{-12})$ and $u_b = 1/20$. Then $\mathrm{KL}(q\|u) = \sum_b q_b \log(q_b/u_b)$, using natural logarithms. Geometry tests use true-front images of Λ_0; M3 uses archive images.

Table 2: Pre-registered synthetic test on C_p. δ is the covering improvement of the proposed lattice against the uniform lattice, scored on a $30{,}000$-point sample disjoint from the 400-point archive.

p	κ	δ_ψ	$\delta_{\psi 0}$	δ_ρ	δ_{d_T}
0.4	0.595	0.125	0.095	0.049	0.114
0.5	0.329	0.072	0.068	0.017	0.067
0.6	0.156	0.038	0.039	0.007	0.032
0.7	0.065	0.019	0.019	0.000	0.013
0.8	0.021	0.007	0.008	0.000	0.002
0.9	0.004	0.001	0.002	0.000	0.000
1	0.000	-0.001	-0.001	0.000	0.000
1.25	0.015	-0.001	-0.000	0.000	0.000
1.5	0.019	-0.001	0.000	0.000	0.000
2	0.030	-0.000	0.001	0.000	-0.004
3	0.042	0.000	0.001	0.000	0.000
4	0.045	0.001	0.002	0.000	-0.003
6	0.060	0.001	0.003	0.000	-0.003

Table 3: Preference-conditioned network on C_p. Covering radius of the trained models on the true front; δ is uniform minus ψ. Medians use five seeds, except the ψ covering and paired gain at $p = 4$, which use three available values.

p	κ	cover, unif.	cover, ψ	δ
0.4	0.595	0.347	0.234	0.114
0.5	0.329	0.199	0.117	0.087
0.7	0.064	0.061	0.045	0.016
1	0.000	0.035	0.037	-0.002
1.5	0.019	0.039	0.051	-0.012
2	0.030	0.044	0.053	-0.010
4	0.045	1.414	1.413	0.001

Table 4: ITM estimators on the 8-problem screen, 15 seeds, IGD Wilcoxon $W/T/L$ against the host, AWA, AdaW, ATM, SSWA, and the shell method.

Estimator	vs MOEA/D	vs AWA	vs AdaW	vs ATM	vs SSWA	vs GAWT2
ITM-psi	0/5/3	1/2/5	1/2/5	1/2/5	1/1/6	0/5/3
ITM-dT	0/3/5	1/1/6	1/2/5	1/2/5	1/1/6	0/3/5
ITM-rho	0/8/0	2/1/5	2/1/5	2/1/5	2/1/5	0/7/1
ITM-alloc	0/8/0	2/1/5	2/1/5	2/1/5	2/1/5	0/7/1

Table 5: Covering of $\{T(\lambda)\}$ on the true front, before search. Λ^* is the M1 (bi-objective) or M2 ($m = 3$) proposal from the true PF. B-overlap is the fraction of each 20-neighbour set shared with the uniform lattice under the identity index.

Problem	$r(T(\Lambda_0), P)$	$r(T(\Lambda^*), P)$	δ	B-overlap
zdt3	9.075e-2	8.975e-2	+9.947e-4	0.75
zdt6	8.164e-3	8.164e-3	+0.000e+0	1.00
dtlz2	1.582e-1	1.346e-1	+2.360e-2	0.23
wfg2	3.209e-2	2.094e-2	+1.115e-2	0.75
wfg4	1.210e-2	1.210e-2	+0.000e+0	1.00
wfg9	1.008e-2	8.997e-3	+1.087e-3	0.82

Table 6: True-front lattice versus the uniform host after search. Entries are median covering radius and IGD (host / oracle); marks are Wilcoxon of the oracle against the host.

Problem	cover host / oracle	IGD host / oracle
zdt3	9.043e-2 / 8.952e-2^{+}	1.205e-2 / 1.224e-2
zdt6	1.242e-2 / 1.242e-2	5.527e-3 / 5.527e-3
dtlz2	1.080e-1 / 2.390e-1^{-}	4.870e-2 / 6.897e-2^{-}
dtlz3	3.325e+0 / 2.112e+0	3.190e+0 / 2.048e+0
dtlz4	1.414e+0 / 1.414e+0	9.475e-1 / 5.379e-1
wfg2	3.569e-1 / 3.569e-1	4.081e-2 / 4.081e-2
wfg4	1.983e-2 / 1.983e-2	7.831e-3 / 7.831e-3
wfg9	8.734e-2 / 9.329e-2	2.423e-2 / 2.429e-2

C Supporting estimator properties

C.1 Scalarisation inverse

Lemma 2 (Exact weight inversion)**.** *Let* $\mathbf{z}$ *be the ideal point and let* $\mathcal{P}$ *be any Pareto front. Suppose* $\mathbf{f}^* \in \mathcal{P}$ *is non-dominated with* $f_i^* > z_i$ *for all* i*. The weight* $\boldsymbol{\lambda}^*$ *with* $\lambda_i^* \propto 1/(f_i^* - z_i)$ *makes* $\mathbf{f}^*$ *a minimiser of* $g_{\mathrm{tch}}(\cdot \mid \boldsymbol{\lambda}^*, \mathbf{z})$ *over* $\mathcal{P}$*. It is the only weight with that property, and* $\mathbf{f}^*$ *is the only minimiser, whenever no other point of* $\mathcal{P}$ *attains the same value of the inner maximum, which holds in particular on* C_p*.*

C.2 Density flattening

Proposition 5 (Flattening the penalty)**.** *On a connected monotone front, if* $\hat{\rho}$ *majorises* $\hat{\rho}'$ *(more mass at the mode, less in the tail), then* $\sup 1/(L\hat{\rho}') \le \sup 1/(L\hat{\rho})$ *and* $\int d\sigma/\hat{\rho}' \le \int d\sigma/\hat{\rho}$*. A move toward the uniform histogram cannot increase either penalty of Theorem 2.*

D Proofs

The chord–arc inequality below is used in Theorems 1 and 2. Throughout, a curve $\gamma : [0, L] \to \mathbb{R}^m$ is parametrised by arclength, so $\|\dot{\gamma}\|_2 = 1$ almost everywhere, and $\mathcal{P} = \gamma([0, L])$.

