

PHASE-MOO: population search over competing local geometries for Pareto front approximation

Marco Paulsen Rodrigues

`marco.paulsen@ufpel.edu.br`

Universidade Federal de Pelotas

Rafaela Moreira Javier Lemos

Universidade Federal de Pelotas

Research Article

Keywords: Multiobjective optimization, Gradient-based heuristic, Memetic algorithm, Variable metric, Algorithm portfolio, Hypervolume

Posted Date: October 19th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-11253023/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

PHASE-MOO: population search over competing local geometries for Pareto front approximation

Marco Paulsen Rodrigues^{1*} and Rafaela Moreira Javier Lemos²

^{1*}Centro de Engenharias, Universidade Federal de Pelotas, Rua Benjamin Constant, 989, Porto, Pelotas, 96010-020, RS, Brazil.

²Centro de Desenvolvimento Tecnológico, Universidade Federal de Pelotas, Rua Gomes Carneiro, 1, Porto, Pelotas, 96010-610, RS, Brazil.

*Corresponding author(s). E-mail(s): marco.paulsen@ufpel.edu.br;
Contributing authors: rafaela.mjl@gmail.com;

Abstract

Gradient-based Pareto front approximation needs reliable local descent and a way to spread search effort over distinct trade-offs. We propose Pareto Heuristic with Alternative Search geometries for Multiobjective Optimization (PHASE-MOO), a population heuristic in which the local search geometry is itself a search variable. Each branch carries a positive definite metric and moves by a metric-consistent multiple-gradient descent step with memory and safeguards. Branches with different metrics compete for evaluations through an explore–select–commit cycle, and the method returns every nondominated point it evaluates. A convergence theorem covers the safeguarded local kernel. The population mechanisms are assessed empirically on six benchmarks with two or three objectives and up to 200 variables. A racing-based calibration identifies the initial population and the direction radius as the parameters that matter. A paired ablation with 20 seeds shows that the multi-geometry population improves hypervolume on all problems over a multistart identity-metric control and over an identity-only population. Adaptive and uniform allocation over a de-duplicated geometry portfolio are practically equivalent, so the portfolio helps through geometric diversity rather than selection. At 20,000 joint objective–Jacobian evaluations (100,000 for the 200-variable problem), median reference hypervolume ratios are 98.30–99.99% on five problems and 94.20% on Kursawe, whose disconnected front is incompletely covered in 4 of 20 runs. Runs with one million evaluations reach 100.00–100.04% on all problems. Against eight established methods under the same allowances, PHASE-MOO is significantly better in 33 of 46 hypervolume comparisons and worse in 6, with the clearest advantage on the 200-variable problem.

Keywords: Multiobjective optimization, Gradient-based heuristic, Memetic algorithm, Variable metric, Algorithm portfolio, Hypervolume

1 Introduction

Smooth multiobjective optimization differs from scalar optimization in two related ways: a descent direction must reconcile several objective gradients, and a useful numerical output is a set of trade-off solutions rather than a single point [1, 2]. Convergence of one trajectory to a first-order critical point therefore says little about coverage of a Pareto front. Common-descent methods provide the local foundation for this task [3–5], with extensions based on projected gradients, Newton models, nonlinear conjugacy, and constrained SQP [6–9]. Front-oriented descent and trust-region algorithms add explicit mechanisms for constructing a distributed approximation [10–12], complementing classical scalarization procedures such as normal-boundary intersection [13].

A descent method also depends on the geometry in which the direction is computed. Variable-metric methods replace the Euclidean geometry with positive definite quadratic models [14], memory-gradient methods combine current and previous directional information [15], and objective scaling affects both the direction-finding problem and the step accepted by a simultaneous line search [16]. In these methods the geometry is a design decision fixed before the run or updated along a single trajectory. Different parts of a front, or different stages of a run, may however be served best by different geometries, and it is rarely known in advance which one to choose.

Population heuristics offer a natural way to hedge this choice. Dominance-based and decomposition-based evolutionary algorithms distribute effort over the front without derivatives [17–19]. Gradient information has been embedded in such algorithms as a local search operator [20–22], memetic front algorithms combine population search with common-descent refinement [23], and predictor–corrector continuation traces connected parts of Pareto sets from critical points [24]. Hypervolume-based methods adapt search distributions or follow hypervolume gradients [25, 26], and in MO-CMA-ES the search geometry is adapted through the covariance matrices of mutation distributions [27]. To our knowledge, none of these approaches treats the metric of a gradient step as a population-level variable on which branches compete.

PHASE-MOO takes this step. Its core is a metric multiple-gradient descent: the same positive definite metric determines the simplex weights and the descent direction. Each initial seed starts one branch, and the metrics, drawn from a family built from trajectory history, Jacobian information, and coordinate transformations, are assigned cyclically across seeds; branching later creates children of a seed with other metrics. Local updates combine the metric direction with history, attraction to the branch’s best merit, momentum, and repulsion from discovered nondominated points, under safeguards that restore common descent when necessary. Periodic exploration creates new position–geometry candidates, the best candidates receive further steps (explore–select–commit), and stage-wise pruning reduces the seed population. The method returns the nondominated set of all points it has evaluated, so pruning redirects effort without discarding coverage already obtained.

Such a design raises a question that competitive benchmarking alone cannot answer: does the portfolio of geometries improve the approximation, and if so, through diversity or through adaptive selection among geometries? The numerical study addresses it directly. A paired ablation shows that the multi-geometry population is much better than multistart identity-metric descent and than the same population restricted to the identity metric. Adaptive allocation, however, is practically equivalent to uniform allocation once mathematically redundant geometries are removed. The benefit of the portfolio therefore comes from complementary geometries rather than from learning which geometry is best. The same study identifies the main limitation: at small budgets, coverage of disconnected parts of a front depends on whether some branch reaches their basin.

The contributions are threefold. First, we formulate a metric-consistent MGDA kernel whose safeguards preserve simultaneous descent after memory and momentum are added, and prove its convergence under interior conditions for uniformly bounded metric sequences. Second, we develop a population heuristic that treats local geometry as a branch-level search variable, with an extensible metric family and an explore–select–commit mechanism. Third, we evaluate the method on six benchmarks with two or three objectives and up to 200 variables, through four experiments: a racing-based calibration with response-surface analysis, a paired ablation with equivalence tests, high-budget case studies, and a comparison with eight evolutionary, memetic, front-descent, and hypervolume-based methods. In the comparison, PHASE-MOO is significantly better in most comparisons, especially on the 200-variable problem, and trails on the disconnected Kursawe front.

PHASE-MOO is a heuristic. The convergence result of Section 4 applies to its safeguarded local kernel, whereas branching, geometry allocation, pruning, and archiving are justified empirically. Following established recommendations for the experimental analysis of heuristics [28, 29], the study isolates these components with paired controls, fixes the evaluation allowance and its accounting for every method, tests equivalence claims against margins set before the runs, and reports failures and limitations alongside medians.

The benchmarks cover disconnected, nonconvex, three-objective, rugged, and high-dimensional settings, reflecting complementary benchmark requirements [30]. Kursawe has a disconnected front and Viennet three nonconvex objectives; Fonseca–Fleming is a smooth two-objective problem with a known Pareto set [31]; DTLZ2 has a scalable spherical front [32]; and two constructed problems add an oscillating landscape and 200 decision variables. Approximations are scored by exact hypervolume and inverted generational distance against fixed reference sets with a common normalization.

The paper is organized as follows. Section 2 introduces the metric MGDA kernel and the geometry family, and Section 3 the population mechanism. Section 4 proves convergence of the local kernel. Section 5 describes the experimental design, Section 6 presents the results, and Section 7 concludes.

2 Problem and local geometry

2.1 Pareto optimality and criticality

Consider

$$\min_{x \in \Omega} F(x) = (f_1(x), \dots, f_m(x)), \quad \Omega = [\ell, u] \subset \mathbb{R}^n. \quad (1)$$

For minimization, y dominates z if $y_i \leq z_i$ for every i and at least one inequality is strict. A feasible point is Pareto optimal when no feasible point dominates its objective vector. In the unconstrained case, or at an interior point, first-order Pareto criticality is

$$0 \in \text{conv}\{\nabla f_1(x), \dots, \nabla f_m(x)\}. \quad (2)$$

At a box boundary, the corresponding necessary condition is instead [6]

$$0 \in J(x)^T \Delta_m + N_\Omega(x), \quad \Delta_m = \{\alpha \geq 0 : \mathbf{1}^T \alpha = 1\}, \quad (3)$$

where N_Ω is the convex normal cone and the Jacobian $J \in \mathbb{R}^{m \times n}$ has gradient rows. This row convention is used throughout; it agrees with the numerical code. For numerical stopping and archive admission we use the projected residual

$$\chi_\Omega(x) = \|x - \Pi_\Omega(x - J(x)^T \alpha_I)\|_2, \quad (4)$$

where α_I solves (5) with $M = I$. A zero residual supplies a feasible simplex multiplier satisfying (3). Thus it is a sufficient box-stationarity certificate, evaluated identically across treatments. At a boundary, a different simplex multiplier may certify stationarity even when this residual is positive; no converse is asserted.

2.2 Metric multiple-gradient descent

For a symmetric positive definite matrix $M \in \mathbb{R}^{n \times n}$, solve

$$\alpha_M \in \arg \min_{\alpha \in \Delta_m} \frac{1}{2} \alpha^T J M J^T \alpha, \quad s_M = J^T \alpha_M, \quad d_M = -M s_M. \quad (5)$$

The weights are optimized using the same metric that generates the direction. Multiplying a Euclidean MGDA direction by an arbitrary positive definite matrix after the simplex problem has been solved does not in general preserve common descent.

Equation (5) also has a direct geometric interpretation. Let

$$G = M^{-1}, \quad \langle u, v \rangle_G = u^T G v, \quad \|s\|_{G^{-1}}^2 = s^T G^{-1} s = s^T M s. \quad (6)$$

Then the simplex subproblem minimizes the squared dual norm of the convex gradient combination,

$$\alpha_M \in \arg \min_{\alpha \in \Delta_m} \frac{1}{2} \|J^T \alpha\|_{G^{-1}}^2, \quad d_M = -G^{-1} J^T \alpha_M. \quad (7)$$

Hence M plays the role of an inverse metric tensor, or cometric, and d_M is the negative G -gradient of the selected convex combination of objectives. Changing M therefore changes both which convex gradient combination is shortest and how that combination is mapped into a search direction. This metric consistency is the local mechanism on which the geometry competition in PHASE-MOO is built.

Proposition 1 *If (5) is solved exactly, then*

$$\nabla f_i(x)^T d_M \leq -s_M^T M s_M = -\|d_M\|_{M^{-1}}^2 \quad (i = 1, \dots, m). \quad (8)$$

Moreover, $d_M = 0$ if and only if (2) holds.

Proof Optimality on the simplex implies $(e_i - \alpha_M)^T J M J^T \alpha_M \geq 0$ for each simplex vertex e_i . Thus $\nabla f_i^T M s_M \geq s_M^T M s_M$, which gives (8). Positive definiteness implies that the minimum quadratic value is zero exactly when a convex combination of the gradients is zero. $\square$

For the two- and three-objective experiments considered here, the implementation solves the two-objective subproblem analytically and enumerates vertices, edges, and an interior candidate for three objectives. Small numerical regularizers stabilize the linear solves.

2.3 Three-level geometry competition

The PHASE-MOO architecture separates three roles that are often merged into a single update rule. At iteration k , branch j carries a geometry-conditioned state

$$\mathcal{B}_{j,k} = (x_{j,k}, M_{j,k}, \mathcal{H}_{j,k}), \quad G_{j,k} = M_{j,k}^{-1}, \quad (9)$$

where $\mathcal{H}_{j,k}$ collects the branch memory and metric state.

1. *Local geometry.* $G_{j,k}$ defines the inner product through which the branch interprets local directions and gradient combinations.
2. *Geometry-conditioned descent.* Given $G_{j,k}$, Eq. (5) produces a metric-consistent MGDA direction, which is then augmented by safeguarded history, attraction, momentum, and repulsion.
3. *Population-level geometry allocation.* Branching forms joint position–geometry candidates and the explore–select–commit mechanism reallocates evaluations among them. Metric counters and stage-wise suppression operate at this level.

Thus the metric family is not a collection of post-hoc preconditioners applied to one common trajectory. It defines concurrent local search processes, each with its own geometry-conditioned dynamics, and the population mechanism decides which of these processes receive computational effort.

Every symmetric positive definite $M_{j,k}$ defines the local inner product in (6). When a metric mode is a smooth function of position alone, $G_j(x) = M_j(x)^{-1}$ is a Riemannian metric on the decision domain in the standard sense [33]. Some implemented modes also depend on accumulated branch history, so the broader term *state-dependent local geometry* is used throughout. This distinction leaves the metric-MGDA equations unchanged.

The same formulation naturally accommodates geometries induced by smooth coordinate transformations. If $z = \Phi(x)$ is locally invertible with Jacobian $A(x) = D\Phi(x)$ and Euclidean geometry is used in z , the pullback metric and corresponding inverse metric are

$$G_\Phi(x) = A(x)^T A(x), \quad M_\Phi(x) = A(x)^{-1} A(x)^{-T}. \quad (10)$$

Such transformed-coordinate geometries enter Eq. (5) as additional competitors without changing the descent kernel. Section 2.5 specifies the implemented extension.

2.4 Implemented metric family

The base family contains seven positive definite inverse-metric modes spanning isotropic scaling, diagonal history, Jacobian-informed anisotropy, secant information, and hybridization. Let $a_{k+1} = \beta_a a_k + (1 - \beta_a)(d_k \odot d_k)$, $D_k = \text{diag}((\sqrt{a_k} + \delta)^{-1})$, and $Q_k = J_k^T J_k$. With $Q_{k,r} = U_r \text{diag}(\sigma_1^2, \dots, \sigma_r^2) U_r^T$ formed from the leading right singular vectors of J_k , the metrics are

$$M_I = I, \quad (11)$$

$$M_D = D_k, \quad (12)$$

$$M_G = (1 - \theta)I + \theta(Q_k + \lambda I)^{-1}, \quad (13)$$

$$M_{G,r} = (1 - \theta)I + \theta(Q_{k,r} + \lambda I)^{-1}, \quad (14)$$

$$M_S = \text{diag}(\max(\zeta_k, \lambda)^{-1}), \quad (15)$$

$$M_T = \frac{\min(1, R)}{\max(1, \sqrt{n^{-1}} \mathbf{1}^T a_k)} I, \quad (16)$$

$$M_H = \frac{1}{2} D_k + \frac{1}{2} M_G. \quad (17)$$

All componentwise inverses use positive numerical floors. Here R is the prescribed direction radius, with the numerator in M_T set to one when no radius is prescribed. On an accepted step, let $y_k = s_{M,k+1} - s_{M,k}$, computing the new metric direction with the accumulator updated at the current attempt. The diagonal secant state ζ_k , distinct from the historical displacement h_k , is $\zeta_{k+1} = \beta_h \zeta_k + (1 - \beta_h) \text{clip}(|y_k|/(|x_{k+1} - x_k| + 10^{-8}), 10^{-6}, 10^6)$. The code labels are, respectively, `identity`, `adagrad_diag`, `grad_gram`, `low_rank_gram`, `damped_hessian_diag`, `trust_scaled_identity`, and `hybrid_diag_gram`. The family deliberately combines information from distinct sources: M_I supplies the Euclidean baseline, M_D uses an exponentially weighted direction accumulator [34], M_G and $M_{G,r}$ use the current objective Jacobian, M_S is a positive diagonal secant model, M_T introduces isotropic radius- and history-dependent scaling, and M_H blends trajectory and Jacobian information. The accumulator uses the final safeguarded direction before radius scaling, including unsuccessful line-search attempts.

For the low-rank metric, the leading right singular vectors of J replace the full spectrum of $J^T J$ before regularization and inversion. The full Gram inverse is evaluated with the Woodbury identity in objective space, and the low-rank inverse uses the thin singular value decomposition. These forms reduce the dense solve from dimension n to at most m and make the same metric definitions practical on HD200. In low-dimensional cases exact coincidences between full- and low-rank constructions can occur; for example, in two-dimensional Viennet runs with rank two they coincide mathematically. Such guaranteed coincidences are removed from the portfolio before a run (Section 2.5).

2.5 Coordinate-induced geometries and experimental portfolios

The base family is augmented with coordinate-induced SPD models in the low- and moderate-dimensional experiments. This extension adds directional information while preserving the metric-consistent descent kernel. Let $x = c + Sy$ normalize the box, where $c = (\ell + u)/2$ and $S = \text{diag}((u - \ell)/2)$. For a locally invertible map $y = \psi(z)$, define $B = SD\psi(z)$. The unregularized inverse metric is BB^T , equivalently Eq. (10). Its implemented, scale-normalized form is

$$M_\psi = U \text{diag}(\text{clip}(\lambda_i, \lambda_-, \lambda_+)) U^T, \quad U \text{diag}(\lambda_i) U^T = \frac{nBB^T}{\text{tr}(BB^T)}. \quad (18)$$

Normalization separates directional anisotropy from an arbitrary scalar magnification of the coordinate map. Spectral bounds $0 < \lambda_- \leq \lambda_+$ are setup parameters. Away from singularities, the unregularized construction has the exact coordinate interpretation; clipping supplies a bounded SPD model for the local kernel.

