mathematics

MDPI

Article

Objective Selection in Many-Objective Optimization via Conditioning-Based Analysis of Objective Redundancy

Tomás Marques, António J. Pontes and António Gaspar-Cunha *

IPC—Institute for Polymers and Composites, University of Minho, 4800-058 Guimarães, Portugal; b13851@uminho.pt (T.M.); pontes@dep.uminho.pt (A.J.P.)
* Correspondence: agc@dep.uminho.pt

Abstract

Evolutionary many-objective algorithms lose selection pressure and become harder to steer as the number of objectives grows, so objective selection is often applied before the search. Most selection methods measure dependence directly on the objective values and therefore cannot distinguish dependence induced by shared decision variables from redundancy between the objectives themselves. This paper proposes a conditioning-based selection criterion in which each objective is regressed on the decision variables, and redundancy is measured on the residuals; in the linear case, the measure is the partial correlation given the design space. A greedy score combining relevance, unexplained variability, and residual redundancy, with standardized components, builds the subset, and an admissibility gate reports when the conditioning model is too weak or the residual variance too small for the criterion to apply. On DTLZ and WFG benchmarks, the complete procedure admits five of twelve configurations and declares the other seven non-applicable; in four of the five admitted configurations, the conditioned subset differs from those of PCA, mRMR-like selection, and clustering. An active NSGA-III validation with paired Wilcoxon tests shows no uniform performance advantage. The admitted subsets are invariant to the cross-validation seed, and the re-seeded nonlinear selector recovers its own modal subset in 93.1% of runs. A controlled example with known redundant and conflicting objectives shows that absolute residual correlation can discard a pure residual conflict, which a sign-aware variant avoids. In an injection-molding case with 35 indicators, an ablation shows that the unexplained-variability term is decisive: removing it changes two of the three selected objectives. The contribution is therefore a selection criterion conditional on an adequacy diagnostic, not a general-purpose objective-selection method, and the analysis states when it should be used instead of marginal-dependence baselines.

Keywords: many-objective optimization; objective selection; objective reduction; evolutionary algorithms; NSGA-III; residual redundancy; partial correlation; injection molding

MSC: 90C29; 90C59; 68W50; 62H20; 62J05

Check for updates

Academic Editor: Sara Pérez-Carabaza

Received: 13 October 2026
Revised: 31 October 2026
Accepted: 3 November 2026
Published: 5 November 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

Many real-world engineering problems are best described as multi-objective optimization problems because they involve many performance indicators that reflect different aspects of system behavior. For example, in injection molding, objectives might include cycle time, temperature uniformity, warpage, shrinkage, residual stress, and thermal efficiency. As the number of objectives grows, the problem becomes a many-objective problem, in

which traditional evolutionary optimization algorithms become less effective. Using Pareto dominance in these cases reduces selection pressure, makes it harder to maintain diversity, and complicates the interpretation and use of the resulting solution sets [1]. This also increases the difficulty of selecting meaningful solutions for engineering decision-making. These challenges have led to the development of specialized many-objective algorithms and methods to reduce the number of objectives before optimization [1–3].

Objective selection or reduction tackles this problem by finding a smaller set of key objectives before optimization. The goal is to keep the most important trade-offs, reduce redundancy, facilitate the use of multi-objective optimization algorithms, and make results easier to interpret. Common methods include correlation-based screening, mutual information, minimum redundancy-maximum relevance (mRMR), principal component analysis (PCA), clustering, and other objective-reduction frameworks [3–6]. These approaches often allow for significant simplification, but it can still be hard to decide which objectives are truly redundant and which are essential to preserving the physical meaning and decision relevance of the problem [3–7].

Objective selection matters for evolutionary many-objective search in a concrete and quantifiable way. Reference-point and decomposition-based algorithms such as NSGA-III distribute their search effort over a structured set of reference directions whose number grows combinatorially with the number of objectives: a simplex lattice with H divisions on m objectives contains $C(H + m - 1, m - 1)$ directions, so that every objective removed reduces either the population required to cover the front or the density with which it is covered [1]. Objective selection is equally a decision-making problem, because the reduced objective vector determines which trade-offs the evolutionary search exposes and, therefore, what the decision maker is subsequently able to trade off. A selection rule intended for use ahead of a nature-inspired heuristic must accordingly satisfy two requirements at once: it must reduce the dimension in which the heuristic operates, and it must return objectives that remain physically interpretable and that can be overridden by decision-maker priorities, as is done for cycle time and total energy consumption in Section 4.5.

However, objective dependence is not the same as objective redundancy. Most current methods use statistical measures such as correlation or mutual information to identify similar objectives, but these measures do not indicate where the dependence arises. Two objectives might seem strongly related not because they represent the same trade-off, but because they are both affected by the same decision variables. In this case, the link is due to shared variables rather than a direct relationship between the objectives. As a result, dependence-based selection might remove objectives that are only jointly influenced by design factors, while missing objectives that have weaker dependence but still capture unique aspects of system behavior [5,8].

This limitation suggests the need for a better way to define redundancy, one that considers how objective relationships arise from the design space. If the association between two objectives disappears once the decision variables are accounted for, the objectives are not redundant in the residual sense: their similarity is explained by the shared inputs, each may still respond to the design in a way the other does not, and their marginal correlation is no reason to discard either. If dependence remains after accounting for the decision variables, the two objectives share variation that the design space does not explain, and one of them may be redundant given the other. Objective relationships should therefore be assessed after conditioning on the decision variables, so that effects of shared variables can be separated from other dependencies [8–10]. In this work, conditioning is used as a statistical tool for that separation, without attempting causal discovery.

Motivated by this observation, this paper proposes a conditioning-based objective selection framework for many-objective optimization, in which each objective is regressed

on the decision variables and redundancy is evaluated on the residuals. Each objective is modeled as a function of the decision variables, and redundancy is evaluated by assessing the residual dependencies remaining after removing their influence. The method does not aim to infer directed causal links between objectives; rather, it uses conditioning to distinguish shared-variable-induced dependence from residual relationships between objectives. Based on this concept, a sequential selection procedure is developed to identify compact, informative, non-redundant, and physically interpretable subsets of objectives.

The proposed approach is evaluated using synthetic benchmark problems and a real engineering application. The benchmark analysis considers some DTLZ and WFG problems, two scalable benchmark suites widely used to assess many-objective optimization methods under different geometric and dependency characteristics [11,12]. The methodology is then applied to the design of conformal cooling channels in injection molding, a many-objective engineering problem with numerous interrelated performance measures and strong practical relevance.

The main contributions of this work are as follows:

1. an objective-selection criterion in which redundancy is defined as the residual dependence remaining after each objective is regressed on the decision variables, together with an admissibility gate (conditioner adequacy and a residual-variance floor) that determines when this definition is applicable; the estimator itself is classical partial-correlation analysis and is not claimed as new, whereas its use as a redundancy criterion for objective selection, and the conditions under which it is valid, are;
2. a sequential selection procedure that extends the standard relevance-redundancy (mRMR) framework with an explicit unexplained-variability term and a scale-standardized score, accompanied by an ablation showing that the added term is the most influential of the three;
3. a comparative analysis of raw and conditioned objective dependencies on DTLZ and WFG benchmark problems;
4. an active optimization validation comparing conditioning-based selections with PCA, mRMR and clustering baselines, including a stochastic re-seeding analysis of the nonlinear conditioning selector;
5. a real injection molding case study showing how engineering knowledge, physical objective categories, and regime effects can be integrated into objective selection.

The remainder of the paper is organized as follows. Section 2 reviews related work on objective selection, dependence analysis, and causally motivated perspectives on redundancy in many-objective optimization. Section 3 presents the proposed methodology. Section 4 presents the results for both benchmark problems and the injection molding case study. Section 5 concludes the paper.

2. Background and Related Work

Objective selection and reduction have been widely studied as a means to address the challenges associated with many-objective optimization. As the number of objectives increases, the difficulty of maintaining selection pressure and diversity in evolutionary algorithms grows significantly, and the resulting Pareto sets become increasingly complex and difficult to interpret [1,2]. Consequently, reducing the number of objectives while preserving the problem's essential characteristics has become an important research direction.

Early approaches to objective reduction focused on identifying redundancy through statistical dependence measures. Correlation-based methods aim to remove objectives that are strongly linearly dependent, whereas more advanced techniques employ nonlinear measures, such as mutual information, to capture broader relationships [5,13]. More recent

objective-reduction approaches have also considered local or population-based structure to avoid over-reduction and preserve relevant objective-space features [14].

To give a structured overview of the main objective selection strategies, Table 1 summarizes the most commonly used approaches and highlights their underlying principles, advantages, and limitations. In particular, it emphasizes the reliance of existing methods on statistical dependence measures that do not explicitly account for whether objective relationships are induced by shared decision variables. This limitation motivates the conditioning-based approach proposed in this work, which is discussed separately after the literature summary.

Table 1. Summary of objective selection methods.

Method Category	Representative Methods	Principle	Advantages	Limitations
Correlation-based	Pearson, Spearman	Linear (Pearson) or monotonic (Spearman) dependence between objectives	Simple, fast	Pearson captures only linear and Spearman only monotonic association; neither captures non-monotonic dependence
Information-theoretic	Mutual information, mRMR [5]	Maximize relevance, minimize redundancy	Captures nonlinear dependence	Does not distinguish dependence induced by shared decision variables
Dimensionality reduction	PCA [4]	Transform objectives into lower-dimensional space	Effective compression	Loss of physical interpretability
Clustering-based	Correlation-based and hierarchical clustering [15,16]	Group similar objectives and select representatives	Intuitive grouping	Depends on distance metric
Graph-based	Laplacian Score, spectral methods [17]	Preserve intrinsic data structure	Captures manifold structure	Computational complexity

The proposed method differs from the approaches summarized in Table 1 by evaluating redundancy after conditioning each objective on the decision variables. Thus, it does not treat marginal dependence alone as evidence of redundancy, but instead focuses on the residual dependence that remains after shared-variable effects are removed.

Another class of approaches relies on dimensionality reduction techniques, such as principal component analysis (PCA), which transform the objective space into a lower-dimensional representation [4]. Despite being effective in capturing variance, these methods typically produce latent variables that lack direct physical interpretation, limiting their usefulness in engineering decision-making contexts. To address this limitation, clustering-based methods have been proposed to group similar objectives and select representative ones from each cluster [15,16]. Graph-based techniques, including Laplacian Score and spectral methods [17], have also been used to preserve the intrinsic structure of the objective space while reducing dimensionality.

A related machine-learning-based selection strategy has also been proposed, in which regression models are used to iteratively score and select the objectives that best preserve the explained variance of the remaining objective set [7]. More recently, ensemble-based objective selection frameworks have been introduced to improve robustness and incorporate multiple selection criteria. These approaches combine methods such as clustering, mutual information, mRMR, and graph-based scores into a consensus mechanism that balances redundancy reduction and relevance [18]. Such frameworks are particularly useful in engineering applications, where objective selection must account not only for statistical properties but also for physical meaning and decision-maker preferences.