D.1 A chord–arc inequality for monotone fronts

Lemma 3 (Chord–arc)**.** *Let* $\gamma : [0, L] \to \mathbb{R}^2$ *be a rectifiable curve parametrised by arclength whose coordinate functions are monotone (one nonincreasing, one nondecreasing), the defining property of a two-objective Pareto front. Then for all* $s, s' \in [0, L]$*,*

$$\|\gamma(s) - \gamma(s')\|_2 \;\le\; |s - s'| \;\le\; \sqrt{2}\,\|\gamma(s) - \gamma(s')\|_2. \tag{10}$$

Both constants are optimal. Straight segments attain the left bound; strictly monotone two-segment curves approach the right bound.

Proof. Write $\gamma = (\gamma_1, \gamma_2)$ and assume without loss of generality $s < s'$.

The left inequality is the triangle inequality in integral form, $\|\gamma(s') - \gamma(s)\|_2 = \left\|\int_s^{s'} \dot{\gamma}\right\|_2 \le \int_s^{s'} \|\dot{\gamma}\|_2 = s' - s$.

For the right inequality, put $\Delta_i = |\gamma_i(s') - \gamma_i(s)|$ for $i = 1, 2$. Because each γ_i is monotone on $[s, s']$, its total variation equals Δ_i, that is $\int_s^{s'} |\dot{\gamma}_i| = \Delta_i$. Using $\sqrt{a^2 + b^2} \le |a| + |b|$ pointwise,

$$s' - s \;=\; \int_s^{s'} \sqrt{\dot{\gamma}_1^2 + \dot{\gamma}_2^2} \;\le\; \int_s^{s'} \big(|\dot{\gamma}_1| + |\dot{\gamma}_2|\big) \;=\; \Delta_1 + \Delta_2.$$

By Cauchy–Schwarz, $\Delta_1 + \Delta_2 \le \sqrt{2}\sqrt{\Delta_1^2 + \Delta_2^2} = \sqrt{2}\,\|\gamma(s') - \gamma(s)\|_2$. Combining the two displays gives the right inequality. To verify optimality of $\sqrt{2}$, take the polyline through $(0, 1 + \varepsilon)$,

$(1,1)$ and $(1+\varepsilon,0)$, where $0<\varepsilon<1$. Both coordinates are strictly monotone along this front. Its arclength-to-chord ratio between endpoints is $2\sqrt{1+\varepsilon^2}/(\sqrt{2}(1+\varepsilon))$, which tends to $\sqrt{2}$ as $\varepsilon \downarrow 0$. □

D.2 Proof of Lemma 1

Proof. Fix $\boldsymbol{\lambda} \in \operatorname{relint}\Delta_m$, so $\lambda_i > 0$ for every i, and recall $C_p = \{\mathbf{f} \in \mathbb{R}^m_{\geq 0} : \sum_i f_i^p = 1\}$ with $\mathbf{z} = \mathbf{0}$. We must minimise $g_{\mathrm{tch}}(\mathbf{f} \mid \boldsymbol{\lambda}, \mathbf{0}) = \max_i \lambda_i f_i$ over $\mathbf{f} \in C_p$.

Step 1: the optimum equalises the weighted coordinates. Let $\mathbf{f} \in C_p$ and suppose $\lambda_j f_j < \max_i \lambda_i f_i =: M$ for some j, attained at index k. Increase f_j by $\varepsilon > 0$ and decrease f_k so as to stay on C_p; this is feasible for small ε because $f_k > 0$ (if $f_k = 0$ then $M = 0$ and $\mathbf{f} = \mathbf{0} \notin C_p$). For ε small enough that $\lambda_j(f_j + \varepsilon) \leq M$, the new maximum is $\lambda_k f_k' < M$. Hence no minimiser can have unequal weighted coordinates, and at the optimum $\lambda_i f_i = c$ for all i and some $c > 0$.

Step 2: solve for c. From $f_i = c/\lambda_i$ and $\sum_i f_i^p = 1$ we get $c^p \sum_i \lambda_i^{-p} = 1$, hence $c = \left(\sum_i \lambda_i^{-p}\right)^{-1/p}$ and

$$T_p(\boldsymbol{\lambda})_i \;=\; \frac{\lambda_i^{-1}}{\left(\sum_j \lambda_j^{-p}\right)^{1/p}}. \tag{11}$$

The point is unique because Step 1 forces the equalisation and Step 2 determines c uniquely.

Step 3: bijectivity and smoothness. The map (11) is a composition of smooth positive operations on $\operatorname{relint}\Delta_m$, hence C^∞. It is injective: if $T_p(\boldsymbol{\lambda}) = T_p(\boldsymbol{\lambda}')$ then $\lambda_i^{-1}/\lambda_j^{-1} = \lambda_i'^{-1}/\lambda_j'^{-1}$ for all i, j, so $\boldsymbol{\lambda}' = \alpha\boldsymbol{\lambda}$ for some $\alpha > 0$, and $\sum_i \lambda_i = \sum_i \lambda_i' = 1$ forces $\alpha = 1$. It is surjective onto C_p°: given $\mathbf{f} \in C_p^\circ$, the weight $\lambda_i \propto 1/f_i$ (normalised) maps to $\mathbf{f}$ by direct substitution. The inverse is smooth by the same formula, so T_p is a diffeomorphism. □

D.3 Proof of Proposition 1

Proof. We compute the pushforward of the uniform distribution on Δ_m under T_p and measure its density with respect to arclength on C_p.