The coordinate family comprises the following constructions.

- *Polar and cylindrical coordinates*: $y_1 = \rho \cos \theta$, $y_2 = \rho \sin \theta$, retaining the remaining Cartesian components in the cylindrical extension. Polar coordinates are used in two dimensions; cylindrical coordinates are available in higher dimensions. In two dimensions these two labels describe the same local geometry.
- *Spherical coordinates*: in three dimensions, $y = r(\sin \theta \cos \varphi, \sin \theta \sin \varphi, \cos \theta)^T$.
- *Radial cubic coordinates*: $z = (1 + a\|y\|^2)y$, $a \geq 0$, with inverse Jacobian $[(1 + a\|y\|^2)I + 2ayy^T]^{-1}$.
- *Nonlinear shear*: $y_1 = z_1$ and $y_j = z_j + \kappa z_{j-1}^2$ for $j \geq 2$. The coefficient κ controls the departure from Cartesian coordinates.
- *Box-logit coordinates*: $x_i = \ell_i + (u_i - \ell_i)\sigma(z_i)$, where $\sigma(t) = (1 + e^{-t})^{-1}$; here $B = \text{diag}((u_i - \ell_i)\sigma(z_i)[1 - \sigma(z_i)])$ directly in physical coordinates.

The radial coefficient, shear coefficient, coordinate floors, and spectral bounds remain configurable. Regularized Jacobians are used near angular singularities and box boundaries; these yield local SPD competitors rather than globally invertible coordinate systems.

The configured portfolio is fixed by dimension. Kursawe and Fonseca–Fleming list the seven base modes plus the radial cubic, box-logit, cylindrical, spherical, and nonlinear-shear modes (12 labels); Viennet and Rugged2D replace the spherical label by polar (12 labels); DTLZ2 omits both polar and spherical modes (11 labels); and HD200 uses the seven base modes only. Two coincidences are guaranteed by construction: the low-rank Gram metric equals the full Gram metric when its rank cutoff is at least $\min(m, n)$, and the polar and cylindrical pullbacks coincide when $n = 2$. Before a run, PHASE-MOO removes such mathematically redundant modes; no rank is inferred from sampled Jacobians. The de-duplicated portfolio used by the method therefore has $k = 11$ modes on Kursawe and Fonseca–Fleming (low-rank Gram removed), $k = 11$ on DTLZ2 (nothing removed, since the cutoff two is below $\min(m, n) =$

3), $k = 10$ on Viennet and Rugged2D (low-rank Gram and polar removed), and $k = 6$ on HD200 (low-rank Gram removed). The calibration used the configured portfolios, which the ablation retains as variant P5 (Section 5.1). The HD200 evidence concerns the base family, whereas the other five problems use the coordinate-enriched portfolio.

For the coordinate models, the reported setups use radial coefficient $a = 0.3$, shear coefficient $\kappa = 0.2$, coordinate floor 10^{-6} , and spectral bounds $[10^{-6}, 10^6]$. These values are configurable, not defining constants of PHASE-MOO. The population-covariance metric available in the implementation is not part of these experimental portfolios; population-informed proposal positions are a separate branching mechanism.

All reported modes use Cartesian safeguarded directions and line search: coordinate information enters through M , not through nonlinear chart stepping or a different acceptance rule. The local theorem therefore applies under its stated assumptions and uniform SPD bounds.

3 Population framework and memory

3.1 Geometry-conditioned branch state and scalar merit

A PHASE-MOO branch is a geometry-conditioned search state of the form (9). In addition to its current and previous decision vectors and metric mode, each branch stores metric state, momentum, and a best-so-far point. A normalized weighted Chebyshev-type branch merit,

$$\phi_b(x) = \max_i \frac{f_i(x) - z_i^{\text{id}}}{\varrho_i(w_{b,i} + \varepsilon)}, \quad \varrho_i = \max\{z_i^{\text{nad}} - z_i^{\text{id}}, 10^{-12}\}, \quad (19)$$

is used for branch ranking and memory management. Here z^{id} is the componentwise minimum of all objective vectors evaluated so far inside the box, and z^{nad} the componentwise maximum over the current nondominated set of these vectors, which is the approximation returned by the method (Section 3.4). Both are updated after every new evaluation, so the merit is invariant to translations and positive rescalings of the individual objectives; before the first evaluation the reference is zero and the ranges are one. Merit values recorded at different times may therefore refer to slightly different normalizations; this affects only heuristic ranking and confidence terms, never the acceptance of a step. In the reported PHASE configurations, each seed receives a positive preference vector sampled from a symmetric Dirichlet distribution with concentration 0.5; descendants inherit this vector. The alternative, used when preference sampling is disabled, takes the absolute initial objective values plus a small offset. Fixed seed preferences provide a consistent merit for comparing descendants while distributing search priorities across trade-offs. The scalar merit allocates population effort, while simultaneous objective tests remain responsible for accepting local steps.

3.2 Historical displacement, attraction, and repulsion

Let $h_k = x_k - x_{k-1}$ and $b_k = x_k^{\text{best}} - x_k$. The candidate is

$$d_k^{\text{raw}} = (1 - \lambda_{h,k} - \lambda_{b,k})d_{M,k} + \lambda_{h,k}h_k + \lambda_{b,k}b_k + \mu r_k. \quad (20)$$

The two memory coefficients are rescaled when their sum exceeds 0.95. Historical displacement is weighted by the confidence in Section 3.5. Attraction is weighted by the maximum of the confidence progress factor and $1 - \exp(-s_k/g)$, where s_k counts successive improvements of the best merit value and $g > 0$ sets the growth scale.

The repulsion vector uses neighbors chosen in objective space from the currently discovered non-dominated pool. In decision space, the implementation forms an inverse-squared-distance weighted displacement, averages it, and normalizes it to unit length when nonzero. The coefficient μ then controls its contribution. The operator is entirely endogenous: it uses the front discovered by PHASE-MOO and requires no externally supplied reference set.

3.3 Safeguards and line search

The raw direction is replaced by $d_{M,k}$ if any component violates

$$J_k d_k^{\text{raw}} \leq \gamma_g \min(J_k d_{M,k}, 0), \quad (21)$$

up to the numerical tolerance in the code. Momentum is updated from this safeguarded direction:

$$p_k = \beta_p p_{k-1} + (1 - \beta_p) d_k^{\text{safe}}, \quad d_k = (1 - \gamma_k) d_k^{\text{safe}} + \gamma_k p_k. \quad (22)$$

The momentum mixture is enabled when a merit comparison and an alignment check pass. After mixing, the final safeguard repeats (21) and requires $\|d_k\| \leq C_d \|d_{M,k}\|$; otherwise it restores the pure metric direction. A positive scaling can additionally cap the direction norm when a radius is specified. Under the compactness and uniform metric bounds in Theorem 1, a fixed positive radius preserves a uniform descent margin, as established below.

The line search, following the sufficient-decrease principle of Armijo [35] and its simultaneous use in vector descent methods [4], starts from a branch-specific initial trial t_k^0 and multiplies t by $\rho \in (0, 1)$ after each rejection. The initial trial is adaptive. A new branch starts with $t^0 = 1$; after an accepted step of length t the next initial trial is $\min\{s_{\max}, 2t\}$, and after a failed search it is $\max\{t_{\text{lo}}, \rho t_k^0\}$, with $t_{\text{lo}} = 10^{-6}$. Hence $t_k^0 \in [t_{\text{lo}}, s_{\max}]$, where the step ceiling $s_{\max}$ is a setup parameter. Each trial is projected onto the box, and sufficient decrease is measured on the move actually taken:

$$x^+ = \Pi_{\Omega}(x_k + t d_k), \quad \Delta_k = x^+ - x_k, \quad \nabla f_i(x_k)^T \Delta_k < 0, \quad f_i(x^+) \leq f_i(x_k) + \sigma \nabla f_i(x_k)^T \Delta_k \quad \forall i. \quad (23)$$

A trial whose projected move is not a predicted descent direction for every objective is rejected without evaluation. When the projection is inactive, $\Delta_k = t d_k$ and (23) reduces to the classical simultaneous Armijo condition. When a direction radius R is prescribed, trials with $\|\Delta_k\| > R$ are also rejected, so accepted moves never exceed R . The implementation uses at most 20 reductions and small additive tolerances; an unsuccessful search leaves the point unchanged.

3.4 Branch generation, geometry selection, archive, and output

At scheduled branching events, PHASE-MOO generates candidate pairs consisting of a decision-space position and an active metric mode. Positions are obtained by interpolation, extrapolation, covariance sampling, and recombination using discovered nondominated neighbors. Metrics are sampled from the active family, each position–geometry candidate receives exploratory local steps, and the best-ranked candidates receive additional commit steps. This explore–select–commit cycle is the population-level geometry competition in operational form: it concentrates evaluations on combinations of location and local geometry that are successful under the branch merit. Elitist nondominated sorting, strength-based ranking, and neighborhood decomposition provide established population-management context [17–19, 36].

Candidate modes are proposed by one of two adaptive rules, selected per problem in the calibration. For the active mode set $\mathcal{A}$, the count rule uses the weights

$$w_j = \max\{\varepsilon_{\text{sel}}, (1 + B_j)^{-p}\}, \quad \pi_j = \frac{w_j}{\sum_{q \in \mathcal{A}} w_q}, \quad j \in \mathcal{A}. \quad (24)$$

Here $p > 0$ and $\varepsilon_{\text{sel}} > 0$ are setup parameters; the floor applies to the unnormalized weight, not directly to the probability. Candidate modes are sampled without replacement. The global counter B_j accumulates unfavorable candidate events across seeds: nonselection adds one, and a merit gap above the equivalence tolerance adds another when the best candidate merit is positive. A single branching sweep can therefore produce multiple increments. Stage-wise suppression additionally compares the selected-to-proposed count ratio with the best ratio, subject to the bad-count threshold and activation fraction; at least one mode is retained. Surviving branches with suppressed modes are removed at this stage, except that a seed’s best branch is retained and its mode reactivated if none of its branches would survive. In the geometric view of Section 2.3, these mechanisms form an online allocation policy over the active geometry family: p controls how rapidly evidence from unfavorable candidate events changes proposal pressure, whereas suppression changes family membership at stage boundaries. From a heuristic-design perspective, Eq. (24) is an adaptive operator selection rule [37, 38] whose operators are metric modes; credit is assigned through counts of unfavorable events, and suppression acts as operator elimination.

The reward rule instead assigns step-level credit. After each local update of a branch with mode j , the quality of the mode is updated as

$$q_j \leftarrow (1 - \beta_q) q_j + \beta_q \frac{g}{u}, \quad g = \max\left\{0, \frac{\phi_b(x_k) - \phi_b(x_{k+1})}{|\phi_b(x_k)| + 10^{-12}}\right\}, \quad (25)$$

where g is the relative merit decrease of an accepted step ($g = 0$ for a rejected one), $u \geq 1$ is the number of evaluations charged by the update, and $\beta_q = 0.05$. Proposal probabilities follow probability matching, $\pi_j = p_{\min} + (1 - Kp_{\min})q_j / \sum_{q \in \mathcal{A}} q_q$ with $K = |\mathcal{A}|$ and $p_{\min} = \min\{0.02, 1/K\}$, and are uniform while all qualities are zero. Uniform proposals, $\pi_j = 1/K$, are the limit $p \rightarrow 0$ of (24) and serve as a control in the ablation.

Pruning operates within seeds and across seeds, using merit and decision-space zoning. A search archive admits eligible branches after its configured activation fraction, filters dominance, and controls occupancy through objective bins; together with the active branches it supplies the nondominated pool used by repulsion and branching proposals. Independently of this search archive, every new evaluation at a point inside the box is offered to an unbounded set $\mathcal{E}$ of nondominated evaluated points, which also provides z^{id} and z^{nad} in (19). The approximation returned by PHASE-MOO is $\mathcal{E}$. It contains the final endpoints together with all nondominated line-search trials, exploratory candidates, and intermediate iterates, so pruning a branch or a seed never removes coverage that has already been found.

Algorithm 1: PHASE-MOO

1. **Input:** box Ω , initial population X_0 , configured metric portfolio $\mathcal{M}$, allocation rule, branch and pruning schedules, archive rules, step ceiling $s_{\max}$, direction radius R , and joint F - J evaluation budget B .
2. Remove the mathematically redundant modes from $\mathcal{M}$. Initialize one branch per seed, assigning metric modes cyclically across seeds; initialize metric state, merit memory, momentum, initial trial $t^0 = 1$, allocation counters or qualities, an empty search archive, and an empty evaluated set $\mathcal{E}$.
3. For each active branch, construct M , solve the geometry-conditioned MGDA problem (5), form (20), apply safeguards, momentum, and the radius cap, and run the projected search (23) from t_k^0 ; then update t_k^0 .
4. Insert every new in-box evaluation into $\mathcal{E}$ and update z^{id} and z^{nad} . At each attempted update, record its starting merit and point; update the current merit after acceptance, the branch-specific momentum and metric state, and, under the reward rule, the quality (25). After ordinary branch updates, insert eligible critical points into the search archive.
5. At within-seed pruning events, retain the best-ranked branch of each seed. At branching events, form joint position-geometry candidates, execute the prescribed explore steps, rank candidates, retain the configured survivors, and execute their commit steps.
6. At stage events, rank seeds, apply the configured zonal or global pruning rule, and update the active metric family. Candidate proposal probabilities use the latest counters or qualities at each branching event.
7. Stop at the evaluation ceiling or a common secondary stopping limit; return $\mathcal{E}$, the nondominated set of all evaluated in-box points.

3.5 Progress confidence for historical displacement

PHASE-MOO modulates historical displacement with a confidence score combining positive local alignment c_k^{align} , consecutive relative merit improvement m_k , tangent agreement t_k , and a trajectory-curvature factor u_k :

$$q_k = c_k^{\text{align}} \min(1, [m_k]_+ / \tau_m) t_k u_k, \quad \lambda_{h,k} = \lambda_h q_k. \quad (26)$$

The tangent proxy is a normalized secant between discovered decision vectors selected by objective-space proximity; when too few neighbors are available, its agreement factor is set to one. The consecutive-progress term is

$$m_k = \frac{\phi_b(x_{k-1}) - \phi_b(x_k)}{|\phi_b(x_{k-1})| + \varepsilon}. \quad (27)$$

The implementation maintains the previous and current merits as distinct state variables, evaluated with the same current branch weights. At each attempted update it records the starting point and merit; an accepted step sets the current merit to the new value, whereas a rejected step records equal consecutive merits. A relocated child starts without a preceding local displacement. Thus sustained progress can activate historical displacement, while rejected or stationary steps yield zero progress at the next update. The same rule is used by ordinary, exploratory, and commit steps. Confidence is allowed to vanish when there is no positive aligned progress; this disables the historical term while retaining the safeguarded metric step.

4 Convergence of the safeguarded local kernel

The following result concerns uninterrupted local sequences with inactive projection and unlimited Armijo backtracking. Population operations act between such local sequences and are evaluated computationally in Section 5.

Theorem 1 *Suppose an uninterrupted sequence lies in a compact set, the objective gradients are Lipschitz continuous on a neighborhood of all trial segments, and one objective is bounded below there. Assume $\underline{\lambda}I \preceq M_k \preceq \bar{\lambda}I$, with $0 < \underline{\lambda} \leq \bar{\lambda} < \infty$, and exact solutions of (5). Set $v_k = \|s_{M,k}\|_{M_k}$. Suppose the final directions satisfy*

$$\nabla f_i(x_k)^T d_k \leq -\kappa v_k^2 \quad \forall i, \quad \|d_k\| \leq C v_k, \quad (28)$$

for fixed $\kappa, C > 0$. Assume all trial points for $t \in (0, t_{\text{hi}}]$ are feasible, so projection is inactive. Let the steps use the first acceptable Armijo length from $\{t_k^0 \rho^j : j \geq 0\}$, without a backtracking cap, where the initial trials satisfy $t_{\text{lo}} \leq t_k^0 \leq t_{\text{hi}}$ for fixed $0 < t_{\text{lo}} \leq t_{\text{hi}}$. Then $v_k \rightarrow 0$, and every accumulation point satisfies (2).