Despite these advances, a common limitation of existing methods is their reliance on statistical dependence as a proxy for redundancy. Statistical dependence does not distinguish direct relationships from dependencies induced by common factors, and in many engineering systems several objectives are functions of the same decision variables, so that

apparent correlations need not reflect any interaction between the objectives. Removing objectives on marginal dependence alone may therefore eliminate objectives that are physically meaningful yet jointly driven by the same design factors. The statistical remedy is classical: partial correlation and conditional-independence analysis evaluate relationships after accounting for covariates [19], and the distinction between correlation and common-cause structure is standard in the causal-inference literature [8–10]. Related ideas appear in conditional mutual-information feature selection, where redundancy is evaluated after accounting for additional variables [20].

In evolutionary multi-objective optimization, recent studies have analyzed the structure of objective dependencies and its implications for optimization performance [21], but they still rely on marginal measures. Causal reasoning has also entered multi-objective optimization through causal Bayesian optimization, where known causal graphs guide intervention decisions [22]; that line of work seeks Pareto-optimal interventions rather than a reduced objective set and is therefore complementary to the present one.

The present work handles this gap by introducing a conditioning-based perspective on objective redundancy. Instead of evaluating dependence directly on the objective values, the proposed approach conditions each objective on the decision variables and analyzes the residual dependencies. This enables the identification of objectives whose dependence is mainly induced by shared decision variables, as opposed to those that retain independent structural information. By integrating this conditioning-based perspective into the objective selection process, the proposed method extends existing dependence-based approaches and provides a more principled basis for reducing the number of objectives in many-objective optimization problems.

Figure 1 illustrates the situation. The decision variables act as common causes of several objectives and induce apparent dependencies between them; after conditioning on the decision variables, only the residual relationships remain, and these form the basis of the selection criterion proposed in this work.

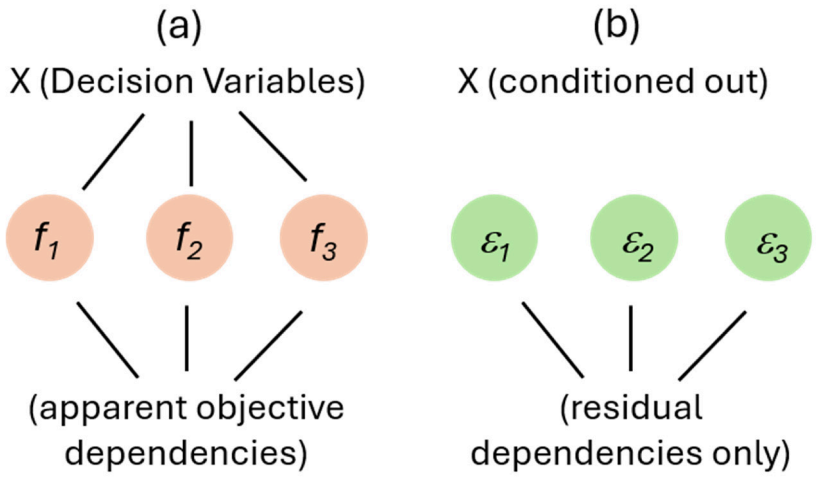


Figure 1. Conceptual illustration of objective dependencies in many-objective optimization: (**a**) before conditioning; (**b**) after conditioning on the decision variables.

3. Conditioning-Based Objective Selection Methodology

3.1. Overview and Notation

Consider a multi-objective optimization problem defined by a set of decision variables $X \in \mathbb{R}^{N \times d}$ and a set of objective functions $F = \{f_1, f_2, \ldots, f_m\}$, where each objective $f_i \in \mathbb{R}^N$ is evaluated over a set of N sampled solutions. In many-objective optimization problems, objectives are often strongly dependent due to shared physical mechanisms or because they

are functions of the same decision variables. This dependence complicates both optimization and interpretation, motivating the need for objective selection or reduction.

3.2. Conditioning of the Objectives on the Decision Variables

To address this issue, we adopt a conditioning-based perspective and evaluate objective redundancy after conditioning on the decision variables. Each objective f_i is modeled as a function of the decision variables:

$$f_i = g_i(X) + \varepsilon_i \tag{1}$$

where $g_i(X)$ is a regression model and ε_i represents the residual component not explained by the decision variables. In this work, linear conditioning was performed using ordinary least-squares regression. Nonlinear conditioning was evaluated using an ensemble tree model with limited depth and minimum leaf size to reduce overfitting. Model quality was assessed using five-fold cross-validation, and both training and validation values were recorded. The nonlinear conditioner is deliberately regularized: tree depth is capped, and a minimum leaf size is enforced, so the ensemble cannot manufacture residual structure from noise, which, in a redundancy test, would falsely retain a genuinely redundant objective. The choice is conservative by design, and the procedure is set to under-detect rather than invent redundancy. Because Extra Trees draws split points at random, the resulting subset could, in principle, vary between runs; Section 4.4 evaluates the stochastic stability of the nonlinear selector under estimator and cross-validation re-seeding, showing that the selected subsets are not artifacts of a particular random state. This formulation is consistent with the logic of structural modeling, in which the residual term captures variation not attributable to the observed decision variables [8,9]. In the present work, the purpose is not to infer directed causal links between objectives, but to distinguish dependence induced by shared decision variables from residual relationships among objectives. This regularized configuration (100 trees, a maximum depth of 8 and at least 20 samples per leaf) is the one used for selection and for the adequacy gate; the more flexible model used only for the diagnostics of Section 4.4.2 is described in Section 4.4.2.

The choice between the linear and the nonlinear variant is not left to the analyst. Both are fitted, and each is first checked against the adequacy gate of Section 3.4 (mean out-of-fold coefficient of determination of at least 0.50, computed on the same five folds). Among the adequate variants, the one used for selection is the one with the lower mean absolute residual dependence. Predictive fit alone is not a sufficient criterion: a more flexible conditioner can absorb the separable part of each objective and leave a shared residual, so that a better fit inflates rather than isolates the redundancy signal. Predictive fit therefore decides adequacy, and residual dependence decides between adequate variants. In the injection-molding case, the linear variant has both the higher out-of-fold value (0.922 against 0.884) and the smaller generalization gap (0.004 against 0.014), and it also has the lower residual dependence over the 35 candidate indicators (0.219 against 0.240), so it is retained under either rule. With the selector configuration (Section 4.4.4), the nonlinear variant is clearly superior out-of-fold on DTLZ2 and on DTLZ3 with three and five objectives, but its residual dependence is higher than that of the linear variant in all five admitted configurations, and in four of them it even exceeds the raw dependence (for DTLZ2 with three objectives, 0.615 against a raw value of 0.371); the linear variant is therefore used in all admitted benchmark configurations. On DTLZ7, the linear variant is superior or equivalent (out-of-fold R^2 of 0.990–1.000 against 0.926–0.9996), but most DTLZ7 objectives then fall below the residual-variance floor, which is the situation handled by Section 3.4. On WFG9, and on DTLZ3 with ten objectives, neither variant reaches the adequacy threshold of Section 3.4.

The residuals are computed as:

$$\varepsilon_i = f_i - \hat{g}_i(X) \tag{2}$$

and the proportion of variance explained by the decision variables is quantified using the coefficient of determination:

$$R_i^2 = 1 - \frac{\sum_{n=1}^{N}(f_i^{(n)} - \hat{g}_i(X^{(n)}))^2}{\sum_{n=1}^{N}(f_i^{(n)} - \bar{f}_i)^2} \tag{3}$$

The complementary quantity

$$U_i = 1 - R_i^2 \tag{4}$$

represents the fraction of variability that remains unexplained after conditioning. A low U_i does not imply that the objective is unimportant for optimization or decision-making. It only indicates that its variability is strongly explained by the decision variables. Such objectives may still be retained when they represent important engineering or decision-maker priorities. In the implementation, $R^2{}_i$ is computed from the out-of-fold predictions of the five-fold cross-validation (Section 3.3), and U_i is truncated to the unit interval, $U_i = \min\{1, \max(0, 1 - R^2{}_i)\}$, so that an objective with a negative out-of-fold $R^2{}_i$ is treated as entirely unexplained.

3.3. Residual Redundancy, Partial Correlation and Relation to Causal Inference

After conditioning, the dependence between objectives is evaluated using the residuals rather than the original objective values. The residual redundancy between two objectives f_i and f_j is defined as:

$$CR_{ij} = D(\varepsilon_i, \varepsilon_j) \tag{5}$$

where $D(\cdot,\cdot)$ is a dependence measure. In this work, we use the absolute value of the Pearson correlation:

$$CR_{ij} = |\, \rho(\varepsilon_i, \varepsilon_j)\, | \tag{6}$$

although other nonlinear measures, such as mutual information or kernel-based dependence [23], could also be employed. The key idea is that CR_{ij} captures the remaining association between objectives after removing the influence of the decision variables. Therefore, low values of CR_{ij} indicate that the observed dependence was primarily explained by shared decision-variable effects, whereas high values suggest residual statistical relationships that are not explained by the fitted conditioning model. Such residual association may reflect dependence beyond the observed inputs, but it may equally reflect conditioning-model error: when both conditioners miss the same input-dependent component, for example an interaction or a nonlinearity that the model cannot represent, the omitted component appears in both residuals and induces a residual correlation even though the dependence is entirely mediated by the decision variables. High values of CR_{ij} are therefore interpreted as dependence not explained by the chosen conditioner, and the adequacy gate and the per-objective diagnostics of Section 3.4 are used to judge whether model error is a plausible explanation.

Equation (6) uses the absolute value of the residual correlation, so that strong positive and strong negative residual associations are penalized alike. For minimization objectives, the two cases are not equivalent. A positive residual association means that, once the design-variable effects are removed, the two objectives tend to improve together, which is redundancy; a negative association means that improving one degrades the other, which is a trade-off that the reduced problem should retain. The absolute form is kept as the default because it is the symmetric quantity underlying partial correlation and the distance-based alternatives of this section, and because strong negative residual associations are rare in the cases studied: no residual pair in the admitted benchmark configurations has a residual correlation of -0.5 or less, and in the injection-molding case nine pairs do, but none reaches the blocking threshold (the minimum is -0.721), so the blocking rule is never triggered by

a conflict. A sign-aware variant is nevertheless defined, $CR^{+}_{ij} = \max\{0, \rho(\varepsilon_i, \varepsilon_j)\}$, which penalizes and blocks only positive residual association. It is recommended as the rule for problems in which all objectives are minimized, or have been brought to a common sense of optimization, because a negative residual association can only represent a trade-off; the absolute form should be used only as a check, or when the sense of optimization of the objectives is not fixed, as in exploratory analyses of indicator sets. The main results of Sections 4.4–4.6 use the absolute form because the active NSGA-III validation was designed and run with it. The complete results with the sign-aware form, including NSGA-III runs for the configurations in which the two forms select different subsets, are given in Supplementary Materials S2 and summarized in Section 4.4.6; they do not change the conclusions of the paper. Its effect is reported in Sections 4.4.6 and 4.6, and the controlled scenarios of Section 4.4.6 show cases in which the absolute form removes a conflicting objective that the sign-aware form retains.