Case $m = 2$, $p = 2$ (exact). Parametrise the simplex by $\boldsymbol{\lambda} = (t, 1-t)$, $t \in (0,1)$, so $t \sim \mathrm{Unif}(0,1)$. By (11) the image direction is proportional to $(1/t,\ 1/(1-t)) = \big(t(1-t)\big)^{-1}\big(1-t,\ t\big)$, and after normalisation in the 2-norm,

$$T_p(t) \;=\; \frac{\big(1-t,\ t\big)}{r(t)}, \qquad r(t) \;:=\; \sqrt{t^2 + (1-t)^2}. \tag{12}$$

Since C_2 is the unit circle quadrant, write $T_p(t) = (\cos\theta, \sin\theta)$; then (12) gives $\tan\theta = t/(1-t)$, and arclength on the unit circle is exactly $\mathrm{d}\sigma = \mathrm{d}\theta$. Differentiating $\theta(t) = \arctan\big(t/(1-t)\big)$,

$$\frac{\mathrm{d}\theta}{\mathrm{d}t} \;=\; \frac{1}{1 + \left(\frac{t}{1-t}\right)^2} \cdot \frac{1}{(1-t)^2} \;=\; \frac{1}{t^2 + (1-t)^2} \;=\; \frac{1}{r(t)^2}.$$

Because t is uniform, the density of the image with respect to θ (equivalently with respect to arclength) is

$$\rho(\theta) \;=\; \left|\frac{\mathrm{d}t}{\mathrm{d}\theta}\right| \;=\; r(t)^2 \;=\; t^2 + (1-t)^2. \tag{13}$$

On $t \in [0,1]$ the quadratic (13) is minimised at $t = \frac{1}{2}$ with value $\frac{1}{2}$ and maximised at the endpoints $t \in \{0,1\}$ with value 1. Therefore

$$\Delta_2 \;=\; \frac{\sup\rho}{\inf\rho} \;=\; \frac{1}{1/2} \;=\; 2$$

exactly, which is claim (ii). The extremes occur at the axes ($t \to 0, 1$, density 1) and at the diagonal ($t = \frac{1}{2}$, density $\frac{1}{2}$), so uniform weights are twice as dense near the axes as near the diagonal. The finite-grid estimate on 20 000 points is 1.990 (Table 7).

Case $m = 2$, general p. The same computation with C_p in place of the circle gives image point $T_p(t) = \big(t^{-1}, (1-t)^{-1}\big)/\big(t^{-p} + (1-t)^{-p}\big)^{1/p}$, and the arclength density is

$$\rho_p(t) \;=\; \frac{1}{\big\|\frac{\mathrm{d}}{\mathrm{d}t} T_p(t)\big\|_2}, \tag{14}$$

because a unit of probability mass $\mathrm{d}t$ is spread over arclength $\|\frac{\mathrm{d}}{\mathrm{d}t} T_p(t)\|_2 \,\mathrm{d}t$. For $p = 1$ the front is the segment $f_1 + f_2 = 1$, on which (11) reduces to $T_p(t) = (1-t, t)$, a constant-speed parametrisation; hence ρ_1 is constant and $\Delta_1 = 1$, which is claim (i) and shows uniform weights give equal-arclength points on this two-objective segment. For $p > 1$, the speed tends to 1 at an endpoint and equals $2^{3/2-1/p} > 1$ at $t = 1/2$, so the density is nonconstant.

Endpoint behaviour for $0 < p < 1$. The equivalent expression $T_p(t) = (1-t, t)/((1-t)^p + t^p)^{1/p}$ gives $\|T_p'(t)\|_2 \sim t^{p-1}$ as $t \to 0^+$. The speed therefore diverges, whereas it is finite and positive at $t = 1/2$. Thus $\inf \rho_p = 0$ and $\Delta_p = \infty$ for every $0 < p < 1$, including $p = 0.5$. The value 24.41 in Table 7 is a different statistic: `theory.py` samples 20,000 weights on $[10^{-6}, 1 - 10^{-6}]$, estimates speed by consecutive chords, and clips speeds at their 0.5th and 99.5th percentiles before taking the clipped speed ratio. Figure 6 plots the unclipped normalized density and shows the vanishing endpoint density explicitly. □

D.4 Proof of Theorem 1

Proof. Parametrise $\mathcal{P}$ by arclength $s \in [0, L]$ via the bijection γ.

Upper bounds (S^*). The equal-arclength set $S^* = \{s_k^* = (2k-1)L/(2N) : k = 1, \dots, N\}$ partitions $[0, L]$ into N cells of length L/N, each centred on one s_k^*. For any $s \in [0, L]$, $\min_k |s - s_k^*| \le L/(2N)$. By Lemma 3 (left inequality),

$$r(S^*) \;=\; \sup_s \min_k \|\gamma(s) - \gamma(s_k^*)\|_2 \;\le\; \sup_s \min_k |s - s_k^*| \;\le\; \frac{L}{2N}.$$

For IGD, each cell contributes average arc-distance $L/(4N)$ to its midpoint (the average of $|s - c|$ over $s \in [c - L/(2N), c + L/(2N)]$ is $L/(4N)$), and by the same left inequality,

$$\mathrm{IGD}(S^*, \mathcal{P}) \;=\; \frac{1}{L}\int_0^L \min_k \|\gamma(s) - \gamma(s_k^*)\|_2 \,\mathrm{d}s \;\le\; \frac{1}{L}\int_0^L \min_k |s - s_k^*| \,\mathrm{d}s \;=\; \frac{L}{4N}.$$

Lower bounds. Write $r_{\mathrm{arc}}(S) = \sup_{s \in [0,L]} \min_j |s - s_j|$ for the covering radius measured in arclength. We first bound r_{arc} from below for every S, then transfer to Euclidean distance.