Proof Let L bound all gradient Lipschitz constants. The descent lemma and (28) give

$$f_i(x_k + t d_k) \leq f_i(x_k) + t \nabla f_i(x_k)^T d_k + \frac{L}{2} t^2 C^2 v_k^2. \quad (29)$$

Thus Armijo holds for $t \leq t_* := 2(1 - \sigma)\kappa/(LC^2)$. The first acceptable backtracking step is bounded below by $t_{\min} = \min\{t_{\text{lo}}, \rho t_*\} > 0$. For any objective bounded below, summing the accepted decreases yields

$$\sigma \kappa t_{\min} \sum_{k=0}^{\infty} v_k^2 \leq f_i(x_0) - \inf_k f_i(x_k) < \infty. \quad (30)$$

Hence $v_k \rightarrow 0$. Define the Euclidean residual $\chi(x) = \min_{\alpha \in \Delta_m} \|J(x)^T \alpha\|$. Since $v_k^2 \geq \underline{\lambda} \chi(x_k)^2$, we obtain $\chi(x_k) \rightarrow 0$. The minimum of this continuous function over the compact simplex varies continuously with x . Therefore $\chi(x_*) = 0$ at every accumulation point, proving (2). $\square$

For the pure metric direction, Proposition 1 gives $\kappa = 1$, and $\|d_M\| \leq \sqrt{\bar{\lambda}} v$ supplies the norm bound. An augmented direction can inherit the theorem if the *final* direction is checked against (28), with a fallback to the pure metric direction. To account for the implemented radius cap, write the pre-scaled direction as $\tilde{d}_k$ and $d_k = \theta_k \tilde{d}_k$, with $\theta_k = \min\{1, R/\|\tilde{d}_k\|\}$ and fixed $R > 0$. Compactness, bounded gradients, the uniform metric bound, and the relative direction safeguard yield $\|\tilde{d}_k\| \leq D$ for some $D > 0$. Consequently $\theta_k \geq \min\{1, R/D\} > 0$: scaling multiplies the pre-cap descent constant by at least this lower bound and cannot increase the norm bound. The same argument covers varying radii bounded away from zero. The adaptive initial trial of Section 3.3 satisfies the hypothesis with $t_{\text{lo}} = 10^{-6}$ and $t_{\text{hi}} = s_{\max}$; with inactive projection the projected test (23) is the classical one, and the rejection of moves longer than R only adds the lower bound $\rho R/D$ to $t_{\min}$. Numerical tolerances and regularized subproblem solves are implementation choices; the theorem assumes the exact inequalities stated above.

Importantly for the geometric architecture, Theorem 1 does not require a fixed matrix M : it permits any uniformly bounded positive definite sequence M_k satisfying its hypotheses. The local guarantee is therefore compatible with variable geometry. Population-level branch replacement and finite projected line searches remain separate finite-budget mechanisms; for box constraints the implementation records (4), and the experiments assess the complete algorithm.

5 Numerical study

5.1 Treatments and evaluation budget

All PHASE-MOO experiments in Sections 6.1–6.3 use the same implementation and setups (Section 5.2) and count distinct joint objective–Jacobian evaluations. The evaluation ceiling is 20,000 for five problems and 100,000 for HD200. Initialization and rejected trials count toward the allowance; identical cached requests do not. Actual counts and stopping reasons are recorded.

The calibration (Section 6.1) comprises 2,172 runs, 362 per problem: 64 Latin-hypercube configurations and a control raced over seeds 1101–1110. The component ablation (Section 6.2) comprises 480 runs of five variants:

- P1 (MGDA-I): multistart identity-metric MGDA from the same initial points, using the box-constrained Armijo kernel without branching, pruning, memory, momentum, repulsion, or preference sampling; it uses a fixed initial trial step (0.08; 0.4 on HD200), the simultaneous Armijo test on the clipped trial, and returns the nondominated union of its endpoints and filtered archive;

- P2 (PHASE-I): the complete population mechanism restricted to the identity geometry;
- P3: the de-duplicated geometry portfolio with uniform allocation and no suppression;
- P4 (PHASE-MOO): the method, that is, the calibrated setup with the de-duplicated portfolio and the calibrated allocation rule;
- P5: P4 without de-duplication, i.e., exactly the calibrated configuration; it links the ablation to the calibration and is reported in the text only.

P3, P4, and P5 use 20 paired seeds (6101–6120); P1 and P2 use the first ten. These seeds were not used for any calibration decision. Initial decision points are shared within each problem and seed, and every variant inherits the setup of Table 1 instead of being tuned separately. Before the ablation, four calibration runs were repeated with the ablation code and reproduced bit for bit.

The six high-budget case studies (Section 6.3) run P4 with seed 6106 and 1,000,000 evaluations per problem, 2,000,000 on HD200. Seed 6106 was chosen from the ablation as the seed without a P4 failure on any problem and with the best mean HV rank. Because stage transitions follow the consumed fraction of the allowance, a high-budget run is a separate run, not a continuation of a 20,000-evaluation run. The comparison with other methods (Section 6.4) uses the same allowances and accounting, seeds 6101–6110, and an additional common time cap described there.

5.2 Implementation, calibration design, and selected setups

All experiments use the implementation described in Sections 2–3: the de-duplicated portfolio (Section 2.5), the normalized merit (19), the adaptive initial trial and projected sufficient-decrease test (23), and the evaluated-set output (Section 3.4). The three-objective MGDA subproblem and the covariance-based branch proposals use faster computations with the same results or sampling distribution.

Calibration varies seven factors per problem: initial population $N_0 \in \{40, \dots, 1280\}$ (doubling), direction radius $R \in \{0.5, 1, 2.5, 5, 10\}$, step ceiling $s_{\max} \in \{1, 2, 4\}$, branching interval $R_{\text{branch}} \in \{2, 4, 8\}$, exploratory survivors $s_b \in \{1, 2, 3\}$ out of three candidate geometries, intermediate seed retention $q \in \{0.5, 0.75, 1\}$, and the allocation rule. The rule is either the count rule (24) or the reward rule (25). The remaining parameters are fixed per problem at the earlier setup; their values are listed in Table 1 and in the common settings below. Configurations are raced by successive halving, in the spirit of racing-based configuration [39]: 64 Latin-hypercube configurations and a fixed control configuration run on seeds 1101–1103; the 32 best by mean HV rank run on 1104–1106 as well; and the 16 best also run on 1107–1110. The selected configuration is the finalist with the highest median HV over the ten seeds; finalists within 0.25 percentage points of it are tied and ranked by smaller interquartile range, then by smaller median IGD. Two documented choices depart from this rule. On HD200, all 16 finalists lie within the tie band, and the interquartile ranges differ only at the 10^{-3} level, so configuration c40 was preferred to the rule’s choice because it had both the higher median and the higher minimum, and it was better on 8 of 10 seeds. On Kursawe, c07 ($N_0 = 640$) was preferred to the configuration with the highest median (96.35%, minimum 89.99%) because its worst seed and interquartile range were more favorable.

The response surfaces in Section 6.1 use the 64 stage-1 configurations (three seeds each; control excluded). A quadratic ridge model in the seven coded factors ($\log_2 N_0$, $\ln R$, $\log_2 s_{\max}$, $\log_2 R_{\text{branch}}$, s_b , q , and the allocation rule as a binary factor) is fitted to the $\log_{10}$ HV deficit $\log_{10}(100 + 10^{-3} - \text{HV})$ and back-transformed for display. This transformation keeps predictions below the reference value and prevents the few collapsed configurations from dominating the fit. The ridge strength is selected by leave-one-configuration-out cross-validation; predictive R_{cv}^2 uses nested holdout of entire configurations. These models describe parameter trends; the selected setups are measured configurations, not surface optima.

Table 1 lists the selected configurations, which are used unchanged by the ablation and the high-budget runs. Initial positions are uniform in the benchmark box, with one mode per seed assigned cyclically across the ordered portfolio. Each branching event proposes three candidate modes, performs three exploratory steps per candidate, and gives six commit steps to each retained candidate. The four stages use seed-retention fractions $(1, q, q, 1)$, subject to the zonal diversity rule. Within-seed pruning retains one branch at its configured interval R_{prune} .

Common settings are Armijo coefficient 10^{-4} , backtracking factor 0.5, and at most 20 reductions; metric regularization $\lambda = 10^{-3}$, Gram mixture $\theta = 0.8$, low-rank cutoff two, and accumulator/secant decay 0.9; history and attraction weights 0.15 and 0.10, momentum decay/mixing 0.85/0.15, descent guard 0.25, and direction-norm factor five. Repulsion has weight 0.03 and eight neighbors; proposal construction uses 16 objective-space neighbors. Suppression is eligible after half of the allowance, with bad-count threshold 2,000 and equivalence tolerance 0.02; it was never triggered in the ablation. The run

Table 1 Selected PHASE-MOO setups (calibration of Section 6.1), used unchanged in the ablation and the high-budget runs. R : direction radius; $s_{\max}$: step ceiling; s_b : exploratory survivors; q : intermediate stage retention; p and ε_{sel} : exponent and floor of (24), unused by the reward rule. Last column: median HV (%) of the configuration over the ten calibration seeds.

Problem	N_0	R	$s_{\max}$	R_{branch}	R_{prune}	s_b	q	allocation	p	ε_{sel}	HV
Kursawe (c07)	640	5	2	4	40	1	0.5	counts	20	10^{-6}	95.71
Viennet (c01)	320	5	2	4	40	2	0.75	counts	20	10^{-60}	99.93
Fonseca–Fleming (c40)	40	2.5	2	2	40	1	0.5	reward	–	–	99.73
DTLZ2 (c02)	640	0.5	1	4	20	1	0.5	counts	20	10^{-60}	98.24
Rugged2D (c06)	80	10	4	2	5	2	0.75	counts	1	10^{-60}	99.84
HD200 (c40)	40	2.5	2	2	5	1	0.5	reward	–	–	99.98

manifests retain the remaining numerical floors, archive settings, source hashes, actual evaluation counts, and stopping reasons.

The comparators of Section 6.4 use default settings fixed before the battery. NSGA-II, MOEA/D, and SMS-EMOA are run in pymoo 0.6.2 [40] with population size 100, initialized with the first 100 points of the common seeded pool; all use simulated binary crossover (probability 0.9, distribution index 15) and polynomial mutation (individual probability one, per-variable probability $1/n$, distribution index 20). MOEA/D uses 100 reference directions, 15 neighbors, neighborhood mating probability 0.9, and the Tchebycheff aggregation; two-objective directions are equally spaced and three-objective directions use the energy construction with seed 19. SMS-EMOA produces one offspring per update and uses exact hypervolume contributions. NSMA and the FD methods use the authors’ defaults with 100 initial points, COMO-CMA-ES uses 100 kernels with initial step size 0.2 in the box-normalized space, and the UHV hybrid uses its default operators with a solution set of 100 points (30 on HD200). These runs are separate from those used to construct the reference fronts.

5.3 Benchmarks

The suite contains six complementary problems. Kursawe uses $[-5, 5]^3$, two objectives, and a disconnected trade-off structure [41]. Viennet uses $[-3, 3]^2$ and three nonconvex objectives [42]. Fonseca–Fleming uses $x \in [-4, 4]^3$ and

$$f_1(x) = 1 - \exp \left[- \sum_{j=1}^3 (x_j - 3^{-1/2})^2 \right], \quad f_2(x) = 1 - \exp \left[- \sum_{j=1}^3 (x_j + 3^{-1/2})^2 \right], \quad (31)$$

with a smooth two-objective Pareto set [31]. DTLZ2 uses $[0, 1]^{12}$ and three objectives to test a scalable spherical front [32]. The implemented Kursawe absolute-power term uses $(x_j^2 + 10^{-12})^{0.4}$ for differentiability. Analytic Jacobians of the implemented objectives are used throughout.

Rugged2D and HD200 use $x \in [-4, 4]^n$, with $n = 2$ and $n = 200$, respectively, and

$$r(x) = 1 + \rho [\sin(\omega z_1) \sin(\omega z_2)]^2, \quad (32)$$

$$f_1(x) = \|x - a\|^2 r(x) + 10^{-8}, \quad f_2(x) = s \|x - b\|^2 r(x) + 10^{-8}. \quad (33)$$

Here $a = (2, 0, \dots, 0)$, $b = (-2, 1, 0, \dots, 0)$, $s = 120$, $\rho = 0.15$, and $\omega = 6$. Rugged2D takes $z = x$. HD200 takes $z = U^T x$, where the two orthonormal columns of U are obtained by a reduced QR factorization of a 200×2 standard-normal matrix generated with NumPy’s default generator and seed zero. The fixed projection defines the problem and is independent of optimization seeds.

5.4 Construction of the Pareto-front references

All quality comparisons use fixed reference sets, independent of the evaluated run. Fonseca–Fleming uses 5,000 samples of its analytic Pareto set, $x = t\mathbf{1}$, $t \in [-1/\sqrt{3}, 1/\sqrt{3}]$. DTLZ2 uses a 70×70 angular grid on the positive octant of the unit sphere. Kursawe and Viennet use archived reference samples containing 4,060 and 5,322 objective vectors, respectively.

For Rugged2D and HD200, separate NSGA-II, MOEA/D, and SMS-EMOA runs use five seeds per algorithm, population size 400, and nominal objective-evaluation budgets of one million and two million,

respectively. Their checkpoint and final objective vectors are pooled and nondominated-filtered, yielding reference sets of 12,776 and 7,814 points. PHASE points are excluded from these two references. These empirical targets supply a common scoring convention, not a certified exact front; reference-construction runs are distinct from the individual-algorithm comparisons at 20,000 evaluations. The fixed samples, normalization bounds, and file hashes are retained with the experimental records.

5.5 Performance indicators

For each PHASE run, P is the unique nondominated set of all points evaluated during the run, which contains the final endpoints (the set $\mathcal{E}$ of Section 3.4). The approximations returned by the comparators are described in Section 6.4. Every treatment on a benchmark is evaluated against the same reference set and coordinate normalization,

$$\hat{f}_i = (f_i - a_i)/(b_i - a_i), \quad a_i = \min_{z \in R} z_i, \quad b_i = \max_{z \in R} z_i. \quad (34)$$

The main comparisons report mean inverted generational distance IGD_1 and exact hypervolume in two or three objectives with the common normalized reference point $r_H = (1.1, \dots, 1.1)$ [43–46]. With the references in Section 5.4, the distance indicator is

$$\text{IGD}_1(P; R) = \frac{1}{|R|} \sum_{z \in \hat{R}} \min_{y \in \hat{P}} \|z - y\|_2. \quad (35)$$

The reference hypervolume ratio is

$$H_{\text{rel}}(P; R) = 100 \frac{\text{HV}(\hat{P}; r_H)}{\text{HV}(\hat{R}; r_H)}. \quad (36)$$

Here $\hat{P}$ and $\hat{R}$ use the same reference-derived normalization. The ratio is relative to a finite sample and need not be bounded above by 100%. Medians and interquartile ranges summarize the runs in each experimental block: 20 seeds (P3–P5) or ten seeds (P1, P2) in the ablation, three seeds per configuration in stage 1 of the calibration and ten for its finalists, and ten seeds (6101–6110) in the comparison with other methods. Reference definitions and normalization are common to all methods within each comparison.

5.6 AI-assisted research and manuscript preparation

OpenAI Codex and Claude (Anthropic) assisted with source-code inspection, programming and numerical checks, figure-generation scripts, manuscript drafting, and language editing. Both authors reviewed and approved the work and take responsibility for the scientific content, reported results, and final manuscript.

6 Results and discussion

6.1 Calibration: which parameters matter

The calibration identifies a small number of influential parameters and a large set of near-equivalent configurations. Fig. 1 reports the stage-1 share of variance in mean HV explained by each factor. The initial population dominates on DTLZ2 ($\eta^2 = 0.92$) and, together with the direction radius, on HD200 (0.51 and 0.43); the radius is the main factor on Fonseca–Fleming (0.43); Viennet responds mainly to N_0 and the branching interval (0.35 and 0.21); and Rugged2D responds to N_0 , R , R_{branch} , and s_b in similar proportions. In contrast, the step ceiling, the retention fraction, and the allocation rule remain at the chance level of the design on every problem. In particular, the choice between count-based and reward-based allocation explains at most 4% of the variation. This observation anticipates the ablation result that the allocation rule has no practical effect once the portfolio is in place.