When the same design matrix X is used for all objectives and g_i is linear, CR_{ij} is exactly the absolute value of the sample partial correlation of f_i and f_j given X; when g_i is nonlinear, it is the corresponding nonparametric residual dependence. It is a property of the triple (f_i, f_j, X) together with a model class, not an intrinsic property of the two objectives. The construction is classical—regression adjustment and partial correlation predate modern causal inference [19]—and what is proposed here is its use as an objective-selection criterion, the diagnostics that determine when that use is admissible (Section 3.4), and the evidence that the resulting subsets differ from those of marginal criteria. The decision variables are available in every evolutionary run and every design of experiments, yet existing objective-reduction methods do not use them, and Section 4.4.2 shows that using them changes the selected subset: in four of the five benchmark configurations admitted by the complete procedure, the conditioned subset differs from those of all three marginal baselines (ten of twelve if the historical subsets of the seven non-applicable configurations are also counted). No causal interpretation is assumed: a residual association would identify a direct effect between two objectives only if the structural model were correctly specified, if X were a sufficient adjustment set for the pair, and if the orientation of the remaining edge were established externally, and none of these conditions is required by the criterion.

Proposition 1. *Let $X \in \mathbb{R}^{N\times d}$ contain an intercept column and have full column rank, let $H = X(X^TX)^{-1}X^T$ and $M = I - H$, and let $\varepsilon_i = Mf_i$ be the in-sample least-squares residuals of objective f_i. Then:*
(a) $corr(\varepsilon_i, \varepsilon_j) = \rho_{ij\cdot X}$, the sample partial correlation of f_i and f_j given X, so that $CR_{ij} = |\rho_{ij\cdot X}|$ when D is the absolute Pearson correlation;
(b) $U_i = \|Mf_i\|^2/\|f_i - \overline{f_i}\|^2$, and $U_i = 0$ if and only if f_i lies in the column space of X; in that case $\varepsilon_i = 0$ and CR_{ij} is undefined for every j;
(c) if $f_i = h_i(X)$ is a deterministic function of the decision variables, as in the DTLZ and WFG problems, and the conditioner class is consistent for h_i, then $U_i \to 0$ as $N \to \infty$.

Proof. (a) Because the intercept belongs to the column space of X, the residuals Mf_i and Mf_j have zero mean, and their correlation is $\langle Mf_i, Mf_j\rangle/(\|Mf_i\|\,\|Mf_j\|)$, which is the sample partial correlation of f_i and f_j given X by the Frisch–Waugh–Lovell construction [19]. (b) M is the orthogonal projector onto the orthogonal complement of the column space of X, so $R^2_i = 1 - \|Mf_i\|^2/\|f_i - \overline{f_i}\|^2$, and $\|Mf_i\| = 0$ exactly when f_i lies in that column space. (c) For a deterministic response, the irreducible error is zero, so a consistent estimator drives the mean squared prediction error, and with it U_i, to zero. □

Proposition 1 has two consequences for the criterion. First, residual redundancy is not an intrinsic property of a pair of objectives. By (c), for deterministic objectives the residual

variance tends to zero as the conditioner approaches the true response function; consequently, CR_{ij} becomes increasingly model- and finite-sample-dependent and ultimately ill-defined as the residual variance vanishes, rather than converging to a meaningful intrinsic dependence. CR_{ij} should therefore be interpreted as the dependence that the chosen conditioner, at the available sample size, cannot attribute to the decision variables; the conditioner is deliberately regularized (Section 3.2) to keep this finite-sample quantity stable and interpretable. Second, (b) makes the degenerate case explicit: as $U_i \to 0$ the residual correlation becomes a ratio of vanishing quantities and carries no information, which is why the residual-variance floor of Section 3.4 is part of the criterion rather than an optional safeguard. The computations in this paper use out-of-fold rather than in-sample residuals, for which (a) holds approximately, with a discrepancy governed by the generalization gap; for the linear conditioner, this gap does not exceed 0.008 in any benchmark configuration (Section 4.4.2) and is 0.004 in the injection-molding case.

Three implementation details determine how closely the out-of-fold quantities follow Proposition 1. First, the out-of-fold coefficient of determination is computed from the pooled out-of-fold predictions of the five folds, that is, Equation (3) is evaluated with $\hat{g}_i(X^{(n)})$ replaced by the out-of-fold prediction $\hat{g}_i^{(-\kappa(n))}(X^{(n)})$, where $\kappa(n)$ is the fold that contains sample n, and U_i is obtained from Equation (4) with the truncation to $[0, 1]$ stated in Section 3.2; averaging the fold-wise values instead changes the mean R^2 of the injection-molding case from 0.9289 to 0.9283 (largest per-objective difference 0.0048). Second, U_i defined in this way is a normalized prediction error, which equals the residual-variance fraction only when the out-of-fold residuals have zero mean. The largest ratio between the absolute mean and the standard deviation of a residual is 0.0067 in the injection-molding case, so the two quantities differ by a relative amount below 10^{-4}, and the Pearson correlation of Equation (6) is in any case invariant to a constant offset. Third, a small generalization gap does not by itself imply that in-sample and out-of-fold residual correlations agree, so the agreement was checked directly: the largest difference between in-sample and out-of-fold residual correlations is 0.026 (mean 0.0066) in the injection-molding case and 0.0012 in the admitted benchmark configurations, and the selected subsets are identical in all cases. A negative out-of-fold R^2 would give $U_i = 1$ after truncation; such an objective is treated as unexplained by the conditioner and fails the per-objective diagnostic of Section 3.4. No negative out-of-fold R^2 occurred in any benchmark configuration or in the injection-molding case.

A further consequence of Equation (6) must be acknowledged. The absolute Pearson correlation detects only monotone linear association, whereas the residuals of a nonlinear conditioner may retain nonlinear structure, so a pair of objectives with strong nonlinear residual dependence could be assigned a low CR_{ij} and wrongly treated as non-redundant. Equation (5) is therefore defined for a general dependence measure D, and the distance correlation is implemented as an alternative to the absolute Pearson correlation. Distance correlation is preferred here to mutual information, which requires density estimation or binning choices that are unreliable at these sample sizes, and to the kernel-based HSIC statistic [23], which requires bandwidth selection; the corresponding population distance correlation is zero if and only if the residual random vectors are independent, under finite first moments [24], so it can detect nonlinear as well as linear dependence without requiring a bandwidth parameter. The subsets obtained with the two measures are compared in Section 4.6.

3.4. Admissibility Gate: Conditioner Adequacy and Residual-Variance Floor

Conditioning does not always produce interpretable residuals, and a criterion built on residuals must be able to report that fact rather than return a subset regardless. Two diagnostics are therefore evaluated before any objective is removed.

The first is conditioner adequacy. If the best available conditioning model explains little of the objective variance, the residuals are almost equal to the objectives themselves and CR_{ij} collapses onto the marginal dependence it was intended to replace; the conditioning step is then vacuous. The gate applied is a mean out-of-fold R^2 below 0.50: in that case the analysis is reported as uninformative and the selection is presented as marginal rather than conditioned. The threshold is deliberately permissive, since its purpose is to exclude vacuous conditioning, not to certify a good model. A single threshold of 0.50 is used throughout the paper, including the re-seeding study of Section 4.4.4. The value 0.50 corresponds to a conditioner that explains at least half of the objective variance on average, so that the residuals are no longer dominated by the objectives themselves; it is a scope condition rather than a tuning parameter. Because a mean can conceal poorly fitted objectives, the per-objective out-of-fold R^2 values are also reported (Section 4.5 for the injection-molding case). Objectives below 0.50 are flagged so that their residual correlations are interpreted with caution, but they remain candidates, because a low R^2 indicates a large unexplained fraction rather than a degenerate residual; among the admitted benchmark configurations, only DTLZ2 with ten objectives contains such objectives (four of ten). For the linear conditioner, which is the variant used in every admitted configuration, thresholds of 0.30, 0.40, 0.50, 0.60, 0.70 and 0.80 let 9, 8, 8, 7, 6 and 4 of the twelve benchmark configurations pass, and 6, 5, 5, 4, 3 and 1 are admitted after the residual-variance floor.

The second is a residual-variance floor, applied directly to U_i. Because $U_i = 1 - R^2{}_i$ is the fraction of the variance of f_i carried by its residual, a very small U_i means that ε_i is numerically indistinguishable from fitting error. The absolute correlation between two such residuals is then a correlation between two near-zero-variance numerical artifacts: it is unstable, it can take almost any value in $[0, 1]$, and it carries no information about the objectives. Objectives with U_i below a floor $U_{\min}$, set to 0.01 in the refined procedure, are therefore flagged as conditioning-degenerate and are removed from the candidate set of the sequential procedure entirely: a flagged objective cannot be the anchor, cannot be selected at any step, and its residual redundancies are excluded from $Red_i(S)$ for the remaining candidates; the requested subset size k counts only non-degenerate objectives. Flagged objectives are reported separately, as objectives whose retention cannot be decided by the residual criterion and must be referred to engineering judgment, in the same way as the forced inclusions of Section 4.5.

When fewer than k candidates satisfy the residual-variance floor, the refined selector returns a non-applicability flag. Objectives below the floor are reported, not selected, so the floor can reduce the number of candidates but never the requested subset size k; when it leaves fewer than k candidates, the procedure stops rather than returning a smaller subset. In that case, and when the adequacy gate fails, the fallback is a marginal method (PCA, mRMR-like selection or clustering), and the benchmark results of Section 4.4 report those configurations as non-applicable. The value $U_{\min} = 0.01$ flags objectives of which the conditioner explains more than 99% of the variance. It does not assert that the remaining 1% is numerical noise; it asserts that a residual correlation computed from so small a residual is dominated by the estimation error of the conditioner (Proposition 1(b)). Its sensitivity is reported in Section 4.6: on the benchmarks, values of 0.001 and 0.05 leave the five admitted subsets unchanged, and in the injection-molding case the three-objective subset is unchanged for $U_{\min}$ between 0.005 and 0.05.

The floor must operate in both directions; a weaker formulation, in which a flagged objective is merely exempted from redundancy-based removal, is not sufficient. Under that formulation, a flagged objective enters the score of Equation (10) with a zero redundancy penalty, and since $U_i \approx 0$ by definition, its score reduces to the relevance term alone. Relevance and unexplained variability are, moreover, structurally anti-correlated: an objective

that is almost determined by the decision variables has a variance driven by those variables. On the injection-molding candidate set, the Spearman correlation between Rel_i and U_i is -0.79, the mean relevance of the seven degenerate indicators is 0.185 against 0.063 for the remaining twenty-eight, and the six most relevant objectives of the set are all degenerate. Exempting degenerate objectives from the redundancy penalty would therefore make them automatic winners of the greedy sequence, which is the opposite of the intended protection; excluding them from the candidate set avoids this.