Let $S = \{s_1, \dots, s_N\}$ and order it as $s_{(1)} \le \cdots \le s_{(N)}$. Write $a_0 = s_{(1)} - 0$, $a_k = s_{(k+1)} - s_{(k)}$ for $1 \le k \le N-1$, and $a_N = L - s_{(N)}$, so that $\sum_{k=0}^{N} a_k = L$. The two boundary gaps and the $N-1$ interior gaps contribute to r_{arc} differently: the endpoint $s = 0$ is at arclength distance a_0 from all of S, the endpoint $s = L$ is at distance a_N, and the midpoint of an interior gap a_k is at distance $a_k/2$. Hence

$$r_{\mathrm{arc}}(S) \;=\; \max\Big\{a_0,\; a_N,\; \max_{1 \le k \le N-1} \tfrac{a_k}{2}\Big\}. \tag{15}$$

Put $\rho = r_{\mathrm{arc}}(S)$. Then $a_0 \le \rho$, $a_N \le \rho$, and $a_k \le 2\rho$ for every interior gap, so

$$L \;=\; a_0 + a_N + \sum_{k=1}^{N-1} a_k \;\le\; \rho + \rho + 2\rho(N-1) \;=\; 2\rho N,$$

that is, $r_{\mathrm{arc}}(S) \ge L/(2N)$ for every N-point set. Applying the right inequality of (10) pointwise, $\min_j \|\gamma(s) - \gamma(s_j)\|_2 \ge \min_j |s - s_j|/\sqrt{2}$, and taking the supremum over s gives

$$r(S) \;\ge\; \frac{r_{\mathrm{arc}}(S)}{\sqrt{2}} \;\ge\; \frac{L}{2\sqrt{2}\,N}.$$

The bound is attained in the arclength metric by the equal-arclength set, for which $a_0 = a_N = L/(2N)$ and $a_k = L/N$, so (15) is exactly $L/(2N)$; no placement does better, and the only loss in the Euclidean statement is the chord–arc factor.

For the IGD lower bound, let $\{I_k\}$ be the Voronoi cells of S on $[0, L]$ (each I_k is the interval of points closer to s_k than to any other s_j). We have $\sum_k |I_k| = L$. By the right inequality in (10), $\min_j \|\gamma(s) - \gamma(s_j)\|_2 \geq \min_j |s - s_j|/\sqrt{2} = d(s, S)/\sqrt{2}$, where $d(s, S)$ is the arclength nearest-neighbour distance. Therefore

$$\mathrm{IGD}(S, \mathcal{P}) \;\geq\; \frac{1}{\sqrt{2}\,L} \int_0^L d(s, S)\,\mathrm{d}s \;=\; \frac{1}{\sqrt{2}\,L} \sum_k \int_{I_k} |s - s_k|\,\mathrm{d}s \;\geq\; \frac{1}{\sqrt{2}\,L} \sum_k \frac{|I_k|^2}{4},$$

where the last step uses $\int_{I_k} |s - s_k|\,\mathrm{d}s \geq |I_k|^2/4$ (since the nearest point s_k is in I_k and the worst case is the cell boundary, giving average distance $|I_k|/4$). By the Cauchy–Schwarz inequality (or Jensen applied to x^2): $\sum_k |I_k|^2 \geq (\sum_k |I_k|)^2/N = L^2/N$. Hence $\mathrm{IGD}(S, \mathcal{P}) \geq L^2/(4\sqrt{2}\,L\,N) = L/(4\sqrt{2}N)$. □

D.5 Disconnected fronts

For disconnected fronts such as ZDT3, allocate weights to individual branches in proportion to their arclengths. The following corollary states the covering bounds for these branchwise allocations.

Corollary 1 (Disconnected monotone fronts)**.** *Let $\mathcal{P} = \bigsqcup_{c=1}^{C} \mathcal{P}_c$ be a disjoint union of rectifiable, coordinatewise-monotone arcs with arclengths $L_1, \ldots, L_C$ and $L = \sum_c L_c$. Then every S with $|S| = N$ satisfies $r(S) \geq L/(2\sqrt{2}N)$. Moreover, if the per-branch quotas N_c satisfy $\sum_c N_c = N$ and $N_c \geq N L_c/L$, and $S^\star$ places N_c equal-arclength points inside branch c, then $r(S^\star) \leq L/(2N)$. The same $\sqrt{2}$ sandwich therefore holds with L the* total *arclength.*

Proof. Lower bound. Let $N_c = |S \cap \mathcal{P}_c|$, so $\sum_c N_c = N$. Restricted to branch c, the argument of Theorem 1 applies verbatim, because $\mathcal{P}_c$ is connected and monotone: with $\rho = r_{\mathrm{arc}}(S)$ measured within branches, $L_c \leq 2\rho N_c$. Summing over c gives $L \leq 2\rho N$, hence $\rho \geq L/(2N)$. A branch holding no point of S has $N_c = 0$ and makes the inequality vacuous in the right direction. Transferring by the right inequality of (10), which holds within each branch and is trivial across branches because the concatenated arclength separation of two points on different branches never exceeds their Euclidean separation by more than the within-branch factor, gives $r(S) \geq L/(2\sqrt{2}N)$.

Upper bound. Inside branch c the equal-arclength set attains $r_{\mathrm{arc}} = L_c/(2N_c) \leq L/(2N)$ by the quota hypothesis, and the left inequality of (10) holds within a branch, so the Euclidean covering radius of $S^\star$ restricted to that branch is at most $L/(2N)$. Taking the maximum over branches preserves the bound. □

The upper bound uses the branch decomposition. Placing points at equal arclength along the *concatenated* parametrisation would straddle the jump between two branches, and there the left inequality of (10) fails: two points adjacent in concatenated arclength can be far apart in the objective space. This is why M1 inverts ψ on gap-aware arclength rather than on a concatenated polyline.