Fig. 2 shows the fitted response maps for the three influential structural parameters. The fitted models are predictive on DTLZ2 ($R_{cv}^2 = 0.92$), Rugged2D (0.82), HD200 (0.75), and Fonseca–Fleming (0.59), as confirmed by the nested holdout predictions in Fig. 4. On DTLZ2, HV increases monotonically with N_0 across the whole radius range; the selected configuration combines a large population ($N_0 = 640$) with the smallest radius. HD200 shows the opposite population effect: the best region combines a small initial population with a moderate radius, and the fitted HV falls by up to about 70 percentage points

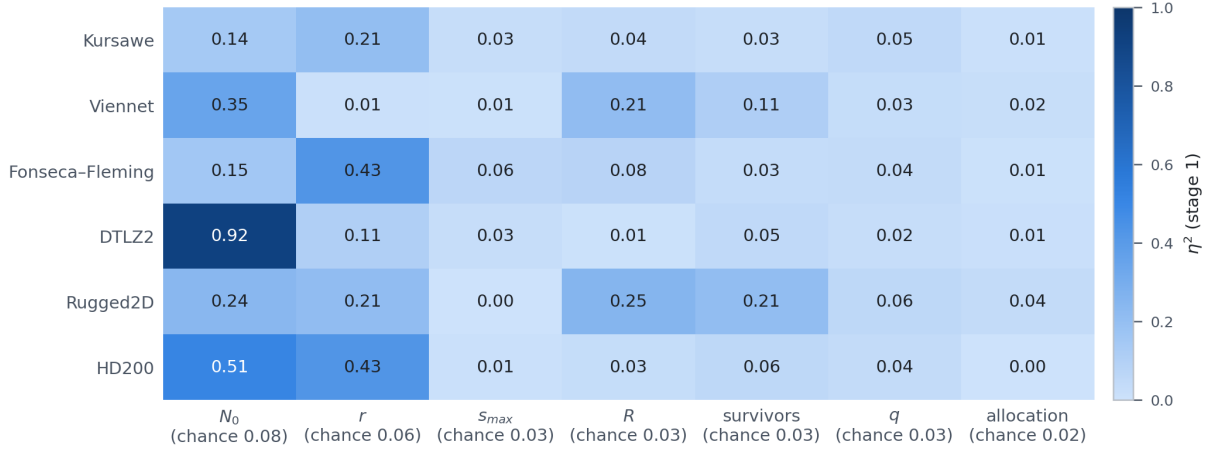


Fig. 1 Stage-1 factor importance in the calibration: η^2 of mean HV for each factor and problem (64 configurations, seeds 1101–1103). Under this design, a factor with L levels and no effect has an expected η^2 of about $(L - 1)/63$, shown under each label. The step ceiling s_{max} , the retention q , and the allocation rule stay at that chance level on every problem. Here r is the direction radius R and R the branching interval R_{branch} .

towards large populations within the design hull, presumably because each of many seeds then receives too few of the 100,000 evaluations to reach the front. On Fonseca–Fleming, the favorable region is again at small N_0 , now with larger radii, whereas Rugged2D varies by less than 0.1 percentage point over the hull, so that its parameters are mainly a matter of fine tuning. For Viennet and Kursawe the models are not predictive ($R_{cv}^2 = 0.30$ and -0.03), and the maps show observed cell means instead. On Viennet all configurations lie between 99.78% and 99.94%, so the response is flat. On Kursawe, observed means range from 83% to 96% without a consistent trend: its variation is driven mainly by whether a run finds the disconnected piece of the front, an event that the parameters influence only weakly (Section 6.2).

Fig. 3 presents the fitted HV as three-dimensional surfaces. The population effect spans tens of percentage points on HD200, a few on DTLZ2, and hundredths on Rugged2D. The star at the observed HV of the selected setup need not lie on the surface, which is a slice at the design center rather than at the selected configuration. Additional radius–branching–interval surfaces confirm the trends in Fig. 2; the fitted $\log_{10}$ IGD maps identify similar favorable regions and are predictive on Viennet ($R_{cv}^2 = 0.67$) and Fonseca–Fleming (0.61), with both analyses omitted here for brevity.

The selected setups (Table 1) attain ten-seed median HV between 98.24% (DTLZ2) and 99.98% (HD200) on five problems. Kursawe reaches 95.71%, with a minimum of 92.19% and an interquartile range of 2.45 percentage points. Its median is almost eight points above the control configuration carried through the racing (87.99%), but the run-to-run variation remains. Two settings recur: HD200 and Fonseca–Fleming share the same configuration, and Kursawe, DTLZ2, and Viennet all use $R_{branch} = 4$ with large initial populations. The calibration therefore narrows the relevant choices to N_0 and R , with N_0 adjusted to the budget per seed.

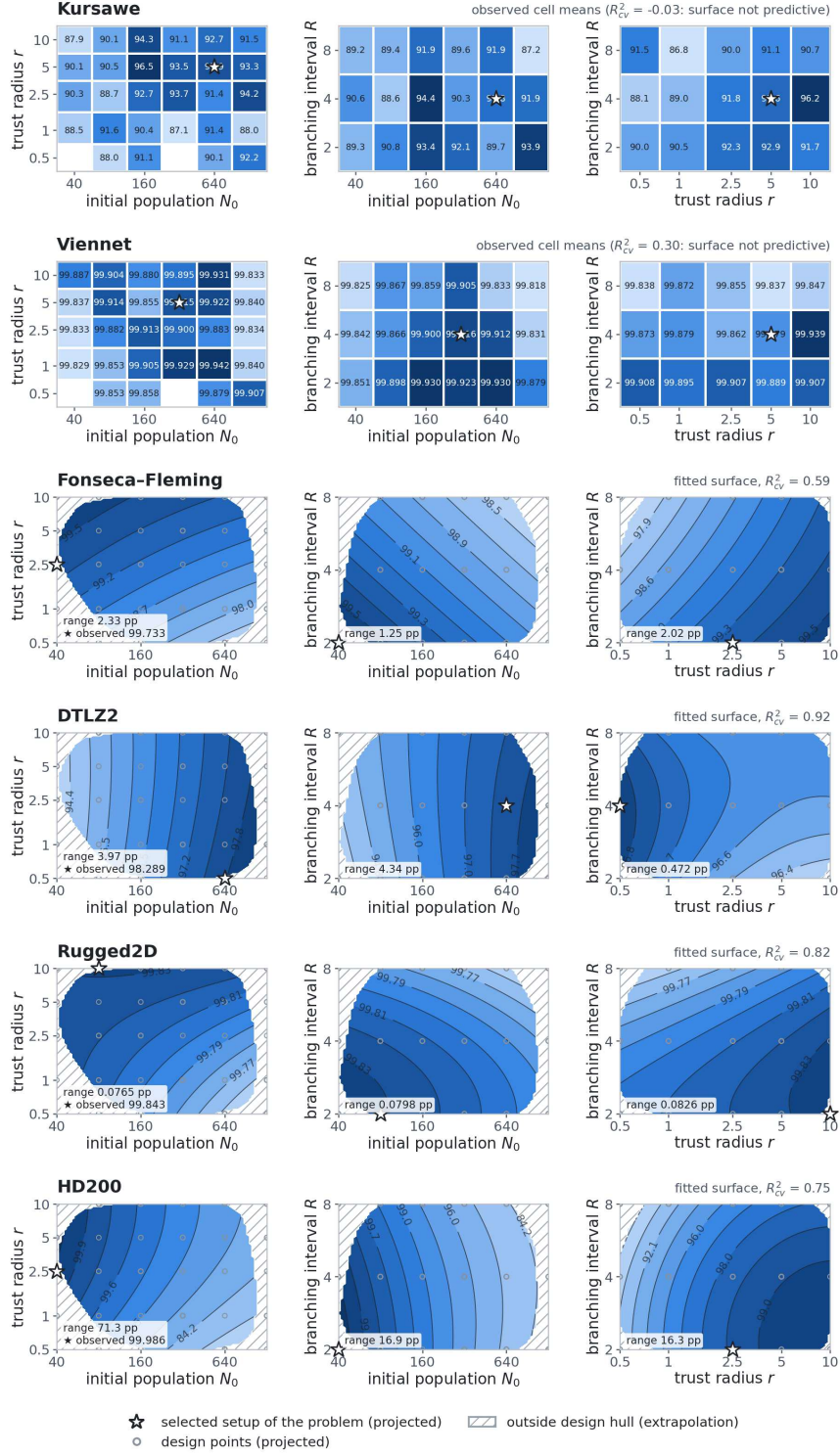


Fig. 2 Fitted response maps of the calibration. Rows: problems; columns: initial population $N_0 \times$ direction radius $r (= R)$, $N_0 \times$ branching interval $R (= R_{\text{branch}})$, and $r \times R$. Colors and labeled contours give the mean HV (% of the reference, darker is better) predicted by a quadratic ridge model of the $\log_{10}$ HV deficit, fitted to the 64 stage-1 configurations (three seeds each) and back-transformed to HV; the remaining factors are held at the design center ($N_0 \approx 226$, $r \approx 2.24$, $R = 4$, $s_{\text{max}} = 2$, two survivors, $q = 0.75$, allocation midway). The star marks the selected setup of each problem in that plane and is annotated with its observed stage-1 mean HV; its other factors differ from the slice, and slices through the selected setups fall outside the design hull. Open circles are design points projected onto each plane; hatched areas lie outside the convex hull of the design. R_{cv}^2 is the nested leave-one-configuration-out predictive R^2 , and “range” is the span of the fitted response inside the hull (color scales are local to each panel). Where $R_{cv}^2 < 0.5$ (Kursawe, Viennet), cells show the observed mean HV of the design points at each level pair, averaged over the other factors. The maps describe calibration trends, not independently validated optima.

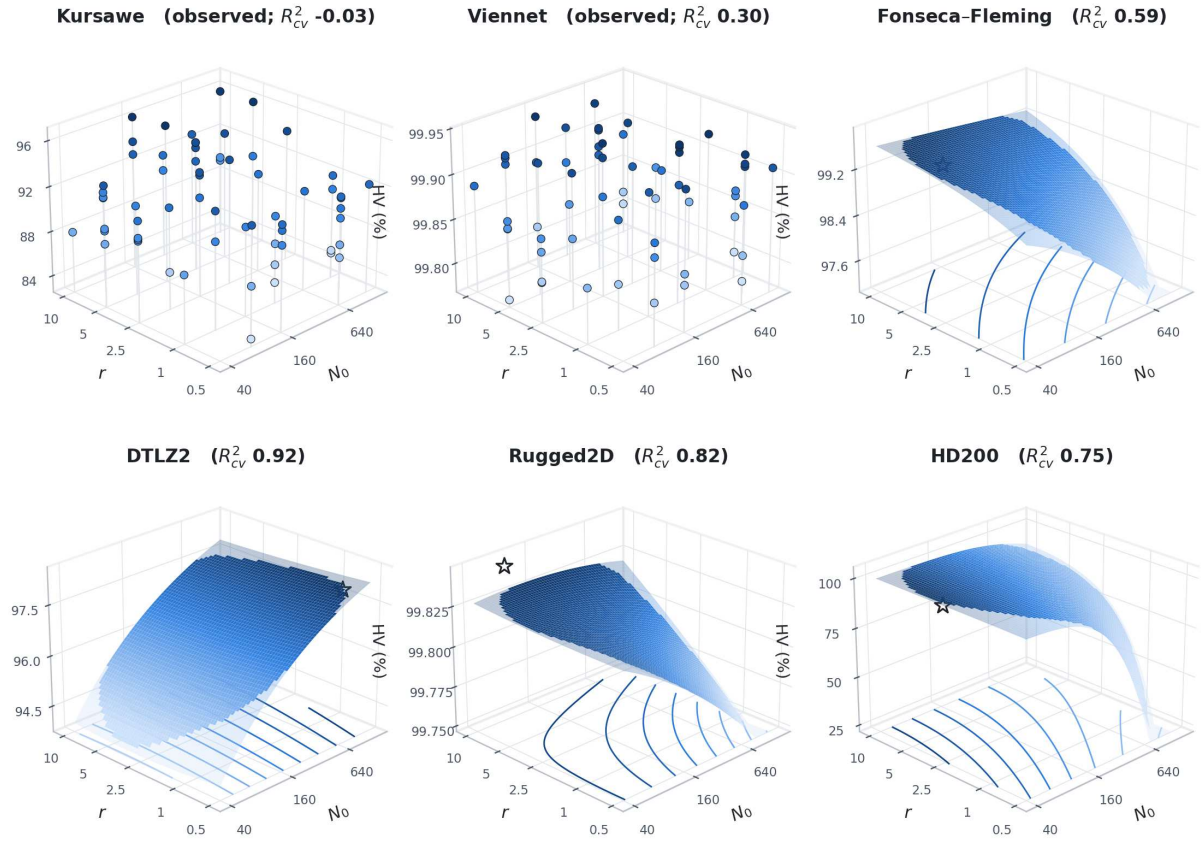


Fig. 3 Fitted mean HV of the calibration as a function of initial population N_0 and direction radius r ($= R$), shown as three-dimensional surfaces with contour projections; other factors at the design center as in Fig. 2. Colors follow the deficit scale; faded facets lie outside the design hull. The star is drawn at the observed stage-1 mean HV of the selected setup. Kursawe and Viennet ($R_{cv}^2 < 0.5$) show the observed configuration means as points.

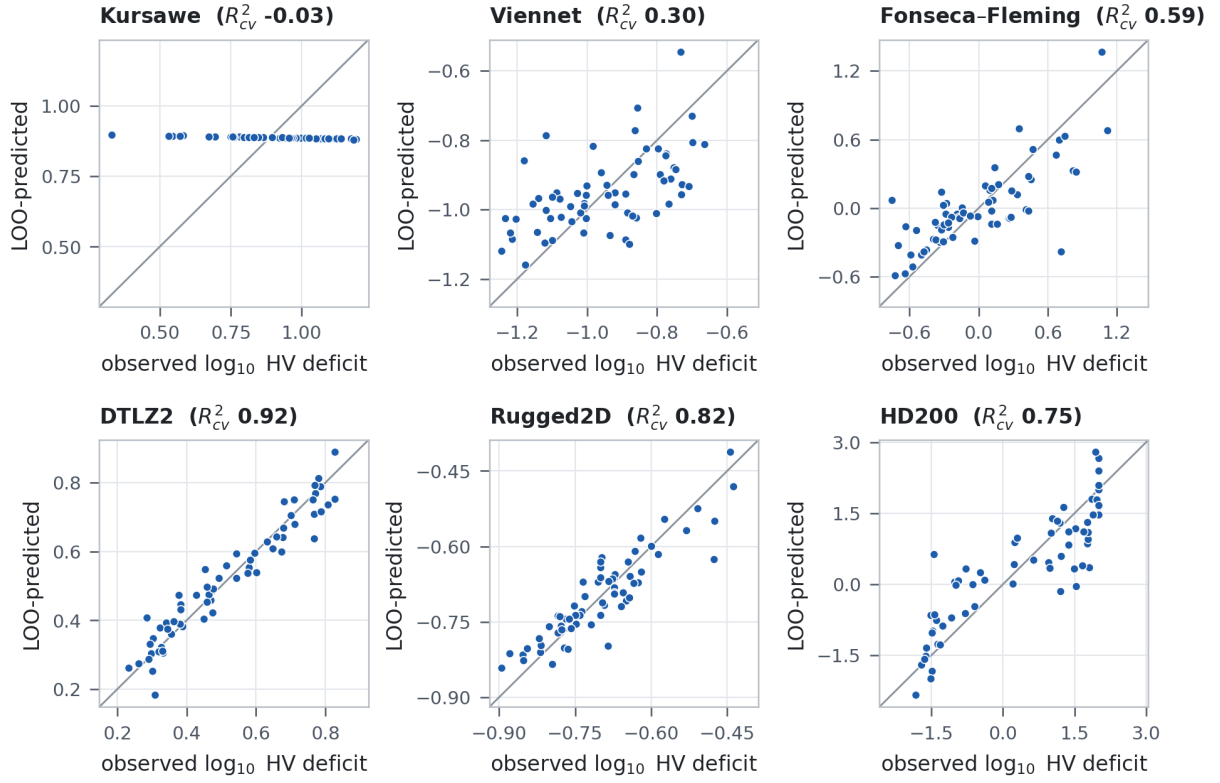


Fig. 4 Predictive quality of the calibration response models: observed $\log_{10}$ HV deficit, $\log_{10}(100 + 10^{-3} - \text{HV})$, of each stage-1 configuration versus its nested leave-one-configuration-out prediction (lower is better). On the HV scale the back-transformed predictions keep the ordering but not the magnitude of collapsed configurations (HD200 has a few configurations with HV below 40%), so the surfaces should be read as trends, not as point predictions of HV.

6.2 Ablation: portfolio diversity rather than adaptive selection

Fig. 5 shows the final HV of every ablation run. The multi-geometry population (P3, P4) is better than both identity controls on all six problems. Relative to P4, the median paired HV difference of MGDA-I (P1) ranges from -2.1 percentage points on Viennet to -84.4 on HD200, and that of PHASE-I (P2) from -0.48 on Rugged2D to -57.6 on HD200. All twelve contrasts are significant (Wilcoxon signed-rank, Holm correction over the four contrasts per problem and indicator, $p_{\text{Holm}} \leq 0.03$), and the corresponding IGD contrasts are significant on all problems except PHASE-I on Kursawe ($p_{\text{Holm}} = 0.08$). The identity-only population already improves on MGDA-I on five problems, but it cannot replace the portfolio: on Fonseca–Fleming its median HV is 57.3% against 99.8%, and on HD200 42.4% against 99.99%. The P4 medians are 99.987% on HD200, 98.30% on DTLZ2, 99.936% on Viennet, 99.78% on Fonseca–Fleming, 99.84% on Rugged2D, and 94.20% on Kursawe, with median IGD between 3.1×10^{-4} (HD200) and 2.3×10^{-2} (Kursawe, DTLZ2). These held-out values agree with the calibration medians in Table 1 to within 0.06 percentage points on the five stable problems and 1.5 points on Kursawe.