This diagnostic identifies exactly the situation that arises in DTLZ7, and explains the large negative reductions reported for that problem in Section 4.4.2. In DTLZ7, the first m $-$ 1 objectives are defined as $f_i(x) = x_i$, so each is an exact function of a single decision variable. Any conditioning model that includes those variables attains $R^2{}_i = 1$ up to floating-point error, U_i tends to zero, and the corresponding residuals are pure numerical noise. The reduction of -2321.5% reported for DTLZ7 with ten objectives is the arithmetic consequence: the denominator is an already small raw dependence of 0.010, while the numerator is the correlation between vectors of machine-precision noise. The percentage is correctly computed and physically meaningless. Under the floor introduced here, those objectives are flagged, the reduction percentage is not offered as evidence about redundancy, and the flag also prevents the criterion from declaring a genuine objective removable because its unexplained variability has been driven to zero. The same mechanism operates, far less severely, on seven of the thirty-five injection-molding indicators, all of them warpage or shrinkage components (Section 4.5).

3.5. Relevance of an Objective and the Role of Normalization

Objective selection requires balancing redundancy reduction with the preservation of informative objectives. To this end, a relevance score Rel_i is assigned to each objective. In its simplest form, relevance can be defined using the variance:

$$Rel_i = Var(f_i) \tag{7}$$

which reflects the ability of the objective to discriminate among solutions. More general formulations may incorporate entropy-based measures or decision-maker preferences:

$$Rel_i = \alpha_1 Var(f_i) + \alpha_2 Ent(f_i) + \alpha_3 w_i \tag{8}$$

where w_i represents optional weights reflecting the importance of each objective in the application context. In the numerical analyses, objective values and relevance terms are normalized before scoring to avoid dominance due to differences in physical units or numerical scales.

The interaction between Equation (7) and this normalization step requires clarification, because if normalization equalized the variances, then Rel_i would be constant and carry no information. The objectives are normalized to the unit interval by a min-max transformation, not standardized to unit variance. A min-max transformation equalizes the range of each objective but not its variance, since the variance of a bounded variable also depends on the shape of its distribution. Empirically, over the thirty-five injection-molding indicators, $Var(f_i)$ after min-max normalization spans $[0.022, 0.198]$, a factor of about nine between the least and the most dispersed objective, with a standard deviation across objectives of 0.060 (Section 4.5). The relevance term is therefore discriminative rather than degenerate. Had the objectives been standardized to zero mean and unit variance, $Var(f_i)$ would equal one for every objective by construction and Rel_i would be exactly non-informative; this is precisely why z-score standardization is not applied to the objective values. Being discriminative is not the same as being influential, however: the ablation of Section 4.6 shows that, in the injection-molding case, relevance is the least influential of the three terms, and removing it changes one objective out of five.

3.6. Selection Criterion and Standardization of Its Components

The selection process is formulated as a sequential optimization problem. Let $S \subset F$ denote the subset of selected objectives. For a candidate objective $f_i \notin S$, its residual redundancy with respect to the current subset is defined as:

$$Red_i(S) = \frac{1}{|S|} \sum_{f_j \in S} CR_{ij} \tag{9}$$

This quantity measures how much information f_i shares with the already selected objectives. The overall selection criterion is defined as:

$$Score_i(S) = \lambda\, Rel_i + \beta\, U_i - \kappa\, Red_i(S) \tag{10}$$

where λ, β, and κ are weighting parameters controlling the importance of relevance, unexplained variability, and redundancy, respectively. This formulation promotes objectives that are informative, not fully explained by the decision variables, and minimally redundant with the current subset. Initially, equal weights are adopted, $\lambda = \beta = \kappa = 1$ so that the three components contribute equally to the selection process. This avoids introducing additional tuning parameters and makes the selection primarily data-driven. Alternative weights may be used when decision-maker preferences or engineering priorities must be emphasized.

Equation (10) combines three quantities measured on different scales, so setting $\lambda = \beta = \kappa = 1$ equalizes the coefficients but not necessarily the influence of the three terms. On the injection-molding candidate set, the observed ranges are $[0.022, 0.198]$ for Rel_i, $[0.001, 0.223]$ for U_i, and $[0.000, 0.993]$ for the pairwise CR_{ij} values that enter $Red_i(S)$ (Section 4.5). Relevance and unexplained variability therefore occupy ranges of similar width, whereas the individual pairwise redundancies span almost the whole unit interval. Because $Red_i(S)$ is an average over the current subset, its own range is narrower than that of the pairwise values—$[0.072, 0.283]$ when averaged over all other objectives—but at the first steps of the greedy sequence, where S contains one or two objectives, it inherits the wide range of the pairwise values and can outweigh the other two terms.

To make the neutral default behave as intended, each of the three components is linearly rescaled to the unit interval over the candidates admitted at the current step, before the score is formed. Writing $z' = (z - \min z)/(\max z - \min z + \delta)$, where δ is a small constant preventing division by zero, the recommended form of the criterion is Equation (10) evaluated on the rescaled components Rel'_i, U'_i and $Red'_i(S)$ in place of their raw values; this standardized form of Equation (10) is used throughout, and no separate expression is required. This rescaling makes the numerical ranges of the three components comparable, so that with $\lambda = \beta = \kappa = 1$ the component with the widest raw range no longer dominates by construction. It does not guarantee that the three components have equal influence on the selection, which also depends on how their values are distributed and correlated across the candidates; the actual influence of each component is therefore measured by the ablation of Section 4.6 rather than assumed. The standardized form of Equation (10) is adopted as the recommended form of the criterion. Section 4.6 reports the subsets obtained with the unstandardized and the standardized forms of Equation (10) for the injection-molding case, together with an ablation in which each of the three terms is removed in turn, so that the contribution of the unexplained-variability term—the element that distinguishes the criterion from the standard relevance–redundancy framework—can be isolated rather than asserted. On the injection-molding case, the empirical effect of the rescaling is modest—the two forms of Equation (10) select the same three objectives and differ in one objective out of five—so the rescaling is retained mainly because it makes the weights act on comparable ranges, not because it changes that particular outcome.

All CI linear benchmark subsets reported in Section 4.4 are produced by the complete procedure of Algorithm 1, including the adequacy gate, the residual-variance floor and the

standardized form of Equation (10). Subsets produced by an earlier implementation of the selector are kept only as references and are labeled as historical.

Algorithm 1. Conditioning-Based Objective Selection

Require: X: decision-variable matrix; $F = \{f_1, \ldots, f_m\}$: objective set; k: number of objectives to select; λ, β, κ: weights; τ: blocking threshold; D: dependence measure; U_{min}: residual-variance floor; $R^2{}_{min}$: adequacy threshold
Ensure: S: selected subset; diagnostics: out-of-fold $R^2{}_i$, U_i, flagged set Φ
1: Normalize every f_i to $[0, 1]$ (min-max)
2: **for** each conditioner class $c \in$ {linear, nonlinear} **do**
3: **for** each $f_i \in F$ **do**
4: Fit $g_i{}^c$ on four folds and predict the held-out fold (five-fold shuffled CV)
5: $\varepsilon_i{}^c \leftarrow f_i - \hat{g}_i{}^c(X)$ (out-of-fold residuals); $R^2{}_{i,c} \leftarrow$ out-of-fold R^2
6: **end for**
7: **end for**
8: $c* \leftarrow$ among the classes with mean$_i$ $R^2{}_{i,c} \geq R^2{}_{min}$, the class with the lower mean residual dependence
9: if no class c satisfies mean$_i$ $R^2{}_{i,c} \geq R^2{}_{min}$ then return "conditioning uninformative" (fallback: marginal selection)
10: **for** each f_i: $U_i \leftarrow 1 - R^2{}_{i,c*}$; $Rel_i \leftarrow \mathrm{Var}(f_i)$
11: $\Phi \leftarrow \{f_i: U_i < U_{min}\}$; $C \leftarrow F \backslash \Phi$ (Φ is reported, not selected)
11a: **if** $|C| < k$ **then return** "non-applicable" (insufficient admissible candidates)
12: **for** each pair $f_i, f_j \in C$: $CR_{ij} \leftarrow D(\varepsilon_i, \varepsilon_j)$
13: $f_a \leftarrow \mathrm{argmin}_{i \in C}$ mean$_{j \neq i}$ CR_{ij} (ties: larger U_i, then lower index); $S \leftarrow \{f_a\}$
14: **while** $|S| < k$ **do**
15: $A \leftarrow \{f_i \in C \backslash S: \max_{j \in S} CR_{ij} \leq \tau\}$
16: **if** $A = \emptyset$ **then** $\tau \leftarrow \tau + 0.05$ and repeat line 15
17: $Red_i(S) \leftarrow$ mean$_{j \in S}$ CR_{ij} for every $f_i \in A$
18: Rescale *Rel*, *U* and *Red* to $[0, 1]$ over A (standardized Equation (10))
19: $f* \leftarrow \mathrm{argmax}_{i \in A}\ \lambda Rel'_i + \beta U'_i - \kappa Red'_i(S)$
20: $S \leftarrow S \cup \{f*\}$
21: **end while**
22: **return** S, Φ, $\{U_i\}$, $\{R^2{}_{i,c*}\}$

3.7. Anchor Selection, Tie Handling and Redundancy Blocking

The selection process is initialized by choosing an anchor objective that is globally least redundant:

$$f_a = \arg\min_{f_i \in F}\left(\frac{1}{m-1}\sum_{j \neq i} CR_{ij}\right) \tag{11}$$

This choice provides a stable starting point by selecting an objective that is broadly independent of the others. The subset is then constructed iteratively by selecting, at each step, the objective that maximizes the score:

$$f^\star = \arg\max_{f_i \notin S} Score_i(S) \tag{12}$$

and updating:

$$S \leftarrow S \cup \{f^\star\} \tag{13}$$

until a predefined number of objectives is reached.

Because the anchor is chosen by a criterion different from the one used in the subsequent steps, its influence is quantified rather than assumed to be benign: Section 4.6 reports the subsets obtained when every objective in turn is forced as the anchor. Ties are also resolved explicitly. When two or more objectives lie within 10^{-6} of the minimum average residual redundancy, the tie is broken by the larger U_i and, if still unresolved, by the lower objective index, so that the procedure remains fully deterministic and reproducible.

To prevent the selection of near-duplicate objectives, a blocking condition is applied. A candidate objective f_i is excluded if its maximum residual redundancy with any selected objective exceeds the threshold:

$$\max_{f_j \in S} CR_{ij} > \tau \tag{14}$$

Here, τ is a redundancy threshold. In this work, $\tau = 0.90$ was used as the default value. This threshold is used only as a near-duplicate blocking rule, not as a statistical significance or causal dependence threshold. If all candidates are blocked, τ is relaxed deterministically in increments of 0.05 until at least one candidate becomes admissible. This step only prevents premature termination before reaching the requested subset size and does not introduce randomness into the selection procedure.