D.6 A shape-free covering bound for $m \geq 3$

For $m \geq 3$, we use a finite ε-net and farthest-point selection to obtain a covering bound. A volume argument then gives an absolute lower bound under a regularity condition; at $m = 2$ it recovers the lower-bound constant in Theorem 1.

Throughout, $\mathcal{P} \subset \mathbb{R}^m$ is compact, $r(S, \mathcal{P}) = \sup_{\mathbf{f} \in \mathcal{P}} \min_{\mathbf{s} \in S} \|\mathbf{f} - \mathbf{s}\|_2$, and

$$r_N^*(\mathcal{P}) \;=\; \inf\big\{ r(S, \mathcal{P}) \;:\; S \subseteq \mathcal{P},\; |S| = N \big\} \tag{16}$$

is the optimal N-point covering radius. A finite $C \subseteq \mathcal{P}$ is an *ε-net* if every $\mathbf{f} \in \mathcal{P}$ has some $\mathbf{c} \in C$ with $\|\mathbf{f} - \mathbf{c}\|_2 \leq \varepsilon$. Given a seed set $S_0 \subseteq C$ with $|S_0| = q$, the *greedy* selection G_N adds points one at a time, each time the element of C farthest from the current set, until $|G_N| = N$. In this paper the construction covers a known front: C is a dense sample of $\mathcal{P}$, and S_0 contains the m axis extremes.

Theorem 4 (Shape-free covering guarantee, any m)**.** *Let $C \subseteq \mathcal{P}$ be an ε-net and let G_N be the greedy selection from C with a seed set of size $q < N$. Then*

$$r_N^*(\mathcal{P}) \;\leq\; r(G_N, \mathcal{P}) \;\leq\; 2\, r_{N-q}^*(\mathcal{P}) + 3\varepsilon. \tag{17}$$

The constants 2 *and* 3 *do not depend on* m*, on* N*, or on the shape, smoothness, connectedness, or monotonicity of* $\mathcal{P}$*.*

Proof. The left inequality is the definition of r_N^*, since $G_N \subseteq \mathcal{P}$.

Write $\rho = \max_{\mathbf{c}\in C} \operatorname{dist}(\mathbf{c}, G_N)$ for the covering radius measured on the net, and let $\mathbf{x} \in C$ attain it.

Step 1 (greedy on the net). For every index $i > q$ the point g_i was chosen as the element of C farthest from G_{i-1}, so $\operatorname{dist}(g_i, G_{i-1}) \geq \operatorname{dist}(\mathbf{x}, G_{i-1}) \geq \operatorname{dist}(\mathbf{x}, G_N) = \rho$. Hence the $N - q + 1$ points $\{g_{q+1}, \ldots, g_N, \mathbf{x}\}$ are pairwise at distance at least ρ. Let $S' \subseteq C$ be any $(N-q)$-point set. By pigeonhole two of those $N - q + 1$ points share a nearest element of S', and since they are at least ρ apart, one of them is at distance at least $\rho/2$ from it. Therefore $\max_{\mathbf{c}\in C} \operatorname{dist}(\mathbf{c}, S') \geq \rho/2$, and taking the infimum over S',

$$\rho \;\leq\; 2\,\mathrm{OPT}_{N-q}(C), \qquad \mathrm{OPT}_k(C) = \min_{S'\subseteq C,\,|S'|=k} \max_{\mathbf{c}\in C} \operatorname{dist}(\mathbf{c}, S'). \tag{18}$$

The seed set costs exactly q centres: without pinning, $q = 1$ and (18) is Gonzalez's 2-approximation.

Step 2 (net to continuum). For $\mathbf{f} \in \mathcal{P}$ pick $\mathbf{c} \in C$ with $\|\mathbf{f} - \mathbf{c}\|_2 \leq \varepsilon$. Then $\operatorname{dist}(\mathbf{f}, G_N) \leq \varepsilon + \operatorname{dist}(\mathbf{c}, G_N) \leq \varepsilon + \rho$, so $r(G_N, \mathcal{P}) \leq \rho + \varepsilon$.

Step 3 (continuum to net). Fix $k = N - q$ and $\delta > 0$, and take $S^* \subseteq \mathcal{P}$ with $|S^*| = k$ and $r(S^*, \mathcal{P}) \leq r_k^*(\mathcal{P}) + \delta$. Replace each $\mathbf{s} \in S^*$ by some $c(\mathbf{s}) \in C$ within ε, giving $S' \subseteq C$ with $|S'| \leq k$. Any $\mathbf{c} \in C$ lies in $\mathcal{P}$, so it is within $r_k^* + \delta$ of some $\mathbf{s}$, hence within $r_k^* + \delta + \varepsilon$ of $c(\mathbf{s})$. Thus $\mathrm{OPT}_k(C) \leq r_k^*(\mathcal{P}) + \varepsilon + \delta$, and δ was arbitrary.

Chaining the three steps, $r(G_N, \mathcal{P}) \leq \rho + \varepsilon \leq 2\,\mathrm{OPT}_{N-q}(C) + \varepsilon \leq 2\big(r_{N-q}^*(\mathcal{P}) + \varepsilon\big) + \varepsilon$. □

Theorem 4 compares the computed set with an optimal $(N-q)$-point set. The next proposition gives an absolute lower bound in terms of front measure, dimension and a regularity constant.