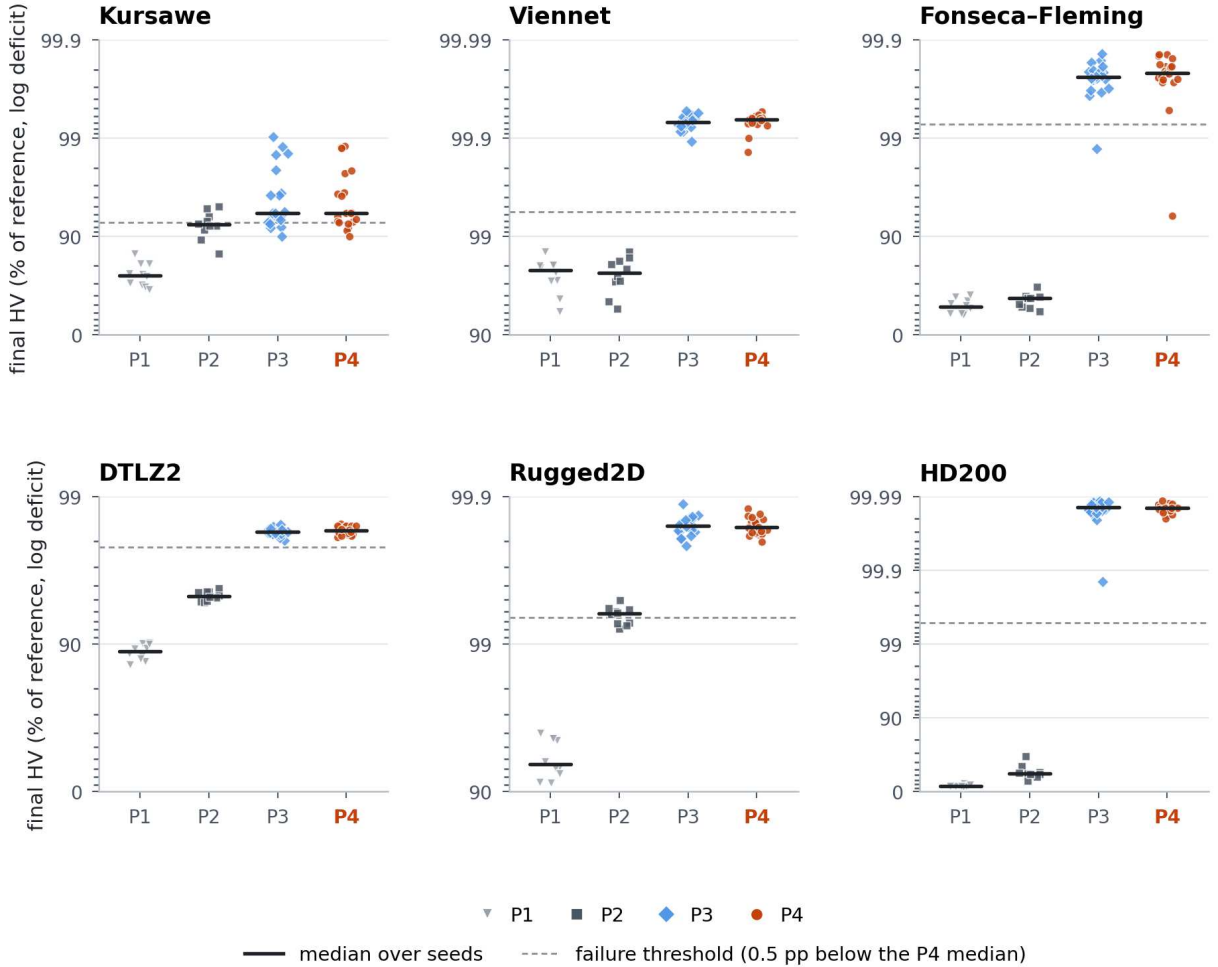


Fig. 5 Ablation without calibration seeds: final HV of each run (% of the reference HV). P1: multistart identity MGDA from the same initial points, with a fixed initial step; P2: PHASE-I (identity geometry only); P3: de-duplicated portfolio with uniform allocation; P4: PHASE-MOO (calibrated setup, de-duplicated portfolio). P3 and P4 use seeds 6101–6120, P1 and P2 seeds 6101–6110; budgets 20,000 joint evaluations (HD200 100,000). The vertical axis is the HV deficit $100 - \text{HV}$ on a logarithmic scale, inverted (better is up) and labeled in HV. Points are seeds; bars are medians; the dashed line is the failure threshold, 0.5 percentage points below the P4 median.

The comparison of P3 with P4 isolates the allocation rule, because both use the same de-duplicated portfolio and setup. None of the twelve P3–P4 contrasts is significant. Since absence of significance does not establish absence of an effect, the claim of equivalence was tested against margins fixed before the run: two one-sided Wilcoxon signed-rank tests (TOST) must both reject at level 0.05, and the 90% bootstrap confidence interval of the median paired difference must lie inside $\pm\delta$. The HV margin is $\delta = \max(0.05, 0.5 \times \text{seed IQR})$ percentage points, and the IGD margin is $\delta = \max(10^{-4}, 0.1 \times \text{median IGD})$,

both from an independent battery (seeds 8201–8210). The resulting HV margins are 0.05 points on four problems, 0.11 on DTLZ2, and 0.70 on Kursawe. Both criteria hold, so uniform and adaptive allocation are equivalent, in HV on all six problems ($p_{\text{TOST}} \leq 0.004$) and in IGD on HD200, DTLZ2, Fonseca–Fleming, and Rugged2D (Fig. 6). Equivalence is not established for IGD on Viennet ($p_{\text{TOST}} = 0.08$), whose interval extends beyond the margin on the side favoring P4, and on Kursawe ($p_{\text{TOST}} = 0.09$). The bridge variant P5, without de-duplication, differs from P4 significantly only in Viennet HV, where it is lower by a median 0.006 points (3 of 20 seeds better, $p_{\text{Holm}} = 0.015$). De-duplication therefore leaves results essentially unchanged, while it shortens the portfolio and removes redundant competitors.

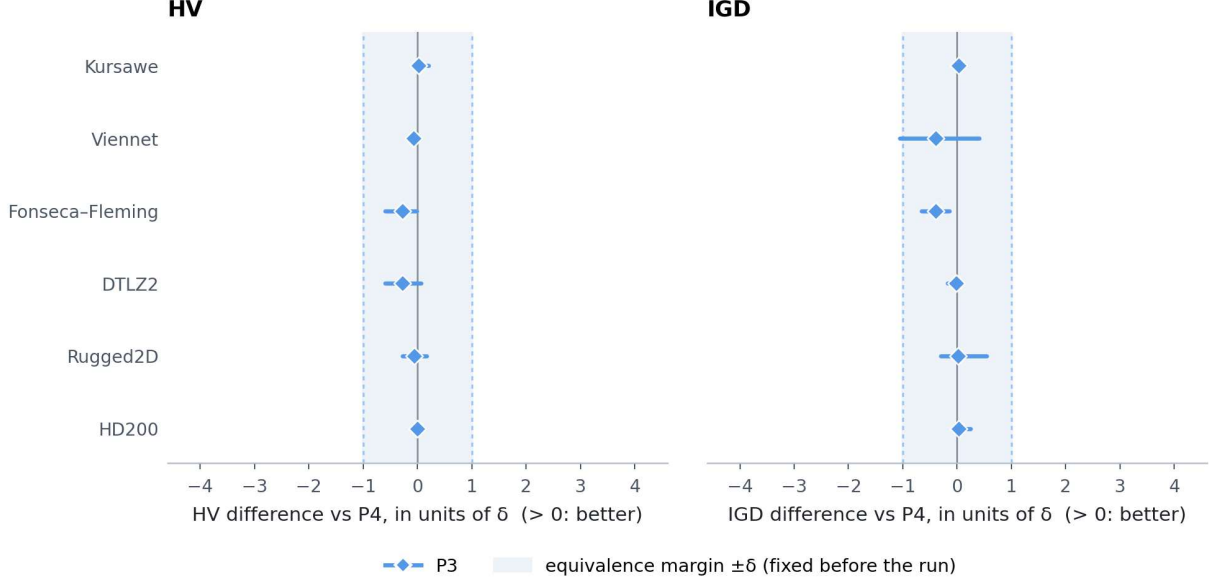


Fig. 6 Equivalence of uniform (P3) and calibrated (P4) allocation. Paired differences P3–P4 divided by the equivalence margin δ fixed before the run (HV: $\max(0.05 \text{ pp}, 0.5 \times \text{seed IQR})$; IGD: $\max(10^{-4}, 0.1 \times \text{median IGD})$, both from an independent battery). Markers: median paired difference over 20 seeds; bars: 90% bootstrap confidence interval. IGD differences are sign-flipped so that positive always means better than P4. The shaded band ± 1 marks the margin. Equivalence is declared only when the two one-sided Wilcoxon tests both reject at level 0.05 and the interval lies inside the band.

The geometry shares of P4 explain this equivalence (Fig. 7). Each geometry receives between 0.71 and 1.27 times its uniform share of branch selections (0.86–1.18 on the four problems with the count rule), and the global suppression rule was never triggered. Diagnostic runs with the same implementation (not shown; seeds disjoint from the ablation seeds) pointed to the cause. When a single geometry is used alone, the quality varies strongly between geometries, but the uniform portfolio is better than the best single geometry on every problem. Removing any one geometry changes the results little. Moreover, the outcome of a candidate competition is determined mainly by its starting position rather than its geometry, so selection credit is confounded with position. The portfolio contributes through the complementarity of geometries, and adaptive reweighting has little left to exploit. We therefore present the allocation rule as an option of the architecture rather than as a source of the observed gains; with the de-duplicated portfolio, uniform allocation is an equally good and simpler default.

The anytime curves (Fig. 8) show how the portfolio uses the budget. P3 and P4 improve steadily up to the full allowance. P2 reaches its final HV within the first 10% of the budget on Viennet, DTLZ2, and HD200 and gains little afterwards. Its identity-only branches apparently settle at a small set of critical points and add few new nondominated points, whereas competing geometries keep generating new trade-offs. On HD200, P4 reaches a median of 76.7% after 10% of the evaluations, 99.76% after 25%, and 99.99% at the end.

Robustness defines the limitations of the method (Fig. 9). A run fails when its HV is more than 0.5 percentage points below the P4 median. With this definition, P4 has no failure on HD200, DTLZ2, Viennet, and Rugged2D. On Fonseca–Fleming, one of 20 runs (seed 6115, HV 93.79%) misses an extreme portion of the front. On Kursawe, four of 20 runs fail, with a worst HV of 89.92%, and the HV of all runs spreads from 89.9% to about 99%. In the median and worst Kursawe runs of Fig. 9, the disconnected piece next to the f_2 -minimizing extreme is represented only by isolated points; the HV of a run depends mainly on how much of this piece it recovers. MGDA-based descent stops at Pareto-critical points, so

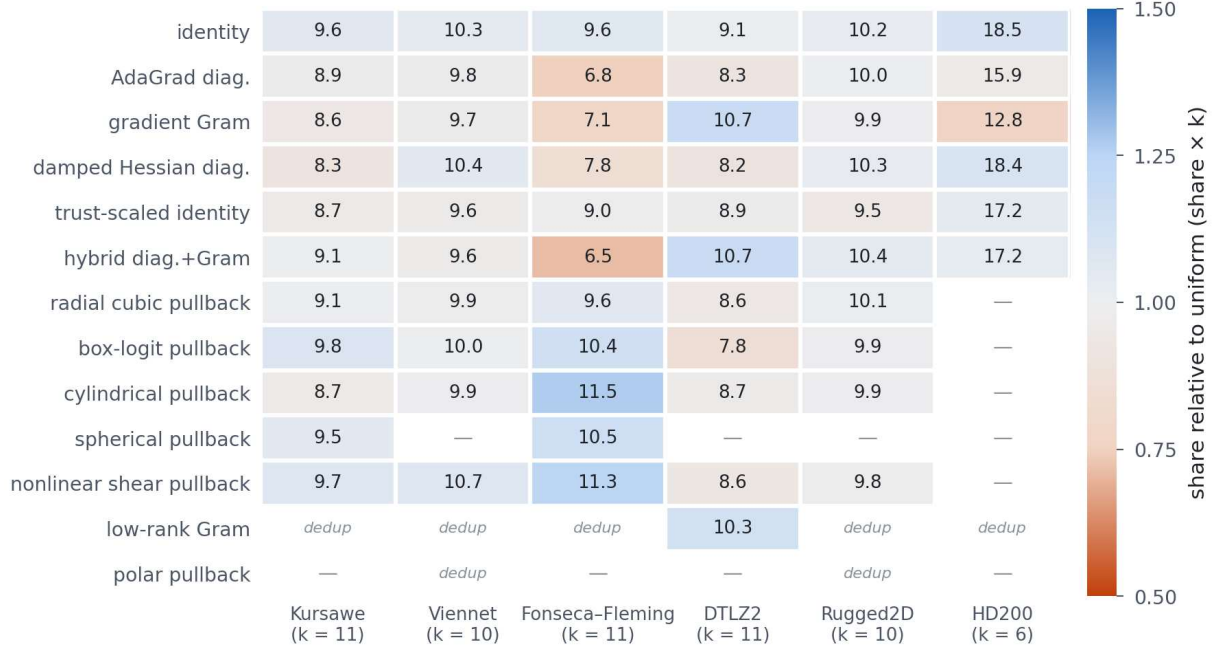


Fig. 7 Share of branch selections per geometry in P4 (numbers: % of selections, mean over 20 seeds; color: share $\times k$, where k is the size of the de-duplicated portfolio, so that 1 means uniform). “dedup” marks geometries removed by de-duplication for that problem (low-rank Gram when its rank is at least $\min(m, n)$; polar pullback when $n = 2$); “—” marks geometries that are not in the portfolio of that problem. Fonseca-Fleming and HD200 use the reward rule, the other problems the count rule (24).

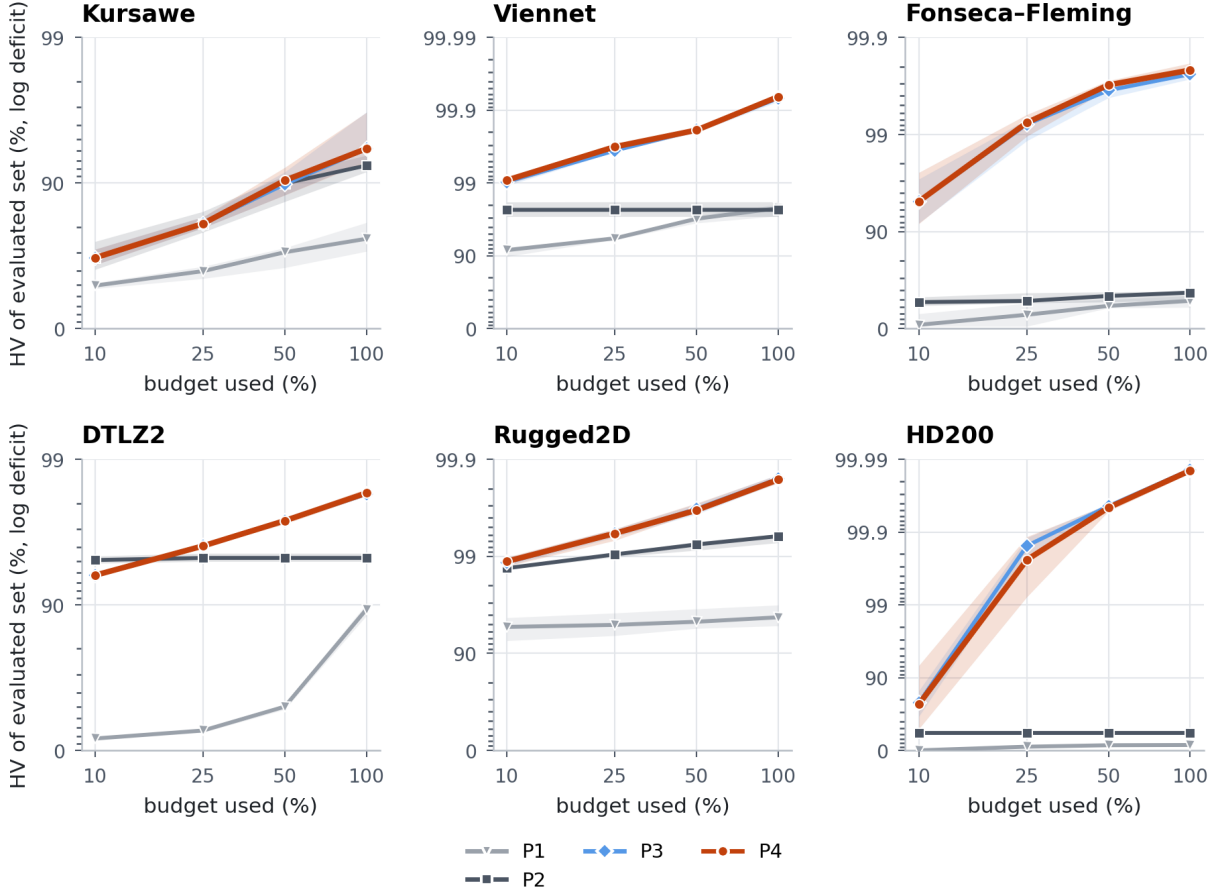


Fig. 8 Anytime behavior in the ablation: HV of all points evaluated up to 10, 25, 50, and 100% of the budget (median and interquartile band over seeds; same logarithmic deficit axis as Fig. 5).

coverage of a disconnected piece or extreme depends on whether some branch starts in, or is relocated to, its basin. P3 has the same failure counts (4/20 and 1/20), confirming that the limitation is not caused by the allocation rule. Section 6.3 shows that the missing piece is found when more evaluations are available.