3.8. Workflow Summary and Algorithm

Figure 2 makes the data flow explicit. The inputs are the decision-variable matrix and the objective values of the same sampled solutions; pre-processing applies the min-max normalization of Section 3.5. Stage 1 fits one conditioning model per objective and returns three quantities: the residual vectors, the unexplained fractions U_i and the out-of-fold R^2 values used to choose between the linear and nonlinear variants. These pass through the admissibility gate of Section 3.4, which either allows the analysis to proceed or reports that conditioning is uninformative, and which flags individual objectives whose residual variance is below the floor. Stage 2 converts the admissible residuals into the residual redundancy matrix CR and the relevance vector. Stage 3 selects the anchor, applies the blocking rule and grows the subset by maximizing the standardized form of Equation (10). The output is the subset itself, expressed in the original objectives, together with the diagnostics required to judge whether it should be trusted. The four control parameters and the stage each one enters are shown separately on the right. Algorithm 1 gives the same procedure in executable form.

The computational cost is dominated by two terms: fitting m independent conditioning models, and forming the m × m residual redundancy matrix. Writing $C(N, d)$ for the cost of one conditioning fit on N samples and d decision variables, the total cost is $O(m\, C(N, d) + m^2\, N)$ for the diagnostic stages plus $O(k\, m^2)$ for the greedy loop, so the procedure scales quadratically in the number of objectives and linearly in the requested subset size. No claim of optimality is made for the greedy sequence, and we believe none would be sound: the standard submodularity arguments do not transfer, because Redi(S) is an average over S rather than a maximum, and the blocking rule of Equation (14) makes the admissible candidate set state-dependent. Instead of asserting a guarantee, the practical consequence of greediness is quantified where it is most likely to matter, namely the dependence on the initialization, which is measured in Section 4.6.

Measured costs are small compared with the optimization itself. On a single CPU, the linear conditioning stage (five-fold out-of-fold residuals for all objectives) takes 0.2–0.8 s for 10,000 samples and ten objectives, the Extra Trees variant 25–71 s, and the greedy selection about 0.002 s; the complete linear pipeline for the 35 injection-molding indicators takes 0.01 s. By comparison, one NSGA-III run of a reduced DTLZ2 formulation takes about 12–17 s for m = 5 and 25–27 s for m = 10, and in the injection-molding application each evaluation requires a process simulation. The benefit of objective selection is therefore not a saving in the selection step but the smaller dimension of the optimization problem and of the set of objectives that the decision maker has to inspect.

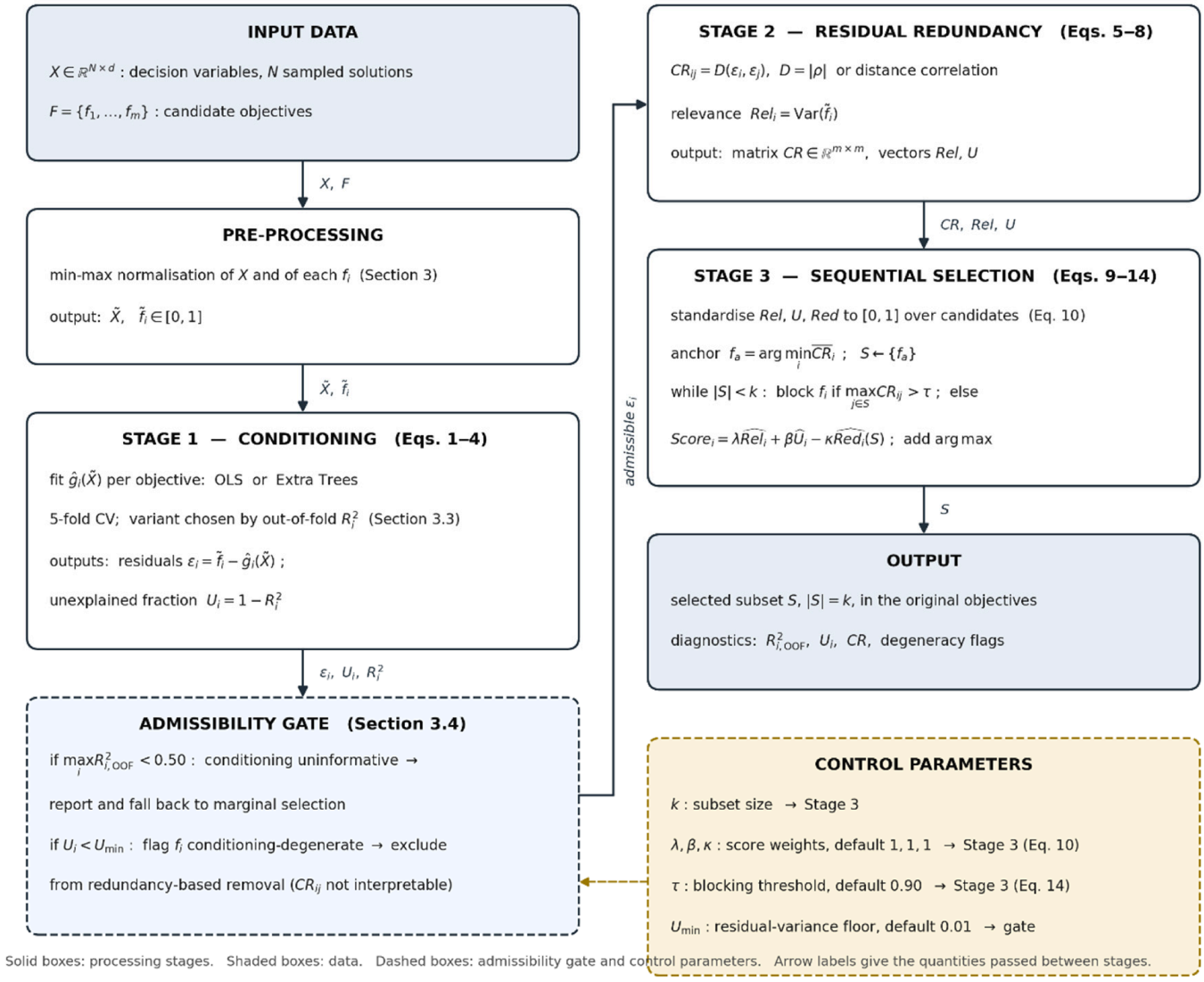


Figure 2. Data-flow diagram of the proposed objective selection method. Solid boxes are processing stages, shaded boxes are data, and dashed boxes are the control parameters and the admissibility gate of Section 3.4. Arrow labels give the quantities passed between stages. In the refined procedure, the gate excludes conditioning-degenerate objectives from the candidate set entirely, not only from the redundancy penalty.

4. Results and Discussion

4.1. Benchmark Problems and Setup

To evaluate the proposed conditioning-based objective selection approach, a set of well-established benchmark problems was considered. The analysis includes problems from the DTLZ and WFG families, which are widely used in the literature to assess many-objective optimization methods due to their controllable properties and diverse dependency structures.

For the benchmark problems, both linear and nonlinear conditioning models were evaluated. Linear conditioning was performed using ordinary least-squares regression, while nonlinear conditioning was used to assess whether nonlinear shared-variable effects remained after residualization. The dependence analysis of Sections 4.2 and 4.3 characterizes the behavior of the criterion; its performance as an objective-reduction method is assessed separately by the active optimization validation of Section 4.4.

The selected DTLZ problems (DTLZ2, DTLZ3, and DTLZ7) provide complementary characteristics. DTLZ2 is characterized by a smooth and continuous Pareto front with strong dependence among objectives induced primarily by shared decision variables. DTLZ3 introduces a more complex landscape with multiple local fronts, while maintaining a similar structural dependence between objectives. In contrast, DTLZ7 presents a disconnected Pareto front and a different dependency pattern, allowing the evaluation of the proposed method under varying structural conditions.

In addition to the DTLZ problems, the WFG9 problem was included to represent a more complex and nonlinear dependency structure. WFG9 is known for introducing non-separable transformations and deceptive features, leading to more intricate relationships between objectives. This makes it particularly suitable for assessing the ability of the proposed approach to distinguish between shared-variable effects and residual structural dependencies.

For each problem, multiple configurations were analyzed by varying the number of objectives. Specifically, cases with 3, 5, and 10 objectives were considered, enabling the evaluation of the method across increasing levels of dimensionality. The number of decision variables was defined according to standard benchmark configurations and adjusted to maintain consistency with the problem definitions and experimental setup.

The datasets used in the benchmark analysis were obtained from evolutionary optimization runs using NSGA-III. For each problem configuration, the population size and the number of generations depend on the number of objectives, as recommended in the literature. For all cases, the population size and the number of generations are as follows: for 3 objectives, 105 individuals and 500 generations; for 5 objectives, 210 individuals and 1000 generations; and for 10 objectives, 275 individuals and 1200 generations. For the WFG9 problem, the number of decision variables varies with the number of objectives as follows: 24 for 3 objectives, 28 for 5 objectives, and 38 for 10 objectives. These sampled solutions were used only for dependence and residual-redundancy analysis, not to compare optimizer performance. Objective and decision-variable values were normalized before regression fitting and dependence estimation. These datasets provide a diverse set of solutions that reflect the structure of the objective space and allow the analysis of both raw and conditioned dependencies. The configurations are summarized in Table 2.

Table 2. Benchmark problems and configurations used in the study.

Problem	Decision Variables	Objectives	Characteristics
DTLZ2	10	3, 5, 10	Smooth Pareto front; strong shared-variable dependence
DTLZ3	10	3, 5, 10	Multimodal landscape; similar dependency structure to DTLZ2
DTLZ7	10	3, 5, 10	Disconnected Pareto front; heterogeneous dependencies
WFG9	24/28/38	3/5/10	Nonlinear, non-separable, deceptive dependencies

Tables 3 and 4 and Figure 3 are computed on these final populations (105, 210 and 275 solutions for three, five and ten objectives), after min-max normalization of objectives and decision variables. The selection analysis of Section 4.4 uses a different sample: 10,000 solutions drawn uniformly in the decision space (sample seed 42 for Tables 5 and 6, seed 2026 for the re-seeding study), also min-max normalized. The NSGA-III populations are concentrated near the Pareto front, whereas the uniform sample covers the whole design space, so the raw dependence values of Tables 3 and 4 and of Table 5 differ for the same configuration and are not directly comparable.

Table 3. Mean absolute objective dependence for the DTLZ benchmark problems before and after conditioning on decision variables. Computed on the final NSGA-III populations of Section 4.1 (min-max normalized); Table 5 reports the analysis on the 10,000-solution uniform sample.