Proposition 6 (Absolute covering lower bound)**.** *Let* $\mathcal{P}$ *be compact with* $\mathcal{H}^{m-1}(\mathcal{P}) = A > 0$*, and suppose it is upper regular with constant* C_{reg}*: for every* $\mathbf{f} \in \mathbb{R}^m$ *and every* $0 < \varrho \leq \operatorname{diam}(\mathcal{P})$*,* $\mathcal{H}^{m-1}\big(\mathcal{P} \cap B(\mathbf{f}, \varrho)\big) \leq C_{\mathrm{reg}}\,\omega_{m-1}\varrho^{m-1}$*, where* ω_{m-1} *is the volume of the unit ball of* $\mathbb{R}^{m-1}$*. Then*

$$r_N^*(\mathcal{P}) \;\geq\; \Big(\frac{A}{C_{\mathrm{reg}}\,\omega_{m-1}\,N}\Big)^{1/(m-1)}. \tag{19}$$

Proof. Let $|S| = N$ and $r = r(S, \mathcal{P})$. Every point of $\mathcal{P}$ lies within r of some $\mathbf{s} \in S$, so $\mathcal{P} \subseteq \bigcup_{\mathbf{s}\in S} B(\mathbf{s}, r)$ and, by subadditivity and the regularity hypothesis, $A \leq \sum_{\mathbf{s}\in S} \mathcal{H}^{m-1}(\mathcal{P} \cap B(\mathbf{s}, r)) \leq N C_{\mathrm{reg}} \omega_{m-1} r^{m-1}$. Solving for r and taking the infimum over S gives (19). □

Recovery of the two-objective lower bound. Take $m = 2$, so $\mathcal{H}^1$ is arclength, $A = L$, and $\omega_1 = 2$. Let $\mathcal{P}$ be a connected monotone front and let $\varrho > 0$. The portion of $\mathcal{P}$ inside $B(\mathbf{f}, \varrho)$ has endpoints at distance at most 2ϱ, so by the right inequality of (10) its arclength is at most $2\sqrt{2}\varrho = \sqrt{2}\,\omega_1\,\varrho$. The regularity hypothesis therefore holds with $C_{\mathrm{reg}} = \sqrt{2}$, and (19) reads

$$r_N^*(\mathcal{P}) \;\geq\; \frac{L}{\sqrt{2}\cdot 2\cdot N} \;=\; \frac{L}{2\sqrt{2}\,N},$$

which is the lower bound of Theorem 1; $\sqrt{2}$ is the chord–arc regularity constant of a monotone curve. For a sphere octant, which is the $m = 3$ geometry of DTLZ2, the same computation gives $C_{\mathrm{reg}} = 1$ exactly, because the area of a spherical cap of Euclidean radius ϱ on the unit sphere is $\pi\varrho^2 = \omega_2\varrho^2$.

D.7 Proof of Theorem 2

Proof. Let $\{s_k\}$ be the N quantile points of ρ on $[0, L]$, i.e. $s_k = F^{-1}(k/N)$ with $F(s) = \int_0^s \rho(u)\,\mathrm{d}u$. The Voronoi cell $I_k = [s_{k-1}, s_k]$ satisfies $\int_{I_k} \rho_S = 1/N$, so by the mean value theorem there exists $\xi_k \in I_k$ with $|I_k| = 1/\big(N\rho(\xi_k)\big)$.

Covering radius. $r(S_N^\rho) = \max_k |I_k|/2 = (1/2N)\max_k 1/\rho(\xi_k)$. As $N \to \infty$, $\{\xi_k\}$ becomes dense in $[0, L]$, so $\max_k 1/\rho(\xi_k) \to \sup_{[0,L]} 1/\rho$. Hence $r(S_N^\rho) \to (1/(2N))\sup_{[0,L]} 1/\rho$. Dividing by $L/(2N)$:

$$\frac{r(S_N^\rho)}{L/(2N)} \;\longrightarrow\; \frac{\sup 1/\rho}{L} = \sup_{[0,L]} \frac{1}{L\rho}.$$

This ratio is ≥ 1 because $\int_0^L (1/L\rho)\cdot(L\rho)\,\mathrm{d}s/L = 1$ and by Jensen's inequality ($1/x$ is convex): $\sup_{[0,L]}(1/(L\rho)) \geq \big(\int_0^L (1/(L\rho))\cdot(L\rho)\,\mathrm{d}s/L\big) = 1$.

IGD.

$$\begin{aligned}\mathrm{IGD}(S_N^\rho, \mathcal{P}) \;&\approx\; \frac{1}{L}\sum_k \frac{|I_k|^2}{4} \;=\; \frac{1}{4L}\sum_k \frac{1}{N^2\rho(\xi_k)^2} \\ &\to\; \frac{1}{4N^2L}\cdot N\int_0^L \frac{\mathrm{d}s}{\rho(s)^2}\cdot\rho(s) \;=\; \frac{1}{4NL}\int_0^L \frac{\mathrm{d}s}{\rho(s)}.\end{aligned}$$

(The approximation uses $|I_k| \approx |I_k|^2/\mathrm{diam}(I_k)$ for centred cells, and becomes exact in the $N \to \infty$ limit by standard Riemann sum arguments.) Dividing by $L/(4N)$ gives $(1/L^2)\int_0^L (\mathrm{d}s/\rho)$. This ratio is ≥ 1 by Cauchy–Schwarz: $\big(\int_0^L \rho^{1/2}\cdot\rho^{-1/2}\mathrm{d}s\big)^2 \leq \big(\int_0^L \rho\,\mathrm{d}s\big)\big(\int_0^L (1/\rho)\,\mathrm{d}s\big) = \int_0^L (1/\rho)\,\mathrm{d}s$, so $\int_0^L (1/\rho)\,\mathrm{d}s \geq L^2$. □

D.8 Proof of Lemma 2

Proof. (a) Set $\lambda_i^* = c_{\mathrm{inv}}/(f_i^* - z_i)$ with $c_{\mathrm{inv}} > 0$ chosen so that $\sum_i \lambda_i^* = 1$; this is well defined because $f_i^* > z_i$ for every i. Every coordinate of the inner maximum is then equal at $\mathbf{f}^*$,

$$\lambda_i^*\big(f_i^* - z_i\big) = c_{\mathrm{inv}} \quad \text{for all } i, \qquad \text{so} \qquad g_{\mathrm{tch}}(\mathbf{f}^* \,|\, \boldsymbol{\lambda}^*, \mathbf{z}) = c_{\mathrm{inv}}.$$

Suppose some $\mathbf{f} \in \mathcal{P}$ had $g_{\mathrm{tch}}(\mathbf{f} \,|\, \boldsymbol{\lambda}^*, \mathbf{z}) < c_{\mathrm{inv}}$. Then $\lambda_i^*(f_i - z_i) < c_{\mathrm{inv}}$ for *every* i, and substituting λ_i^* gives $f_i - z_i < f_i^* - z_i$, that is $f_i < f_i^*$ for every i. Such an $\mathbf{f}$ would strictly dominate $\mathbf{f}^*$, contradicting non-dominance. Hence $\mathbf{f}^*$ is a minimiser, and no model of $\mathcal{P}$ was used.