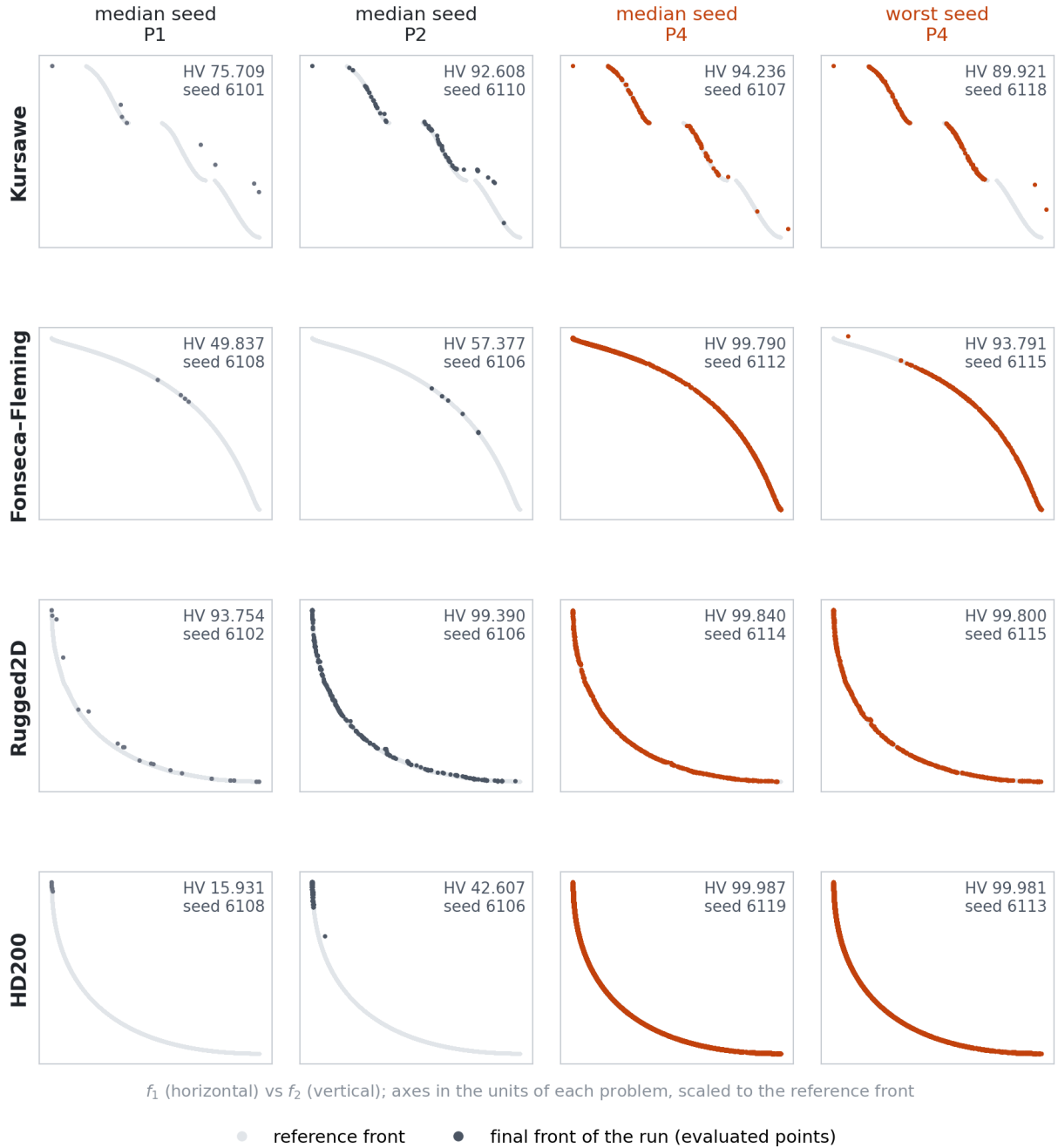


Fig. 9 Final fronts of the bi-objective ablation problems against the reference front (gray): median-HV seed of P1, P2, and P4, and the worst P4 seed. Axes are scaled to the reference front of each problem. The worst Kursawe seed misses the disconnected piece next to the f_2 -minimizing extreme, and the worst Fonseca-Fleming seed misses part of the upper-left extreme; these are the limitations discussed in the text.

6.3 High-budget case studies

The high-budget runs show what P4 converges to when the allowance is not limiting (Table 2). With seed 6106 and one million evaluations (two million on HD200), the reference HV ratio is between 100.003% and 100.04% on all six problems, and IGD is 8.7×10^{-3} on DTLZ2 and at most 1.7×10^{-4} elsewhere. Values above 100% are possible because the references are finite samples of earlier runs (Section 5.4); they indicate that the approximation dominates parts of the space left uncovered by the reference sample.

On Kursawe, all disconnected parts of the front are covered at one million evaluations, so the limitation observed at 20,000 evaluations is a budget effect for this seed. The anytime values in Table 2 show that most of the gain is obtained early: at 5% of the allowance, HV is already at least 98.97% on all problems.

Table 2 High-budget P4 runs (seed 6106). HV and IGD of the returned approximation; ND: number of nondominated points returned; branches: active branches at the end; HV at 20k and 5%: HV of all points evaluated after 20,000 evaluations and after 5% of the allowance; time: wall-clock minutes on the Windows workstation used for all runs.

Problem	budget	HV (%)	IGD	ND	branches	HV at 20k	HV at 5%	time
Kursawe	10^6	100.004	1.12×10^{-4}	83,420	164	98.790	99.512	5
Viennet	10^6	100.018	1.68×10^{-4}	334,439	380	99.933	99.987	419
Fonseca–Fleming	10^6	100.020	3.17×10^{-6}	581,367	11	99.903	99.999	71
DTLZ2	10^6	100.041	8.74×10^{-3}	314,630	224	98.234	98.973	364
Rugged2D	10^6	100.003	4.47×10^{-5}	125,007	99	99.848	99.941	11
HD200	2×10^6	100.009	2.81×10^{-5}	636,175	15	99.779	99.987	114

Figs. 10–15 show, for each problem, the initial points and the complete final nondominated set in decision space (a) and objective space (b). Each start–end link joins an initial point either to an active branch at the end of the run or to a nondominated point whose seed of origin was recovered from the evaluation log. All final branches and all nondominated points are drawn without subsampling. The links are schematic: they connect origin and end point, not the path followed by the optimizer. The returned set is much larger than the set of active branches. On Kursawe, for example, 83,420 nondominated points remain at the end, of which 88 coincide with the 164 active endpoints. A further 46,669 of these points descend from seeds that were later removed from the population. The same pattern holds on all problems (for HD200, 636,175 points against 15 active branches). It reflects how the evaluated-front output works: intermediate steps, exploratory candidates, and line-search trials remain in the approximation as long as they are nondominated, even after their branch or seed has been pruned. Pruning therefore redirects evaluations without discarding the coverage that pruned lineages have already produced. The origins of all 2,075,038 nondominated points were recovered; none was assigned by geometric proximity.

On Kursawe (Fig. 10), links from 617 of the 640 initial points reach all disconnected parts of the front (an isolated extreme point and three curves), and the decision-space panel shows the corresponding separated components of the Pareto set. On Viennet (Fig. 11), initial points across the whole domain feed a nondominated set whose decision-space image consists of several thin, separated components. On Fonseca–Fleming (Fig. 12), only 11 branches remain active from 40 initial points, yet the archive covers the whole front, and the decision-space panel recovers the known Pareto set, the segment $x = t\mathbf{1}$. On DTLZ2 (Fig. 13), the nondominated points cover the octant of the sphere; in the reduced decision coordinates they lie on the plane $r_{\text{tail}} = 0$, i.e., $x_{3:12} = 0.5$, which characterizes the Pareto set. On Rugged2D (Fig. 14), the nondominated points in decision space form several segments separated by the oscillating factor, while the objective-space front is continuous. On HD200 (Fig. 15), the logarithmic axes reveal both extremes of the front: over all 200 coordinates, the distance to anchor a falls to about 10^{-7} and that to b below 10^{-3} , starting from initial points at distances of about 30.

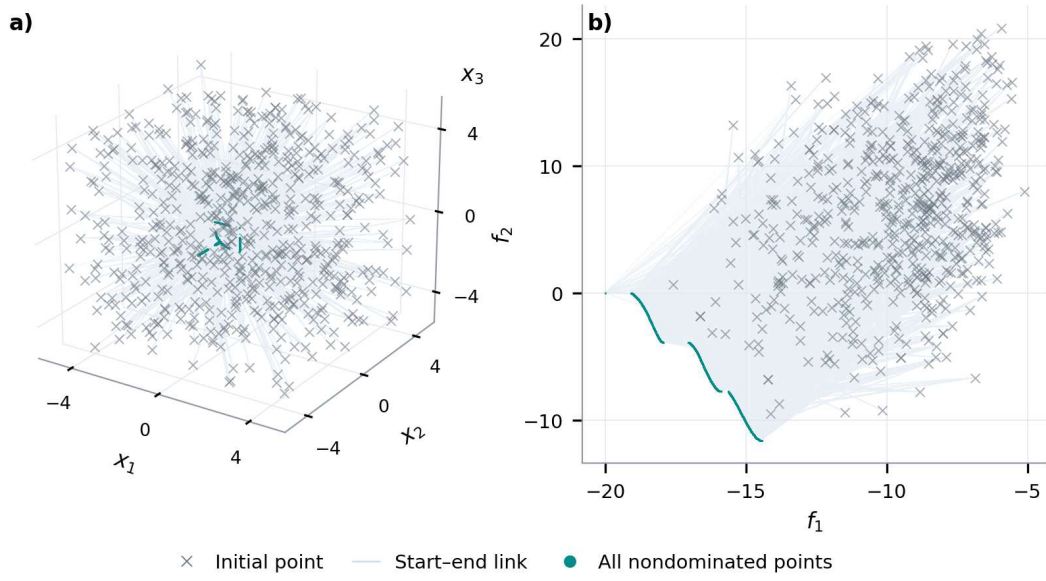


Fig. 10 High-budget PHASE-MOO (P4) on Kursawe, seed 6106, 1,000,000 distinct joint objective–Jacobian evaluations. (a) Decision space (x_1, x_2, x_3); (b) objective space. Crosses: initial points that originate at least one final branch or nondominated point; light lines: schematic start–end links of all 164 final branches and of all 83,420 nondominated points to their seed of origin, recovered from the evaluation log (not optimization paths; all links drawn, 90% transparency); teal: the complete final nondominated set.

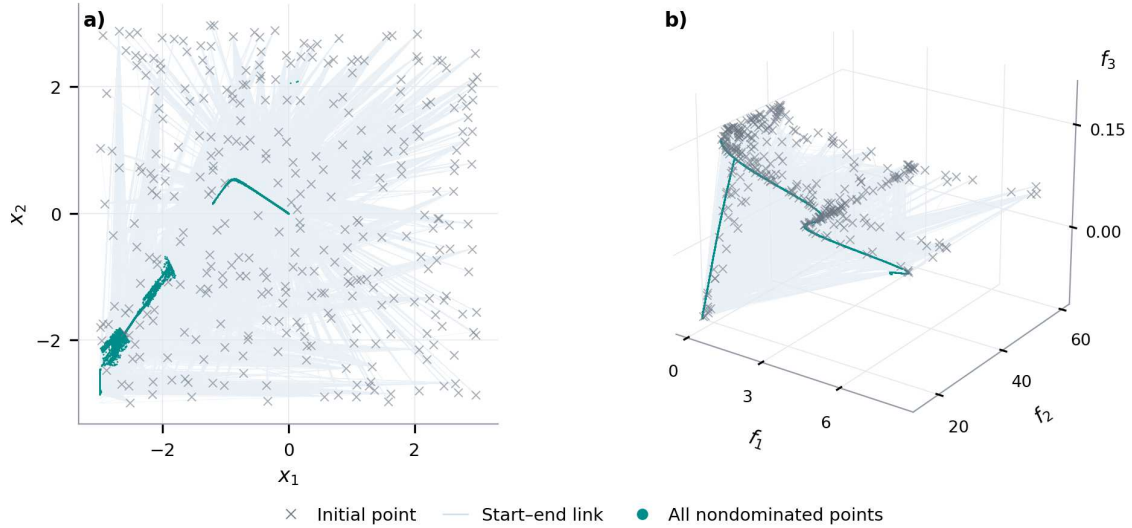


Fig. 11 As Fig. 10, for Viennet (1,000,000 evaluations): 320 initial points, 380 final branches, 334,439 nondominated points. (a) Two-dimensional decision space; (b) three-objective space.

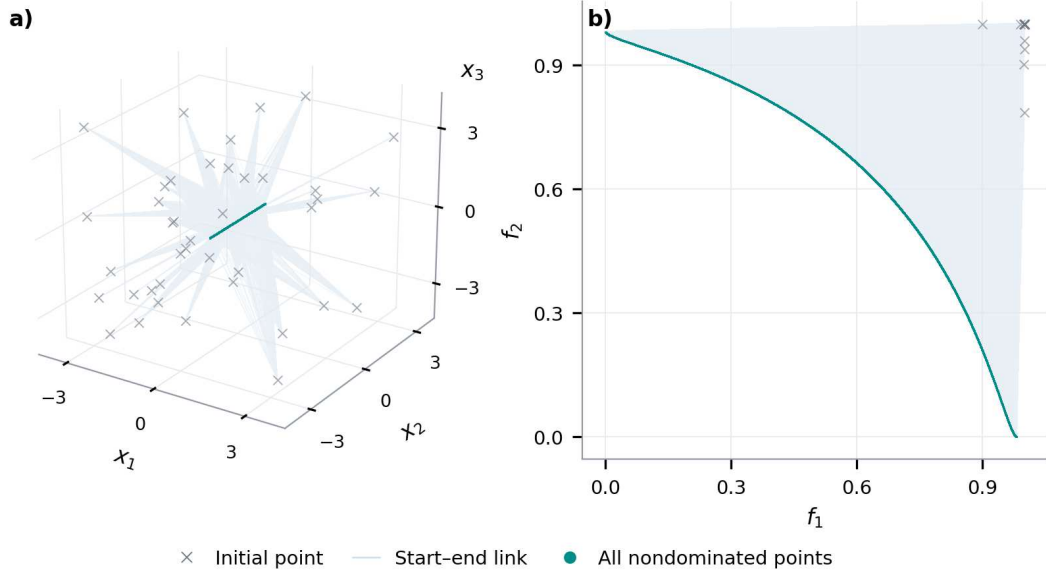


Fig. 12 As Fig. 10, for Fonseca–Fleming (1,000,000 evaluations): 40 initial points, 11 final branches, 581,367 nondominated points. (a) Three-dimensional decision space; (b) objective space.

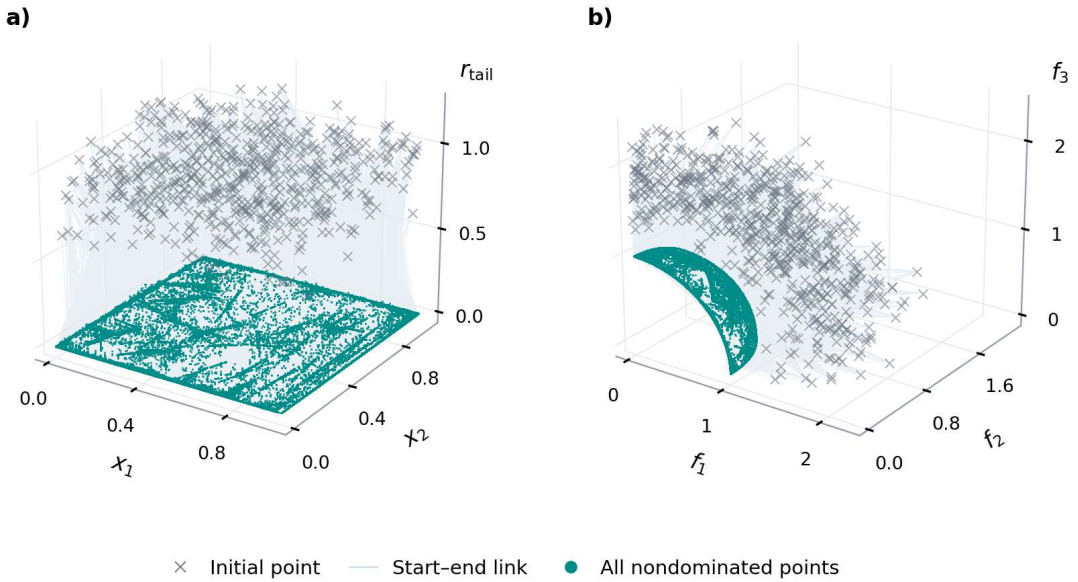


Fig. 13 As Fig. 10, for DTLZ2 (1,000,000 evaluations): 640 initial points, 224 final branches, 314,630 nondominated points. (a) Reduced decision coordinates $(x_1, x_2, r_{\text{tail}})$, where $r_{\text{tail}} = \|x_{3:12} - 0.5\|_2$ summarizes the remaining ten variables (not a complete view of $\mathbb{R}^{12}$); (b) three-objective space.