Problem	Objectives	Raw	Linear Conditioned	Nonlinear Conditioned	Mean Linear (R^2)	Mean Nonlinear (R^2)
DTLZ2	3	0.451	0.326	0.293	0.907	0.999
DTLZ2	5	0.221	0.161	0.263	0.838	0.996
DTLZ2	10	0.100	0.097	0.061	0.768	0.981
DTLZ3	3	0.451	0.323	0.335	0.899	0.999
DTLZ3	5	0.221	0.132	0.233	0.848	0.996
DTLZ3	10	0.100	0.101	0.067	0.767	0.984
DTLZ7	3	0.442	0.334	0.144	0.945	0.999
DTLZ7	5	0.225	0.219	0.044	0.988	0.998
DTLZ7	10	0.076	0.196	0.050	0.994	0.991

Table 4. Mean absolute objective dependence for the WFG9 benchmark problem before and after conditioning on decision variables. Computed on the final NSGA-III populations of Section 4.1 (min-max normalized); Table 5 reports the analysis on the 10,000-solution uniform sample.

Problem	Objectives	Raw	Linear Conditioned	Nonlinear Conditioned	Mean Linear (R^2)	Mean Nonlinear (R^2)
WFG9	3	0.456	0.430	0.410	0.653	0.975
WFG9	5	0.223	0.212	0.127	0.479	0.923
WFG9	10	0.099	0.104	0.071	0.366	0.832

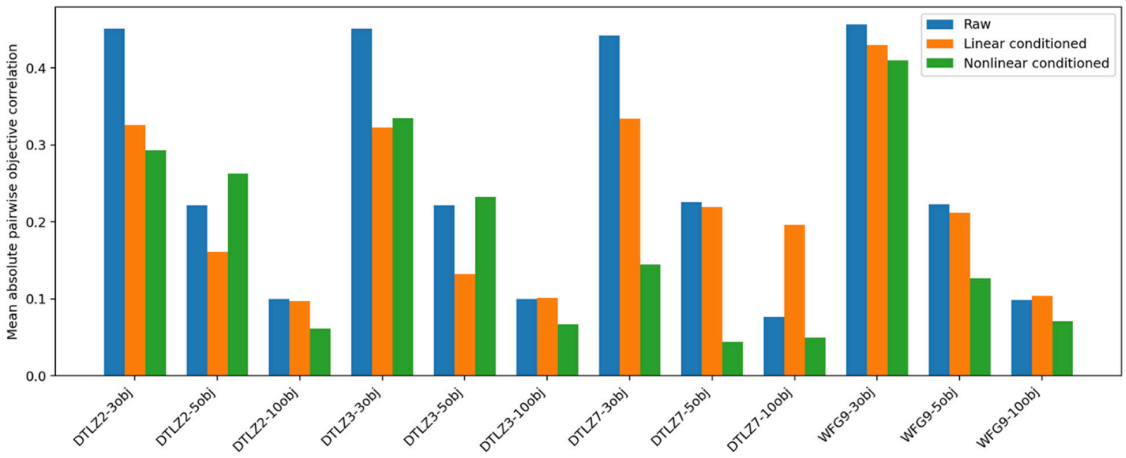


Figure 3. Mean absolute pairwise objective dependence for the benchmark problems before and after linear and nonlinear conditioning.

Table 5. Conditioning-model diagnostics for the benchmark cases. Train/CV R^2 values are averaged over objectives. Raw/residual dependence reports the mean absolute Pearson correlation before and after conditioning. Linear rows: subsets returned by the complete procedure (Algorithm 1); n.a. denotes a non-applicable configuration (gate: mean out-of-fold R^2 below 0.50 for every variant; floor: fewer than k objectives with $U_i \geq 0.01$), with the historical subset evaluated as an exploratory comparison in brackets. Nonlinear rows: subsets of the CI nonlinear robustness diagnostic used in the validation. All values are computed on the 10,000-solution uniform sample (seed 42).

Problem	m	Model	Train/CV R^2	Gap	Raw/Residual Dep.	Red. (%)	Selected Objectives
DTLZ2	3	Linear	0.841/0.840	0.000	0.371/0.325	12.4	f1, f2
DTLZ2	3	Nonlinear	0.995/0.971	0.024	0.371/0.632	−70.3	f3, f2
DTLZ2	5	Linear	0.755/0.755	0.001	0.251/0.147	41.4	f3, f1, f2

Table 5. *Cont.*

Problem	m	Model	Train/CV R^2	Gap	Raw/Residual Dep.	Red. (%)	Selected Objectives
DTLZ2	5	Nonlinear	0.995/0.976	0.020	0.251/0.538	−114.2	f2, f5, f4
DTLZ2	10	Linear	0.592/0.591	0.001	0.248/0.133	46.4	f7, f3, f8, f6, f5
DTLZ2	10	Nonlinear	0.993/0.964	0.029	0.248/0.280	−13.1	f2, f10, f1, f9, f3
DTLZ3	3	Linear	0.745/0.744	0.001	0.280/0.357	−27.6	f1, f2
DTLZ3	3	Nonlinear	0.966/0.834	0.132	0.280/0.590	−110.9	f3, f2
DTLZ3	5	Linear	0.657/0.656	0.001	0.206/0.224	−8.6	f1, f4, f2
DTLZ3	5	Nonlinear	0.965/0.835	0.130	0.206/0.529	−156.6	f5, f4, f3
DTLZ3	10	Linear	0.323/0.321	0.002	0.275/0.321	−16.8	n.a. (gate) [f1, f10, f2, f9, f3]
DTLZ3	10	Nonlinear	0.916/0.590	0.326	0.275/0.441	−60.4	f1, f10, f9, f2, f8
DTLZ7	3	Linear	0.990/0.990	0.000	0.101/0.179	−78.0	n.a. (floor) [f3, f1]
DTLZ7	3	Nonlinear	0.994/0.964	0.030	0.101/0.010	90.3	f2, f1
DTLZ7	5	Linear	0.997/0.997	0.000	0.036/0.174	−385.3	n.a. (floor) [f5, f4, f1]
DTLZ7	5	Nonlinear	0.998/0.988	0.010	0.036/0.008	76.8	f2, f3, f1
DTLZ7	10	Linear	1.000/1.000	0.000	0.010/0.233	−2321.5	n.a. (floor) [f10, f8, f3, f5, f7]
DTLZ7	10	Nonlinear	1.000/1.000	0.000	0.010/0.010	−5.6	f5, f8, f9, f3, f4
WFG9	3	Linear	0.116/0.111	0.005	0.459/0.421	8.1	n.a. (gate) [f1, f2]
WFG9	3	Nonlinear	0.937/0.254	0.683	0.459/0.414	9.6	f1, f2
WFG9	5	Linear	0.119/0.113	0.006	0.341/0.280	18.0	n.a. (gate) [f4, f5, f3]
WFG9	5	Nonlinear	0.936/0.218	0.718	0.341/0.291	14.7	f4, f3, f2
WFG9	10	Linear	0.113/0.106	0.008	0.320/0.260	19.0	n.a. (gate) [f9, f8, f10, f1, f7]
WFG9	10	Nonlinear	0.936/0.174	0.762	0.320/0.283	11.8	f9, f8, f7, f2, f6

Table 6. Objective subsets considered in the complete active optimization validation. Reduced formulations select two objectives for m = 3, three objectives for m = 5 and five objectives for m = 10. CI linear: subsets of the complete procedure (Algorithm 1); n.a. denotes a non-applicable configuration (gate or floor), with the historical subset evaluated as an exploratory comparison in brackets. CI nonlinear: robustness diagnostic. mRMR denotes the mRMR-like baseline.

Problem	m	k	CI Linear	CI Nonlinear	PCA	mRMR	Clustering
DTLZ2	3	2	f1, f2	f3, f2	f3, f2	f2, f1	f1, f2
DTLZ2	5	3	f3, f1, f2	f2, f5, f4	f5, f4, f3	f4, f3, f5	f1, f4, f3
DTLZ2	10	5	f7, f3, f8, f6, f5	f2, f10, f1, f9, f3	f4, f10, f9, f8, f7	f10, f9, f8, f7, f6	f5, f10, f8, f7, f6
DTLZ3	3	2	f1, f2	f3, f2	f3, f2	f3, f2	f3, f2
DTLZ3	5	3	f1, f4, f2	f5, f4, f3	f5, f4, f3	f5, f4, f3	f5, f1, f3
DTLZ3	10	5	n.a. [f1, f10, f2, f9, f3]	f1, f10, f9, f2, f8	f4, f9, f10, f6, f8	f10, f9, f8, f7, f6	f6, f10, f9, f8, f7
DTLZ7	3	2	n.a. [f3, f1]	f2, f1	f3, f2	f1, f2	f1, f2
DTLZ7	5	3	n.a. [f5, f4, f1]	f2, f3, f1	f5, f1, f2	f3, f1, f4	f4, f2, f3
DTLZ7	10	5	n.a. [f10, f8, f3, f5, f7]	f5, f8, f9, f3, f4	f10, f8, f1, f3, f4	f8, f9, f5, f3, f4	f8, f4, f5, f9, f3
WFG9	3	2	n.a. [f1, f2]	f1, f2	f3, f1	f3, f2	f3, f1
WFG9	5	3	n.a. [f4, f5, f3]	f4, f3, f2	f5, f4, f3	f5, f4, f3	f1, f5, f3
WFG9	10	5	n.a. [f9, f8, f10, f1, f7]	f9, f8, f7, f2, f6	f3, f10, f9, f8, f7	f10, f9, f8, f7, f6	f4, f10, f8, f7, f6

4.2. Raw Versus Conditioned Objective Dependence

This section evaluates how objective dependence changes after conditioning on the decision variables. The objective is to quantify how much of the observed dependence is

attributable to shared decision variables and how much reflects residual structural relationships not explained by the observed decision variables. Three levels of dependence were considered: raw objective dependence, dependence after linear conditioning, and dependence after nonlinear conditioning. Raw dependence was computed directly from the objective values, whereas conditioned dependence was computed from residuals obtained after modeling each objective as a function of the decision variables.

Figure 3 shows the mean absolute pairwise objective dependence for all benchmark cases. The results indicate that conditioning generally reduces the apparent dependence between objectives, confirming that a significant part of the observed objective relationships is induced by shared decision variables rather than by residual relationships not captured by the conditioning model.

The detailed numerical results for the DTLZ problems are reported in Table 3. For DTLZ2 with three objectives, the raw dependence is 0.451 and decreases to 0.326 after linear conditioning and to 0.293 after nonlinear conditioning. A similar trend is observed for DTLZ7, where the three-objective case decreases from 0.442 to 0.334 after linear conditioning and to 0.144 after nonlinear conditioning. These results indicate that most of the dependence in DTLZ problems is induced by shared-variable effects, which are effectively removed through conditioning.

For higher-dimensional cases, Table 3 shows that dependence decreases significantly as the number of objectives increases, and that nonlinear conditioning further reduces the remaining dependence. This reinforces the interpretation that DTLZ problems exhibit predominantly shared-variable-induced redundancy rather than strong residual dependencies under the fitted conditioning model.

In contrast, the results for the WFG9 problem, presented in Table 4, show a different behavior. In the three-objective case, the raw dependence is 0.456 and remains relatively high after linear conditioning (0.430) and nonlinear conditioning (0.410). This indicates that WFG9 retains stronger residual dependencies between objectives compared to the DTLZ problems.