For the uniqueness clause, suppose no other point of $\mathcal{P}$ attains the value c_{inv}. Then $\mathbf{f}^*$ is the unique minimiser. Conversely, if $\boldsymbol{\lambda}$ is any weight for which $\mathbf{f}^*$ is the unique minimiser, the components $\lambda_i(f_i^* - z_i)$ must all be equal: if some coordinate j were strictly below the maximum, decreasing f_j slightly along the front while increasing another coordinate would leave the maximum unchanged and produce a second minimiser. Equality of all components forces $\lambda_i \propto 1/(f_i^* - z_i)$. On C_p the required tie-freeness holds because T_p is a diffeomorphism (Lemma 1), so the classical statement is recovered. On a piecewise linear front it fails at the vertices, where an entire cone of weights shares the same minimiser.

□

Remark 1 (Scalarisation scope)**.** *The inversion lemma concerns Tchebycheff scalarisation. Penalty-based boundary intersection (PBI) minimises $d_1 + \theta d_2$, where d_1 is the displacement along the reference ray and d_2 the distance to it; setting $\boldsymbol{\lambda}^* \propto \mathbf{f}^* - \mathbf{z}$ makes $d_2(\mathbf{f}^*) = 0$ but does not make $\mathbf{f}^*$ a minimiser, because for finite θ a point further along the front can trade a small d_2 against a smaller d_1. Only in the limit $\theta \to \infty$, where PBI degenerates to a pure ray-intersection rule, does the inversion hold. Results for the implementation's PBI branch are summarized in Appendix G.*

D.9 Proof of Proposition 3

Proof. The map $\psi = s \circ T$ is a strictly monotone continuous bijection onto $(0, L)$. For ordered arclength coordinates $0 < s_1 < \cdots < s_N < L$, put $a_0 = s_1$, $a_N = L - s_N$, and $a_k = s_{k+1} - s_k$ for $1 \leq k < N$. Their arclength covering radius is $R = \max\{a_0, a_N, a_1/2, \ldots, a_{N-1}/2\}$, so $L = \sum_{k=0}^N a_k \leq 2NR$. Equality requires $a_0 = a_N = R$ and every interior gap to equal $2R$, giving $s_k = (k - \frac{1}{2})L/N$. Inverting ψ gives the unique lattice. In Euclidean distance, Theorem 1 gives $r(S_N, \mathcal{P}) \leq L/(2N)$ and $\inf_{|S|=N} r(S, \mathcal{P}) \geq L/(2\sqrt{2}N)$, hence the stated ratio. Exact inversion supplies these quantile targets; setting $T = T_p$ gives the construction on C_p. □

D.10 Proof of Proposition 4

Proof. Identify candidates with the same image, so selection under d_T is greedy selection on the finite Euclidean metric space A. Let R be the final covering radius. If $R = 0$ the claim holds. Otherwise the N selected images and a farthest image form $N + 1$ points at pairwise distance at least R. Two share a nearest center in any optimal N-center cover of A; the triangle inequality gives $R \leq 2R^*$. Choosing an owning weight for each selected image preserves all distances. □

D.11 Proof of Proposition 5

Proof. Both maps $\rho \mapsto \sup 1/(L\rho)$ and $\rho \mapsto \int d\sigma/\rho$ are Schur-convex on the simplex of strictly positive densities (they are symmetric and convex). If $\hat{\rho}$ majorises $\hat{\rho}'$ then $f(\hat{\rho}') \leq f(\hat{\rho})$ for every Schur-convex f, which is the claim. A histogram closer to uniform in the majorisation order is therefore never worse on either penalty of Theorem 2. □

D.12 Proof of Theorem 3

Proof. Clause 1 ($\mathcal{E}$: gate never fires). We argue by induction on generation g. At $g = 0$ both algorithms share the same initial population and weight set W_0. If the gate has not fired by generation g, the weight set remains W_0 and all mating, replacement, and evaluation steps in both algorithms use the same weights, the same Tchebycheff values, and the same neighbourhood structure. Given the same pseudo-random seed, all random choices (crossover parents, mutation noise) are identical. Hence P_{g+1} and A_{g+1} are the same in both algorithms. By induction, the full trajectories coincide, establishing Clause 1.

Clause 2 (gate fires at g^*). At generation g^*, (9) holds on the held-out folds of Ω_{g^*}:

$$\hat{r}(W_{\mathrm{cur}}, S_{\mathrm{ho}}) \;>\; \gamma \cdot \hat{r}(\tilde{W}, S_{\mathrm{ho}}),$$

so the proxy covering radius of the new weight set is at least γ times smaller than that of the current weights on the archive score defined in Section 5.3. This establishes Clause 2. □

D.13 Proof of Proposition 2

Proof. Write $X = (x^1, \ldots, x^N)$ for the current population and Λ for the current lattice. One generation Φ_Λ draws parents from $B_\Lambda(i)$, the T nearest weights to λ^i in Δ_m, and replaces in that neighbourhood by Tchebycheff comparison. The covering radius $r(T(\Lambda), \mathcal{P})$ depends only on the images $\{T(\lambda) : \lambda \in \Lambda\}$ and on $\mathcal{P}$. If some incumbent fails to attain $T(\lambda^i)$, then $X \neq T(\Lambda)$ and $\mathrm{IGD}(X, \mathcal{P})$ is a property of X, not of $T(\Lambda)$. A remesh $\Lambda \mapsto \Lambda^*$ rebuilds every $B(i)$ and reassigns incumbents, so the next application of Φ is a different map. Nothing in the definition of Φ forces $\mathrm{IGD}(\Phi_{\Lambda^*}(X), \mathcal{P})$ to decrease when $r(T(\Lambda^*), \mathcal{P})$ decreases.