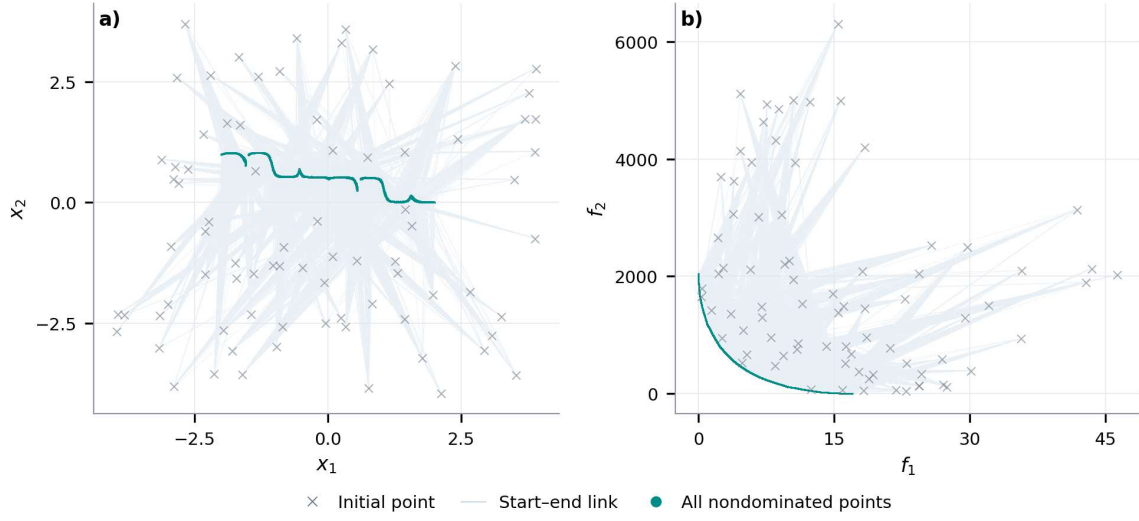


Fig. 14 As Fig. 10, for Rugged2D (1,000,000 evaluations): 80 initial points, 99 final branches, 125,007 nondominated points. (a) Two-dimensional decision space; (b) objective space.

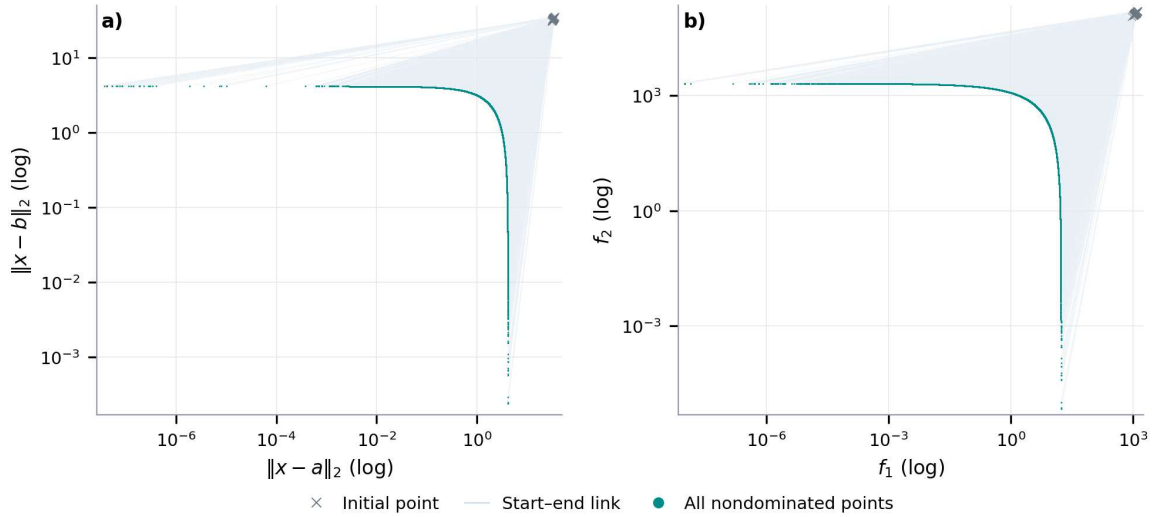


Fig. 15 As Fig. 10, for HD200 (2,000,000 evaluations): 40 initial points, 15 final branches, 636,175 nondominated points. (a) Euclidean distances, over all 200 coordinates, to the anchors $a = (2, 0, \dots, 0)$ and $b = (-2, 1, 0, \dots, 0)$; (b) objective space. Both panels use logarithmic axes, so the links are schematic in the displayed coordinates.

6.4 Comparison with eight methods

PHASE-MOO (P4) was compared with eight methods on the six problems, with ten seeds (6101–6110) per method and problem: the evolutionary algorithms NSGA-II [18], MOEA/D [19], and SMS-EMOA [47]; the memetic algorithm NSMA [23]; the front-descent methods FD-BB and FD-SD [10, 12]; COMO-CMA-ES [25]; and the UHV hybrid of [26], which allocates effort dynamically between an uncrowded-hypervolume evolutionary algorithm and hypervolume gradient ascent. The last five were run from their authors’ implementations at fixed commits. No comparator was tuned: all use the defaults of their authors or of pymoo (Section 5.2), whereas PHASE-MOO uses the setups calibrated on seeds 1101–1110. The 60 PHASE-MOO runs reproduce the P4 ablation runs of the same seeds bit for bit.

All methods share the accounting of Section 5.1: one evaluation per new point at which the objectives and the Jacobian are computed jointly, repeated requests free, and ceilings of 20,000 evaluations (HD200: 100,000). PHASE-MOO, NSMA, the FD methods, and the UHV hybrid use the Jacobian; the evolutionary algorithms and COMO-CMA-ES use only the objectives. The allowance therefore equalizes the number of points queried, not the information obtained per point or the computational cost. Methods that accept initial points start from the same seeded pool of uniform random points. Following the time-based comparisons of [12], every run also stops after 5 minutes of wall-clock time. This cap stopped all completed FD-BB runs on HD200 after 40,000–89,000 of the 100,000 evaluations and COMO-CMA-ES on Viennet after about 14,000 evaluations. Because the cap is checked only between iterations, one FD-SD run on HD200 ended after 43 minutes. The returned approximation is the one each method returns: for PHASE-MOO the nondominated set of all evaluated points (Section 5.5), for the evolutionary algorithms and NSMA their final population of at most 100 points, for the FD methods their final front list, for COMO-CMA-ES the incumbents of its 100 kernels, and for the UHV hybrid its solution set.

Four adaptations were made to the comparators and fixed before their runs. (i) The official FD code lets iterates leave the box; the direction subproblem and the Armijo search were therefore restricted to the box, rejecting rather than clipping infeasible trials. (ii) The FD code has no default stopping rule, and being deterministic, it can cycle at a fixed point without new evaluations; such a run stops when the same accepted front recurs at an unchanged evaluation count. This rule ended three FD-BB runs on Kursawe and, for both FD variants, the DTLZ2 run of seed 6103 after 122 evaluations. (iii) The UHV hybrid is defined for two objectives only and is not applied to Viennet and DTLZ2; its boundary repair is enabled, and its set size is reduced from 100 to 30 on HD200 because of memory requirements. (iv) COMO-CMA-ES and the UHV hybrid receive the scoring reference point r_H in raw objective units. One FD-BB run on HD200 (seed 6108) stopped with an error in its box-direction subproblem and is missing. Pareto Tracer [24] was prepared but excluded before the battery, because its available Octave port would have taken most of the computing time. For each problem and indicator, PHASE-MOO is compared with each method by the two-sided Wilcoxon signed-rank test paired by seed, with Holm’s correction over the comparators of that problem and indicator.

Table 3 and Fig. 16 summarize the results. Of the 46 HV comparisons, PHASE-MOO is significantly better in 33, worse in 6, and not significantly different in 7; for IGD the counts are 33, 3, and 10. The advantage is largest on HD200, where PHASE-MOO is better than every method in both indicators. Its median HV of 99.99% compares with 99.91% for FD-BB, whose runs were stopped by the time cap at about half the allowance, and 99.35% for NSMA. Its median IGD is 2.9×10^{-4} , against 1.3×10^{-3} for FD-BB. COMO-CMA-ES does not approach the HD200 front within the allowance. On Viennet, Fonseca–Fleming, DTLZ2, and Rugged2D, PHASE-MOO is better than the three evolutionary algorithms, NSMA, and COMO-CMA-ES in both indicators. On Rugged2D it is also better than the two FD methods and the UHV hybrid.

The front-descent methods are the strongest competitors on the smooth problems. On Fonseca–Fleming they are significantly better (median HV 100.01% against 99.82%, IGD 3.7×10^{-5} against 7.3×10^{-4}), with front lists of about 18,000 points. On Viennet and DTLZ2 their medians are also higher (HV 100.00% and 100.50%), but the differences are not significant. On DTLZ2 this is due to the single run per FD variant that stopped at a fixed point after 122 evaluations (HV 41.5%); in the other nine seeds the FD methods exceed PHASE-MOO. HV ratios above 100% indicate that these fronts dominate parts of the space not covered by the finite reference samples (Section 5.4). Kursawe is the main weakness of PHASE-MOO, as anticipated by the ablation (Section 6.2). The three evolutionary algorithms and NSMA reach a median HV of 99.1–99.5% against 94.0%, which reflects the incomplete coverage of a disconnected piece of the front. The FD methods and COMO-CMA-ES are not significantly different from PHASE-MOO on Kursawe, and the UHV hybrid is worse.

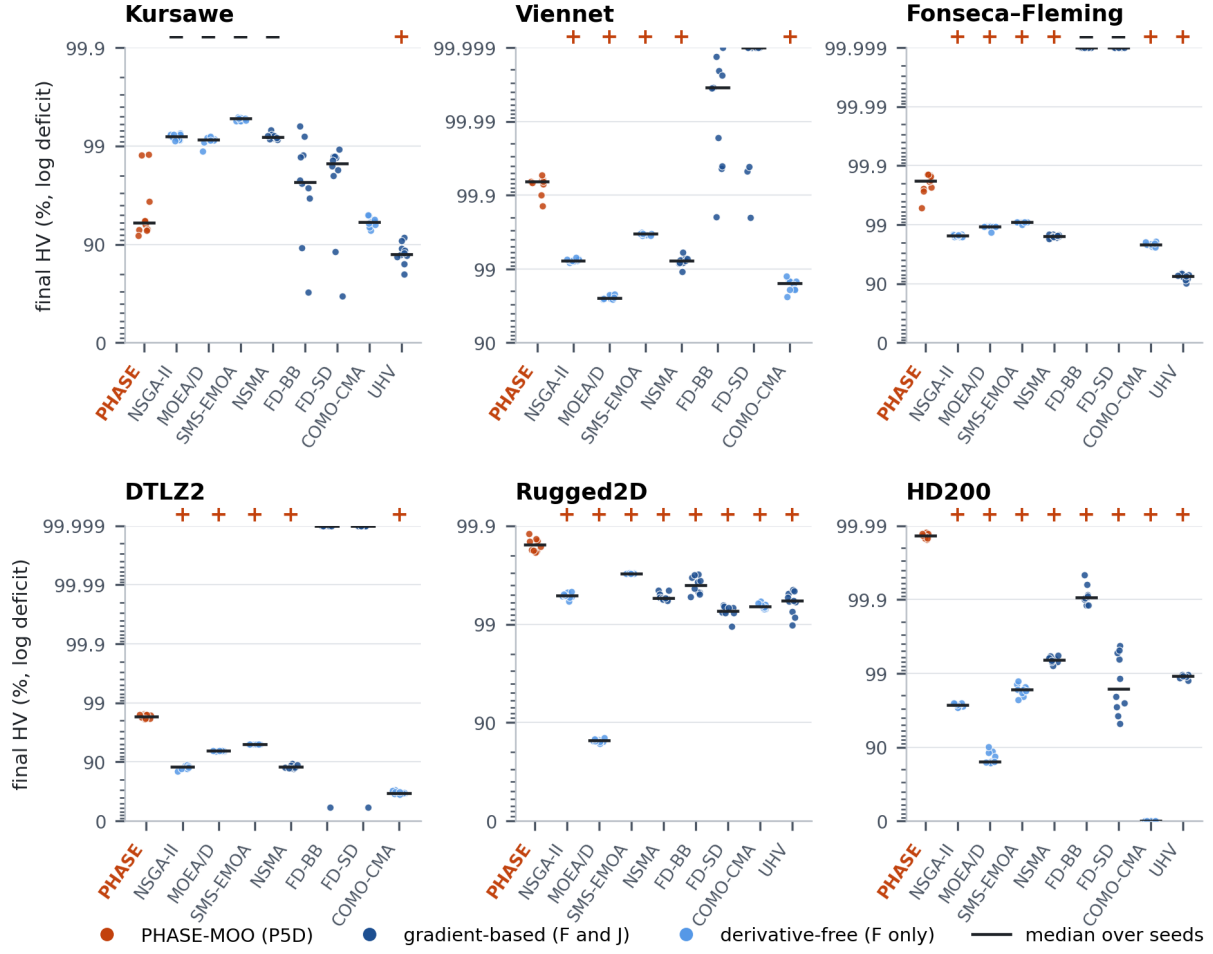


Fig. 16 Final HV of each run in the comparison (seeds 6101–6110; 20,000 joint evaluations, HD200 100,000; 5-minute cap). The vertical axis is the HV deficit $100 - \text{HV}$ on a logarithmic scale, inverted (better is up) and labeled in HV; values of 99.999% and above, including the values above 100% of the FD methods, are drawn at the upper edge. Bars are medians. Marks above the panels: + PHASE-MOO significantly better, – significantly worse (Wilcoxon, Holm, 0.05). Dark blue: gradient-based methods (objectives and Jacobian); light blue: derivative-free methods (objectives only).

Table 3 Comparison with eight methods at 20,000 joint evaluations (HD200: 100,000) and a common 5-minute cap: medians over seeds 6101–6110 of the reference HV ratio (% , higher is better) and of IGD ($\times 10^{-3}$, lower is better) of each method’s returned approximation. Superscripts: + PHASE-MOO significantly better, – significantly worse (two-sided paired Wilcoxon test, Holm correction over the comparators of each problem and indicator, level 0.05); no mark: not significant. FD-BB on HD200: nine runs. UHV is defined for two objectives only.

Problem		PHASE	NSGA-II	MOEA/D	SMS-EMOA	NSMA	FD-BB	FD-SD	COMO	UHV
Kursawe	HV	93.97	99.20 [–]	99.14 [–]	99.47 [–]	99.19 [–]	97.68	98.50	94.05	87.42 ⁺
	IGD	27.2	4.08	4.98	3.17 [–]	4.19	16.0	10.5	22.2	47.8 ⁺
Viennet	HV	99.94	99.24 ⁺	97.52 ⁺	99.67 ⁺	99.23 ⁺	100.00	100.00	98.44 ⁺	–
	IGD	2.46	8.29 ⁺	99.1 ⁺	13.4 ⁺	8.12 ⁺	1.01	0.627	21.6 ⁺	–
Fonseca–Fleming	HV	99.82	98.47 ⁺	98.93 ⁺	99.09 ⁺	98.43 ⁺	100.01 [–]	100.01 [–]	97.84 ⁺	92.63 ⁺
	IGD	0.727	5.20 ⁺	3.76 ⁺	3.84 ⁺	5.38 ⁺	0.0370 [–]	0.0373 [–]	7.72 ⁺	20.6 ⁺
DTLZ2	HV	98.31	87.84 ⁺	93.49 ⁺	94.98 ⁺	87.77 ⁺	100.50	100.49	66.35 ⁺	–
	IGD	22.3	67.5 ⁺	51.5 ⁺	72.4 ⁺	67.6 ⁺	5.42	5.45	172.1 ⁺	–
Rugged2D	HV	99.85	99.49 ⁺	84.82 ⁺	99.70 ⁺	99.46 ⁺	99.60 ⁺	99.27 ⁺	99.34 ⁺	99.42 ⁺
	IGD	1.73	5.29 ⁺	92.6 ⁺	4.97 ⁺	5.38 ⁺	4.40 ⁺	6.38 ⁺	8.35 ⁺	7.92 ⁺
HD200	HV	99.99	97.31 ⁺	84.38 ⁺	98.35 ⁺	99.35 ⁺	99.91 ⁺	98.38 ⁺	0.00 ⁺	98.92 ⁺
	IGD	0.286	15.3 ⁺	103.4 ⁺	10.0 ⁺	5.72 ⁺	1.26 ⁺	29.9 ⁺	29,379 ⁺	15.6 ⁺

The approximations returned by the methods differ in size, from at most 100 points for the population-based methods to thousands for PHASE-MOO and the FD methods. Two further views therefore qualify the results. First, the secondary analysis applies the indicators to the nondominated set of all feasible points evaluated by each method. This view leaves PHASE-MOO unchanged, because it already returns this set, but it raises the evolutionary algorithms substantially: on Kursawe, for example, to 99.8–99.9%, and on DTLZ2 to 98.9–99.8%. On this basis PHASE-MOO remains the best method on HD200, but it is worse than SMS-EMOA and NSGA-II on Kursawe, Viennet, DTLZ2, and Rugged2D, and worse than NSMA on three of these problems. The evolutionary algorithms thus evaluate good points that their final populations do not retain, whereas retention of all evaluated nondominated points is part of PHASE-MOO. Second, when every returned approximation is reduced to at most 100 points by greedy hypervolume subset selection, PHASE-MOO remains significantly better in HV than NSGA-II, MOEA/D, NSMA, COMO-CMA-ES, and the UHV hybrid on every problem except Kursawe, while SMS-EMOA becomes better on DTLZ2 and Rugged2D by at most 0.17 percentage points (results not shown). The advantage in Table 3 is therefore not merely an effect of the number of returned points.