For the five- and ten-objective cases, Table 4 shows that nonlinear conditioning reduces dependence more substantially. However, residual dependence is still present, particularly when compared to the DTLZ problems. This suggests that WFG9 includes more complex structural relationships between objectives that are not fully explained by the decision variables.

Three groups of cases in Tables 3 and 4 run against this general trend and are addressed here explicitly rather than passed over. For DTLZ2 with five objectives, the mean dependence is 0.221 raw, 0.161 after linear conditioning, and 0.263 after nonlinear conditioning; DTLZ3 with five objectives behaves in the same way (0.221, 0.132, 0.233); and for WFG9 the reduction is modest at every dimensionality. In none of these cases does the conditioning model fail numerically—on DTLZ3 with five objectives, the nonlinear model reaches a mean R^2 of 0.996 in Table 3. What increases is the dependence between the residuals, for the following reason.

In DTLZ2 and DTLZ3, every objective contains the same distance term, which enters multiplicatively. A flexible conditioner absorbs the separable angular part of each objective far more effectively than this shared multiplicative factor, so the residuals retain a common component while their variance has been substantially reduced. The correlation of the residuals then rises even though the model is accurate, and the increase is informative: it states that the residual similarity between these objectives is not attributable to the individual decision variables and that the objectives should not be treated as interchangeable. For WFG9, the explanation is the opposite one. There the conditioning model is too weak—a mean linear R^2 of 0.366 to 0.653, and an out-of-fold R^2 of only 0.12 to 0.16 for the nonlinear variant

(Section 4.4.4)—so the residuals remain close to the raw objectives and the conditioned and raw dependences are necessarily close as well.

These cases therefore delimit the scope of the criterion instead of contradicting it. Conditioning separates shared-variable dependence from residual dependence whenever the conditioning model is adequate and the residual variance is not degenerate; where either condition fails, the criterion is required to report that fact through the diagnostics of Section 3.4 rather than to return a reduction. In Section 4.4 the gate and the floor are applied, as part of the complete procedure, to all twelve benchmark configurations.

4.3. Residual Objective–Objective Structure

To examine the structure of the residual dependencies, they are represented as weighted graphs in which the nodes are objectives and the edge weights are absolute residual correlations, without thresholding. Figure 4 shows two representative DTLZ cases at the lowest dimensionalities, where the graphs remain legible; they are illustrations, and the general statements rest on the scalar statistics of Tables 3–5. After conditioning, most connections in DTLZ2 become weak, indicating that the observed dependence was induced by shared decision variables, whereas DTLZ7 retains some residual connections in the lower-dimensional cases. WFG9 retains stronger and more persistent residual connections at every dimensionality, consistent with Table 4; as shown in Section 4.4, however, this persistence reflects the inadequacy of the conditioning model rather than intrinsic structure. A comparison of the initial and final sampled populations further showed that the dependence structure is not static during optimization, so a subset selected from an initial sample may not represent dependencies that emerge later in the search.

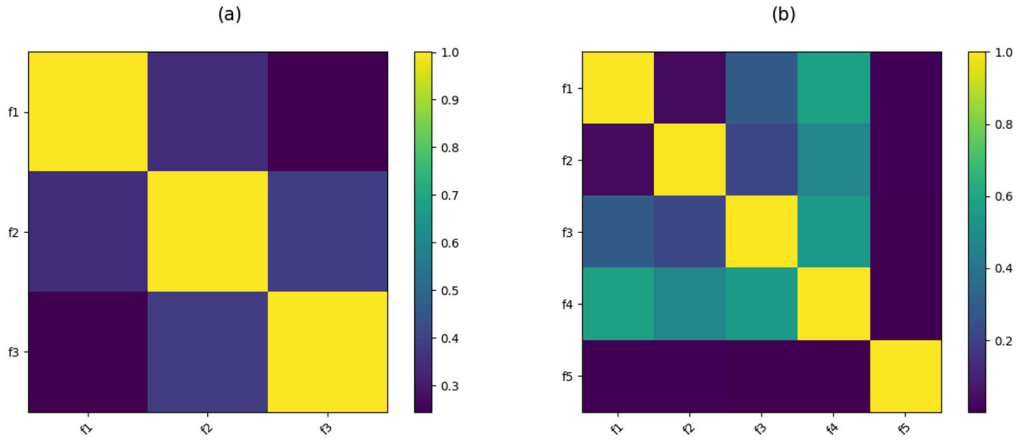


Figure 4. Residual objective–objective dependence for representative DTLZ cases after conditioning on decision variables: (**a**) DTLZ2 with three objectives and (**b**) DTLZ7 with five objectives.

4.4. Optimization Validation and Stochastic Stability of Selected Objective Subsets

The previous sections evaluate objective redundancy through raw and conditioned dependence. This analysis identifies objectives whose apparent similarity is due to shared decision-variable effects, but it does not, by itself, demonstrate that the selected subsets remain useful when used directly within an optimizer. To address this limitation, an active optimization validation is performed. NSGA-III was run on the full objective set and on reduced objective subsets obtained using the proposed conditioning-based selector and classical objective-reduction baselines. All final populations were then re-evaluated in the original objective space.

This validation answers three questions: whether a reduced subset selected by the proposed method can be used as the objective vector of an evolutionary optimizer; how

much full-space performance is lost when the optimizer sees only a subset of objectives; and how the conditioning-based selector compares with PCA, mRMR and correlation-based clustering under the same optimization protocol.

4.4.1. Active Optimization Validation Protocol

Let the original m-objective problem be written as $F(x) = (f_1(x), f_2(x), \ldots, f_m(x))$, where all objectives are treated in minimization form. For a reduced subset $S \subset \{1, \ldots, m\}$, NSGA-III optimizes only $F_S(x) = (f_j(x): j \in S)$. After optimization, however, each final population obtained from F_S is re-evaluated as $F(x)$ in the original m-objective space. The reduced formulation is therefore used only during optimization, whereas performance assessment is performed in the full objective space.

The validation was carried out on DTLZ2, DTLZ3, DTLZ7 and WFG9 using $m = 3$, $m = 5$ and $m = 10$ objectives. DTLZ2, DTLZ3 and DTLZ7 used 10 decision variables. WFG9 used 24, 28 and 38 decision variables for the three-, five- and ten-objective cases, respectively. Reduced formulations retained two objectives when $m = 3$, three objectives when $m = 5$ and five objectives when $m = 10$.

The compared formulations were:

- the full objective formulation, denoted Full objectives;
- the proposed conditioning-based selector with linear residualization, denoted CI linear;
- a nonlinear conditioning variant, denoted CI nonlinear, included as a robustness diagnostic;
- PCA-based objective selection (returning original objectives, not principal components);
- mRMR-like objective selection (relevance–redundancy with Pearson correlation in place of mutual information);
- correlation-based clustering.

The objective subsets were generated from 10,000 sampled solutions before optimization. For each problem, objective dimensionality and method, 30 independent NSGA-III runs were performed. The population sizes and generation budgets were 105/500, 210/1000 and 275/1200 for the three-, five- and ten-objective cases, respectively. NSGA-III [1] was chosen because it is the reference-direction algorithm designed for many-objective problems and is widely used as the reference optimizer for the DTLZ and WFG families; because its reference directions are generated in the dimension of the problem being optimized, the comparison between formulations of different dimensions is well defined. For each run, the reference directions were generated with the Riesz s-energy method in the dimension of the formulation (m for the full problem and k for a reduced one), with as many directions as the population size and the run seed as the generator seed. The full and reduced formulations of a configuration therefore used the same population size, the same number of generations and the same 30 seeds (1–30). That is, identical evaluation budgets, and the reduced problems were evaluated in the full objective space only after optimization. All runs used pymoo 0.6.2.

The CI linear subsets evaluated here are those of the complete procedure (Algorithm 1). For DTLZ2 with three objectives and DTLZ3 with three and five objectives, they coincide, as sets, with the subsets of the earlier implementation, so the archived runs remain valid; for DTLZ2 with five and ten objectives the subsets changed (Table 6), and 30 new NSGA-III runs were performed for each with the settings above. In the seven non-applicable configurations, the procedure returns no subset, and the fallback is a marginal method; the historical CI linear subsets of these configurations are still evaluated, but only as exploratory comparisons. The baseline configurations are specified below so that the comparison can be reproduced.

PCA-based selection, which returns original objectives rather than latent variables, used k principal components of the standardized objective values, where k was the requested subset size [4]. For each component, the objective with the greatest absolute loading was

retained without duplication; remaining positions, if any, were filled by scores based on absolute loadings weighted by explained-variance ratios. The mRMR-like baseline, so called because it uses Pearson correlation instead of mutual information and is labeled mRMR in the tables, used the variance of min-max normalized objectives as relevance and the mean absolute Pearson correlation with previously selected objectives as redundancy; the next objective maximized relevance minus redundancy with unit weights [5,18]. The clustering baseline grouped objectives using absolute correlation-based distances and retained one representative per group [18].

All four methods receive the identical 10,000 sampled solutions and the identical normalized objective values.

4.4.2. Conditioning-Model Diagnostics and Selected Subsets

Before interpreting the optimization results, the conditioning models themselves were analyzed. Each objective was modeled as $f_i(x) = g_i(x) + \varepsilon_i$, where $g_i(x)$ is the fitted conditioning model, and ε_i is the residual component used in the conditioned-dependence analysis. Linear and nonlinear residualization were compared using mean training R^2, mean cross-validated R^2, the generalization gap, raw objective dependence, residual dependence, and the corresponding selected objective subset.

The residual dependence reported in Table 5 is the mean absolute Pearson correlation between cross-validated residuals. The reduction percentage compares residual dependence with raw dependence. Positive values indicate that conditioning reduced average objective-objective dependence, whereas negative values indicate that the residuals became more mutually dependent after conditioning.

For all benchmark problems, nonlinear diagnostics were recomputed with independent Extra Trees models for each objective (300 trees, minimum leaf size 3, five-fold cross-validation). These diagnostics use a more flexible model than the historical subset selector (100 trees, maximum depth 8, minimum leaf size 20). These diagnostics do not affect the selected subsets, which are produced with the selector configuration.

Table 5 shows that the effect of conditioning is problem-dependent. In DTLZ2, linear conditioning reduces residual dependence at five and ten objectives, while the independent nonlinear models achieve higher predictive R^2 but increase residual dependence at all three dimensionalities. In DTLZ3, nonlinear conditioning also improves predictive R^2 but increases residual dependence at every dimensionality. DTLZ7 requires caution because its first $m - 1$ objectives are direct functions of the decision variables, so near-zero residual variance can make residual correlations uninterpretable. In WFG9, the more flexible diagnostic model also remains weak out of fold; WFG9 is therefore non-applicable at every dimensionality.