Table 5 and Table 6 supply a concrete pair. On DTLZ2, $r(T(\Lambda^*), P) = 0.135 < 0.158 = r(T(\Lambda_0), P)$ while the evolutionary archive covering rises from 0.108 to 0.239 and IGD rises from 4.87×10^{-2} to 6.90×10^{-2}. The mean neighbour overlap under the identity index is 0.23. The same sign split holds on ZDT3 and WFG2 at $m = 2$. □

E Shell-based comparator

The C_p-shell comparator GAWT estimates front geometry before inverting equal-arclength targets. On 70 two-task fronts, same-archive covering estimates indicated a 2.6 to 4.0 fold gain, compared with about 1.0 out of sample. Replacing the shell with an empirical polyline (GAWT-emp) gives $0/8/0$ IGD against the shell method.

F Implementation Details

At $m = 3$, $p = 2$ the analytic bound of Proposition 6 applies with $A = \pi/2$ and $C_{\mathrm{reg}} = 1$, giving $r^*_{105} \geq (1/210)^{1/2} = 0.0690$. That is tighter than the packing bound 0.0527, and against it the greedy set is within a factor $0.1106/0.0690 = 1.60$ of optimal, below the bound of 2. The uniform Das–Dennis lattice is worse than the greedy set by a factor of 1.6 to 2.5 across the table.

Table 7: Numerical covering ratios and clipped density statistics on C_p ($N = 200$, $m = 2$). The density column clips estimated speeds at the 0.5th and 99.5th percentiles; the last column gives continuous analytical values when stated. **Bold** marks the lowest measured ratio.

p	Arclength L	Cover ratio	Pred. penalty	Clipped distortion	Δ_p
0.5	1.623	15.82×	–	24.41	∞
1.0	1.414	**1.01**×	1	1.00	1
2.0	1.571	1.28×	$4/\pi \approx 1.27$	1.99	2 (exact)
3.0	1.686	1.34×	–	2.23	–
5.0	1.799	1.39×	–	2.48	–

Table 8: Numerical check of Theorem 4 and Proposition 6 on C_p shells at $m = 3$ and $m = 4$. N is the Das–Dennis lattice size. The packing bound $\mathrm{sep}_{N+1}/2$ is a valid lower bound on r_N^*, so $r(G_N)/\mathrm{lb}$ upper-bounds the true approximation ratio. Every row satisfies $r(G_N) \le 2\,\mathrm{lb} + 3\varepsilon$. For $p = 0.5$, the measured $r(G_N)$ matches the net resolution ε to five decimal places.

m	p	N	ε	lb (packing)	r uniform	$r(G_N)$	$r(G_N)/\mathrm{lb}$
3	0.5	105	0.0728	0.0243	0.3505	0.0728	3.00
3	1.0	105	0.0129	0.0428	0.1724	0.0885	2.07
3	2.0	105	0.0214	0.0527	0.1728	0.1106	2.10
3	3.0	105	0.0255	0.0566	0.1910	0.1155	2.04
4	0.5	120	0.1691	0.0301	0.4126	0.1691	5.61
4	1.0	120	0.0507	0.0716	0.3732	0.1480	2.07
4	2.0	120	0.0647	0.1009	0.4321	0.2140	2.12
4	3.0	120	0.0804	0.1174	0.4898	0.2401	2.05

Prespecified comparisons. The eight-problem screen uses the tests in Section 6. Its prespecified criteria are zero significant IGD losses to the host, no DTLZ2 regression, and fewer combined losses against AWA and AdaW than the shell comparator.

F.1 Proposal scoring

$\hat{T}(\lambda)$ is the archive point minimising $\max_i \lambda_i(f_i - z_i)$ in the normalised box $(f - z)/(z^{\mathrm{nad}} - z)$. The foldwise covering score uses the same map on each training fold. A proposal that worsens covering of the full archive Ω is discarded before the held-out test.

F.2 Settings and software

All decomposition variants use 20 neighbours, neighbourhood mating probability 0.9, DE/rand/1 with $F = 0.5$ and $\mathrm{CR} = 1.0$, and polynomial mutation with $\eta = 20$ and probability $1/n$ per decision variable (n variables). Software is Python 3.9, `pymoo` 0.6.1 (Blank & Deb, 2020), NumPy 1.24 and SciPy 1.11. The PyTorch preference network has two tanh hidden layers of width 64 and four outputs. Training uses 2,500 Adam updates at learning rate 3×10^{-3}, sampling 32 training weights per update.

G Additional experimental results

Online estimators on the standard suite. Across the eight problems, the correlation of true-front κ with covering improvement is not significant for any online estimator. On three-objective DTLZ2, pullback k-center and an occupancy-hole variant both trigger, with higher IGD than the host.

Shell inversion on two-teacher networks. Two pre-registered C_p-shell tests each use 70 bottleneck two-teacher fronts and an independent reference. The predictor $|\hat{p} - 1|$, averaged over the fitted exponents, has correlation $\rho_S = 0.09$ with covering gain under 40-epoch Adam and $\rho_S = -0.18$ under L-BFGS. On ZDT3, the PBI branch has significantly higher IGD than its host (Remark 1).