The median wall-clock time per run is 8–71 s for PHASE-MOO, 3–21 s for NSGA-II, 50–276 s for SMS-EMOA, and up to the 300 s cap for the FD methods on HD200. These times depend on the implementations and include the Jacobian computed by the shared oracle also for the derivative-free methods, so they are reported for completeness rather than as a comparison of efficiency. Overall, under a common allowance of evaluated points, PHASE-MOO is competitive with or better than established evolutionary, memetic, front-descent, and hypervolume-based methods on five of the six problems. It is clearly superior on the 200-variable problem, and its limitation is the coverage of disconnected fronts at small budgets.

7 Conclusions

PHASE-MOO treats the local search geometry as a population-level variable rather than as a fixed preprocessing choice. Each branch computes a safeguarded common-descent step in its own positive definite metric, branches with different metrics compete for evaluations through an explore–select–commit cycle, and the method returns every nondominated point it evaluates. The convergence result covers the local kernel under interior conditions for uniformly bounded metric sequences, so the geometric variation of the population does not compromise the local guarantee; the population mechanisms themselves were assessed empirically.

The experiments answer the question that motivated the design. The multi-geometry population is much better than multistart identity-metric descent and than the same population restricted to the identity metric, on all six benchmarks. Adaptive allocation, however, is practically equivalent to uniform allocation once mathematically redundant geometries are removed, and the selection shares stay close to uniform. The portfolio therefore helps through the complementarity of its geometries, not through learning which geometry is best. The high-budget runs show a second mechanism. Only a small fraction of the returned nondominated points are final branch endpoints, and 21–56% of them descend from seeds that were later pruned. Retaining all evaluated nondominated points therefore lets pruning redirect effort without discarding coverage.

At 20,000 joint evaluations (100,000 on the 200-variable problem), median reference HV ratios are between 98.30% and 99.99% on five problems, and runs with one million evaluations reach 100.00–100.04% on all six. Against eight evolutionary, memetic, front-descent, and hypervolume-based methods under the same allowances of evaluated points, PHASE-MOO is significantly better in 33 of 46 hypervolume comparisons. Its advantage is largest on the 200-variable problem. It trails the evolutionary algorithms on Kursawe and the front-descent methods on Fonseca–Fleming.

The study also yields practical guidance. Of the seven calibrated parameters, only the initial population and the direction radius had a substantial effect; the step ceiling, the retention fraction, and the allocation rule can be left at default values. The initial population should be matched to the budget per seed: the 200-variable problem preferred 40 seeds for 100,000 evaluations, whereas DTLZ2 preferred 640 seeds for 20,000. Uniform allocation over a de-duplicated portfolio is a simpler default that is as good as the adaptive rule.

The limitations define future work. PHASE-MOO requires objective Jacobians and was evaluated on box-constrained problems with two or three objectives. At small budgets, coverage of disconnected or extreme parts of the front depends on whether some branch reaches their basin. On Kursawe, 4 of 20 runs miss part of a disconnected piece (worst HV 89.9%), and on Fonseca–Fleming one of 20 runs misses an extreme portion (HV 93.8%); with one million evaluations, all parts are covered. Defect-triggered front extension or repair is therefore a natural complement, left to separate work. The setups were calibrated per

problem, so transfer across problem classes and automatic configuration deserve study. In the comparison, gradient-based methods received the Jacobian with each charged evaluation, comparators were not tuned, and the returned approximations differ in size. Its conclusions therefore concern returned approximations under a common allowance of evaluated points, not equal information or computing time. Extensions to general constraints and to many-objective problems complete the agenda.

Statements and Declarations

Funding

The authors received no specific funding for this research.

Competing interests

The authors declare no financial or non-financial competing interests relevant to this work.

Data and code availability

The data and code supporting the reported results are publicly available in the PHASE-MOO reproducibility archive on Zenodo [48], <https://doi.org/10.5281/zenodo.23123548>. The archive contains per-run outputs, initial and final populations, frozen reference fronts, parameter manifests, source snapshots, plotting scripts, and file-integrity checks. PHASE-MOO results use the nondominated set of all evaluated points (Section 3.4); comparator results use the conventions of Section 6.4. Original code is released under the MIT license; generated data, figures, and documentation are released under CC BY 4.0, with third-party exclusions specified in the archive.

Author approval and accountability

Both authors approved the manuscript and its submission and accept responsibility for the accuracy and integrity of the work.

Ethics approval and consent

Not applicable: this study uses mathematical benchmark problems and does not involve human participants, personal research data, or animal experiments.

References

- [1] K. Miettinen, *Nonlinear Multiobjective Optimization*, Kluwer Academic Publishers, Boston, 1998. <https://doi.org/10.1007/978-1-4615-5563-6>.
- [2] M.T.M. Emmerich, A.H. Deutz, A tutorial on multiobjective optimization: fundamentals and evolutionary methods, *Nat. Comput.* 17 (2018) 585–609. <https://doi.org/10.1007/s11047-018-9685-y>.
- [3] J. Fliege, B.F. Svaiter, Steepest descent methods for multicriteria optimization, *Math. Methods Oper. Res.* 51 (2000) 479–494. <https://doi.org/10.1007/s001860000043>.
- [4] L.M. Graña Drummond, B.F. Svaiter, A steepest descent method for vector optimization, *J. Comput. Appl. Math.* 175 (2005) 395–414. <https://doi.org/10.1016/j.cam.2004.06.018>.
- [5] J.-A. Désidéri, Multiple-gradient descent algorithm (MGDA) for multiobjective optimization, *C. R. Math.* 350 (2012) 313–318. <https://doi.org/10.1016/j.crma.2012.03.014>.
- [6] L.M. Graña Drummond, A.N. Iusem, A projected gradient method for vector optimization problems, *Comput. Optim. Appl.* 28 (2004) 5–29. <https://doi.org/10.1023/B:COAP.0000018877.86161.8B>.
- [7] J. Fliege, L.M. Graña Drummond, B.F. Svaiter, Newton’s method for multiobjective optimization, *SIAM J. Optim.* 20 (2009) 602–626. <https://doi.org/10.1137/08071692X>.
- [8] L.R. Lucambio Pérez, L.F. Prudente, Nonlinear conjugate gradient methods for vector optimization, *SIAM J. Optim.* 28 (2018) 2690–2720. <https://doi.org/10.1137/17M1126588>.

- [9] J. Fliege, A.I.F. Vaz, A method for constrained multiobjective optimization based on SQP techniques, *SIAM J. Optim.* 26 (2016) 2091–2119. <https://doi.org/10.1137/15M1016424>.
- [10] M. Lapucci, P. Mansueto, Improved front steepest descent for multi-objective optimization, *Oper. Res. Lett.* 51 (2023) 242–247. <https://doi.org/10.1016/j.orl.2023.03.001>.
- [11] A. Mohammadi, A.L. Custódio, A trust-region approach for computing Pareto fronts in multiobjective optimization, *Comput. Optim. Appl.* 87 (2024) 149–179. <https://doi.org/10.1007/s10589-023-00510-2>.
- [12] M. Lapucci, P. Mansueto, D. Pucci, Effective front-descent algorithms with convergence guarantees, *SIAM J. Optim.* 36 (2026) 597–625. <https://doi.org/10.1137/25M1726856>.
- [13] I. Das, J.E. Dennis, Normal-boundary intersection: a new method for generating the Pareto surface in nonlinear multicriteria optimization, *SIAM J. Optim.* 8 (1998) 631–657. <https://doi.org/10.1137/S1052623496307510>.
- [14] J. Chen, G.-X. Li, X.-M. Yang, Variable metric method for unconstrained multiobjective optimization problems, *J. Oper. Res. Soc. China* 11 (2023) 409–438. <https://doi.org/10.1007/s40305-022-00447-z>.
- [15] W. Chen, X. Yang, Y. Zhao, Memory gradient method for multiobjective optimization, *Appl. Math. Comput.* 443 (2023) 127791. <https://doi.org/10.1016/j.amc.2022.127791>.
- [16] J. Chen, L. Tang, X. Yang, A Barzilai–Borwein descent method for multiobjective optimization problems, *Eur. J. Oper. Res.* 311 (2023) 196–209. <https://doi.org/10.1016/j.ejor.2023.04.022>.
- [17] E. Zitzler, L. Thiele, Multiobjective evolutionary algorithms: a comparative case study and the strength Pareto approach, *IEEE Trans. Evol. Comput.* 3 (1999) 257–271. <https://doi.org/10.1109/4235.797969>.
- [18] K. Deb, A. Pratap, S. Agarwal, T. Meyarivan, A fast and elitist multiobjective genetic algorithm: NSGA-II, *IEEE Trans. Evol. Comput.* 6 (2002) 182–197. <https://doi.org/10.1109/4235.996017>.
- [19] Q. Zhang, H. Li, MOEA/D: A multiobjective evolutionary algorithm based on decomposition, *IEEE Trans. Evol. Comput.* 11 (2007) 712–731. <https://doi.org/10.1109/TEVC.2007.892759>.
- [20] P.A.N. Bosman, On gradients and hybrid evolutionary algorithms for real-valued multiobjective optimization, *IEEE Trans. Evol. Comput.* 16 (2012) 51–69. <https://doi.org/10.1109/TEVC.2010.2051445>.
- [21] A. Lara, G. Sanchez, C.A. Coello Coello, O. Schütze, HCS: a new local search strategy for memetic multiobjective evolutionary algorithms, *IEEE Trans. Evol. Comput.* 14 (2010) 112–132. <https://doi.org/10.1109/TEVC.2009.2024143>.
- [22] K. Sindhya, K. Miettinen, K. Deb, A hybrid framework for evolutionary multi-objective optimization, *IEEE Trans. Evol. Comput.* 17 (2013) 495–511. <https://doi.org/10.1109/TEVC.2012.2204403>.
- [23] M. Lapucci, P. Mansueto, F. Schoen, A memetic procedure for global multi-objective optimization, *Math. Program. Comput.* 15 (2023) 227–267. <https://doi.org/10.1007/s12532-022-00231-3>.
- [24] A. Martín, O. Schütze, Pareto Tracer: a predictor–corrector method for multi-objective optimization problems, *Eng. Optim.* 50 (2018) 516–536. <https://doi.org/10.1080/0305215X.2017.1327579>.
- [25] C. Touré, N. Hansen, A. Auger, D. Brockhoff, Uncrowded hypervolume improvement: COMO-CMA-ES and the Sofomore framework, in: *Proc. Genetic and Evolutionary Computation Conference (GECCO 2019)*, ACM, 2019, pp. 638–646. <https://doi.org/10.1145/3321707.3321852>.
- [26] D.M.F. Ha, T.M. Deist, P.A.N. Bosman, Hybridizing hypervolume-based evolutionary algorithms and gradient descent by dynamic resource allocation, in: G. Rudolph et al. (Eds.), *Parallel Problem Solving from Nature – PPSN XVII*, LNCS 13399, Springer, 2022, pp. 179–192. https://doi.org/10.1007/978-3-031-14721-0_13.

- [27] C. Igel, N. Hansen, S. Roth, Covariance matrix adaptation for multi-objective optimization, *Evol. Comput.* 15 (2007) 1–28. <https://doi.org/10.1162/evco.2007.15.1.1>.
- [28] R.S. Barr, B.L. Golden, J.P. Kelly, M.G.C. Resende, W.R. Stewart, Designing and reporting on computational experiments with heuristic methods, *J. Heuristics* 1 (1995) 9–32. <https://doi.org/10.1007/BF02430363>.
- [29] J.N. Hooker, Testing heuristics: we have it all wrong, *J. Heuristics* 1 (1995) 33–42. <https://doi.org/10.1007/BF02430364>.
- [30] S. Huband, P. Hingston, L. Barone, L. While, A review of multiobjective test problems and a scalable test problem toolkit, *IEEE Trans. Evol. Comput.* 10 (2006) 477–506. <https://doi.org/10.1109/TEVC.2005.861417>.
- [31] C.M. Fonseca, P.J. Fleming, An overview of evolutionary algorithms in multiobjective optimization, *Evol. Comput.* 3 (1995) 1–16. <https://doi.org/10.1162/evco.1995.3.1.1>.
- [32] K. Deb, L. Thiele, M. Laumanns, E. Zitzler, Scalable multi-objective optimization test problems, in: *Proc. IEEE Congress on Evolutionary Computation*, 2002, pp. 825–830. <https://doi.org/10.1109/CEC.2002.1007032>.
- [33] P.-A. Absil, R. Mahony, R. Sepulchre, *Optimization Algorithms on Matrix Manifolds*, Princeton University Press, Princeton, 2008.
- [34] J. Duchi, E. Hazan, Y. Singer, Adaptive subgradient methods for online learning and stochastic optimization, *J. Mach. Learn. Res.* 12 (2011) 2121–2159. <https://jmlr.org/papers/v12/duchi11a.html>.
- [35] L. Armijo, Minimization of functions having Lipschitz continuous first partial derivatives, *Pacific J. Math.* 16 (1966) 1–3. <https://doi.org/10.2140/pjm.1966.16.1>.
- [36] E. Zitzler, M. Laumanns, L. Thiele, SPEA2: Improving the strength Pareto evolutionary algorithm, *TIK Report 103*, ETH Zurich, 2001. <https://hdl.handle.net/20.500.11850/145755>.
- [37] A.E. Eiben, R. Hinterding, Z. Michalewicz, Parameter control in evolutionary algorithms, *IEEE Trans. Evol. Comput.* 3 (1999) 124–141. <https://doi.org/10.1109/4235.771166>.
- [38] K. Li, Á. Fialho, S. Kwong, Q. Zhang, Adaptive operator selection with bandits for a multiobjective evolutionary algorithm based on decomposition, *IEEE Trans. Evol. Comput.* 18 (2014) 114–130. <https://doi.org/10.1109/TEVC.2013.2239648>.
- [39] M. López-Ibáñez, J. Dubois-Lacoste, L. Pérez Cáceres, M. Birattari, T. Stützle, The irace package: iterated racing for automatic algorithm configuration, *Oper. Res. Perspect.* 3 (2016) 43–58. <https://doi.org/10.1016/j.orp.2016.09.002>.
- [40] J. Blank, K. Deb, pymoo: Multi-Objective Optimization in Python, *IEEE Access* 8 (2020) 89497–89509. <https://doi.org/10.1109/ACCESS.2020.2990567>.
- [41] F. Kursawe, A variant of evolution strategies for vector optimization, in: H.-P. Schwefel, R. Männer (Eds.), *Parallel Problem Solving from Nature*, LNCS 496, Springer, 1991, pp. 193–197. <https://doi.org/10.1007/BFb0029752>.
- [42] R. Viennet, C. Fonteix, I. Marc, Multicriteria optimization using a genetic algorithm for determining a Pareto set, *Int. J. Syst. Sci.* 27 (1996) 255–260. <https://doi.org/10.1080/00207729608929211>.
- [43] E. Zitzler, L. Thiele, M. Laumanns, C.M. Fonseca, V.G. da Fonseca, Performance assessment of multiobjective optimizers: an analysis and review, *IEEE Trans. Evol. Comput.* 7 (2003) 117–132. <https://doi.org/10.1109/TEVC.2003.810758>.
- [44] C.M. Fonseca, L. Paquete, M. López-Ibáñez, An improved dimension-sweep algorithm for the hypervolume indicator, in: *Proc. IEEE Congress on Evolutionary Computation*, 2006, pp. 1157–1163.

<https://doi.org/10.1109/CEC.2006.1688440>.

- [45] N. Beume, C.M. Fonseca, M. López-Ibáñez, L. Paquete, J. Vahrenhold, On the complexity of computing the hypervolume indicator, *IEEE Trans. Evol. Comput.* 13 (2009) 1075–1082. <https://doi.org/10.1109/TEVC.2009.2015575>.
- [46] K. Shang, H. Ishibuchi, L. He, L.M. Pang, A survey on the hypervolume indicator in evolutionary multiobjective optimization, *IEEE Trans. Evol. Comput.* 25 (2021) 1–20. <https://doi.org/10.1109/TEVC.2020.3013290>.
- [47] N. Beume, B. Naujoks, M. Emmerich, SMS-EMOA: Multiobjective selection based on dominated hypervolume, *Eur. J. Oper. Res.* 181 (2007) 1653–1669. <https://doi.org/10.1016/j.ejor.2006.08.008>.
- [48] M.P. Rodrigues, R.M.J. Lemos, PHASE-MOO: reproducibility data and code for calibrated geometric population optimization, Zenodo (2026). <https://doi.org/10.5281/zenodo.23123548>.