Three patterns help interpret the reduction column. Linear conditioning reduces dependence in the five- and ten-objective DTLZ2 cases (41.4% and 46.4%), whereas conditioning on DTLZ7 produces near-zero residual variance for the first $m - 1$ objectives, making residual correlations unsuitable for selection. The independent nonlinear fits provide high cross-validated R^2 for DTLZ2 and DTLZ3, but their residual correlations need not decrease: accurate prediction and low residual redundancy are different properties. For WFG9, the diagnostic Extra Trees fit and the more strongly regularized historical selector have different predictive adequacy; the applicability assessment uses the selector configuration, under which both variants fail the gate (mean out-of-fold $R^2 \leq 0.16$).

Applying the complete procedure to the twelve configurations gives the outcome shown in the linear rows of Table 5. The adequacy gate fails for WFG9 at every dimensionality (mean out-of-fold $R^2 \leq 0.16$ for both variants) and for DTLZ3 with ten objectives (0.321 linear and 0.498 nonlinear). For DTLZ7 both variants pass the gate, but 2, 4 and 10 objectives fall

below the residual-variance floor for m = 3, 5 and 10, leaving fewer than k candidates, so these configurations are also non-applicable. The remaining five configurations, DTLZ2 at m = 3, 5 and 10 and DTLZ3 at m = 3 and 5, are admitted, and in all five the linear variant is used because it has the lower residual dependence (Section 3.2). Compared with the earlier implementation, the subsets are unchanged as sets for DTLZ2 with three objectives and DTLZ3 with three and five objectives, and they change for DTLZ2 with five and ten objectives. Among the admitted configurations, only DTLZ2 with ten objectives has objectives with an out-of-fold R^2 below 0.50 (four of ten), and no objective has a negative value.

The stability of the final procedure was examined in three ways. Re-seeding the five-fold partition 30 times returns the same subset in every repetition for all five admitted configurations. The subsets are nested in k, because the greedy sequence is deterministic: for k = m − 1 the selection order is (f3, f1, f2, f4) for DTLZ2 with five objectives and (f7, f3, f8, f6, f5, f9, f2, f10, f4) with ten, so the subsets for smaller k are its leading elements. Drawing ten new 10,000-solution samples changes the subset more often: the modal subset is obtained in 10, 5, 5, 8 and 5 of the 10 samples for DTLZ2 (m = 3, 5, 10) and DTLZ3 (m = 3, 5). In DTLZ2 the objectives play nearly symmetric roles, so several subsets are statistically almost equivalent and the sample determines which of them is returned; the sample size, rather than the estimator seed, is the main source of variability of the final procedure.

Table 6 reports the exact subsets used in the active optimization validation. The differences between the conditioning-based subsets and those produced by PCA, mRMR and clustering confirm that conditioning-based redundancy is not simply reproducing a marginal dependence criterion.

This can be stated quantitatively. In the five admitted configurations, the subset returned by the complete procedure differs from all three baseline subsets in four (DTLZ2 at m = 5 and 10, DTLZ3 at m = 3 and 5); for DTLZ2 with three objectives it coincides with the mRMR-like and clustering subsets. Over all twelve configurations, counting the historical subsets of the non-applicable ones, the linear subsets differ from all three baselines in ten. Since the baselines use only the objective values while the proposed criterion also uses the decision variables, this shows that the selectors declare different objectives redundant. Because the baselines also differ from the proposed selector in their score and search, the contribution of conditioning itself is isolated by the matched comparison of Section 4.4.6, which shows that conditioning alone changes the subset in three of the five admitted configurations. It does not, by itself, establish that the conditioned subsets are better; that question is addressed by the optimization results below and answered without a claim of dominance.

Under the complete procedure, five configurations are admitted (DTLZ2 at m = 3, 5 and 10; DTLZ3 at m = 3 and 5) and seven are non-applicable (all DTLZ7 and WFG9 configurations and DTLZ3 with ten objectives). For the latter, the recommended action is a marginal method; their CI linear rows in Tables 7 and 8 are retained as exploratory comparisons and are counted separately below.

Table 7. Active optimization validation in the original objective space. Values are means over 30 independent NSGA-III runs. HV is maximized, while IGD+, spacing and removed-objective loss are minimized. For DTLZ2 with m = 5 and 10, all rows were re-evaluated after adding 30 runs of the CI linear subset of the complete procedure, CI linear (final); the subset of the earlier implementation is listed as CI linear (historical). In the non-applicable configurations (DTLZ3 with m = 10, DTLZ7 and WFG9), the CI linear rows are exploratory. mRMR denotes the mRMR-like baseline.

Problem	m	Method	HV	IGD+	Spacing	ND	Removed Loss
DTLZ2	3	Full objectives	1.0692	0.0012	0.0286	105	–
DTLZ2	3	CI linear	0.1511	0.1999	2.279×10^{-4}	105	0.5200
DTLZ2	3	CI nonlinear	0.6885	0.4317	0.0088	6.77	0.3049

Table 7. *Cont.*

Problem	m	Method	HV	IGD+	Spacing	ND	Removed Loss
DTLZ2	3	PCA	0.6885	0.4317	0.0088	6.77	0.3049
DTLZ2	3	mRMR	0.1511	0.1999	2.279×10^{-4}	105	0.5200
DTLZ2	3	Clustering	0.1511	0.1999	1.683×10^{-4}	105	0.5200
DTLZ2	5	Full objectives	1.5930	0.0045	0.0307	210	–
DTLZ2	5	CI linear (historical)	0.9598	0.2346	0.0382	43.60	0.1892
DTLZ2	5	CI linear (final)	0.9122	0.1938	2.407×10^{-4}	210	0.0106
DTLZ2	5	CI nonlinear	1.0334	0.3890	0.0310	69.77	0.1543
DTLZ2	5	PCA	0.9930	0.4214	0.0392	48.50	0.2107
DTLZ2	5	mRMR	1.0292	0.4048	0.0339	86.33	0.1874
DTLZ2	5	Clustering	0.9122	0.1938	3.394×10^{-4}	210	0.0848
DTLZ2	10	Full objectives	2.5694	0.0272	0.0803	275	–
DTLZ2	10	CI linear (historical)	0.3913	0.6289	3.391×10^{-4}	275	0.0450
DTLZ2	10	CI linear (final)	0.3913	0.6289	3.584×10^{-4}	275	0.0427
DTLZ2	10	CI nonlinear	0.9056	0.5288	0.0343	275	0.0365
DTLZ2	10	PCA	0.9325	0.6348	0.0208	244.60	0.0355
DTLZ2	10	mRMR	1.3474	0.5895	0.0579	264.67	0.1338
DTLZ2	10	Clustering	0.8497	0.5217	0.0262	271.50	0.0249
DTLZ3	3	Full objectives	1.3309	8.078×10^{-5}	0.0227	105	–
DTLZ3	3	CI linear	0.5095	0.2412	2.600×10^{-5}	105	0.3612
DTLZ3	3	CI nonlinear	1.1456	0.1528	0.0256	1.70	0.1529
DTLZ3	3	PCA	1.1456	0.1528	0.0256	1.70	0.1529
DTLZ3	3	mRMR	1.1456	0.1528	0.0256	1.70	0.1529
DTLZ3	3	Clustering	1.1456	0.1528	0.0256	1.70	0.1529
DTLZ3	5	Full objectives	1.6105	6.228×10^{-5}	0.0122	210	–
DTLZ3	5	CI linear	1.5927	0.0079	1.611×10^{-6}	210	−0.1241
DTLZ3	5	CI nonlinear	1.4558	0.1233	0.0524	54.80	0.1253
DTLZ3	5	PCA	1.4558	0.1233	0.0524	54.80	0.1253
DTLZ3	5	mRMR	1.4558	0.1233	0.0524	54.80	0.1253
DTLZ3	5	Clustering	1.6034	0.0054	0.0223	210	0.1537
DTLZ3	10	Full objectives	2.5937	3.876×10^{-4}	0.0093	275	–
DTLZ3	10	CI linear	2.5591	0.0151	0.0026	275	−0.0154
DTLZ3	10	CI nonlinear	2.5893	0.0029	0.0025	275	−0.0299
DTLZ3	10	PCA	2.5886	0.0028	0.0087	265.97	4.268×10^{-4}
DTLZ3	10	mRMR	2.5907	0.0034	0.0082	266.27	0.0019
DTLZ3	10	Clustering	2.5904	0.0036	0.0061	266.33	0.0011
DTLZ7	3	Full objectives	1.2311	0.0073	0.0198	105	–
DTLZ7	3	CI linear	0.2057	0.2936	0.0034	105	0.4760
DTLZ7	3	CI nonlinear	0.7244	0.4496	–	1.27	0.4196
DTLZ7	3	PCA	0.1876	0.2865	0.0034	105	0.4994
DTLZ7	3	mRMR	0.6858	0.4816	–	1.27	0.4498
DTLZ7	3	Clustering	0.6858	0.4816	–	1.27	0.4498

Table 7. *Cont.*

Problem	m	Method	HV	IGD+	Spacing	ND	Removed Loss
DTLZ7	5	Full objectives	1.4792	0.0191	0.0754	210	–
DTLZ7	5	CI linear	0.0843	0.3948	0.0123	210	0.5053
DTLZ7	5	CI nonlinear	0.8503	0.3723	0.0633	9.40	0.1719
DTLZ7	5	PCA	0.0832	0.3914	0.0129	210	0.5070
DTLZ7	5	mRMR	0.8313	0.3906	0.0505	9.90	0.1847
DTLZ7	5	Clustering	0.8075	0.4237	0.0687	9.23	0.2007
DTLZ7	10	Full objectives	2.3204	0.0389	0.2777	275	–
DTLZ7	10	CI linear	0.0026	0.8232	0.0681	275	0.6174
DTLZ7	10	CI nonlinear	0.9287	0.3550	0.0749	133.07	0.2357
DTLZ7	10	PCA	0.0025	0.8289	0.0671	275	0.6187
DTLZ7	10	mRMR	0.9185	0.3384	0.0751	137.33	0.2423
DTLZ7	10	Clustering	0.9265	0.3269	0.0782	112.70	0.2217
WFG9	3	Full objectives	0.7223	0.0292	0.0345	105	–
WFG9	3	CI linear	0.1152	0.3384	1.506×10^{-4}	105	0.5139
WFG9	3	CI nonlinear	0.1152	0.3384	1.506×10^{-4}	105	0.5139
WFG9	3	PCA	0.4246	0.1297	0.0030	105	0.1742
WFG9	3	mRMR	0.5013	0.1847	0.0093	105	0.1439
WFG9	3	Clustering	0.4246	0.1297	0.0030	105	0.1742
WFG9	5	Full objectives	1.1342	0.0484	0.0577	210	–
WFG9	5	CI linear	0.8814	0.1105	0.0507	210	−0.0522
WFG9	5	CI nonlinear	0.1362	0.5472	0.0020	210	0.1925
WFG9	5	PCA	0.8870	0.1055	0.0521	210	−0.0601
WFG9	5	mRMR	0.8870	0.1055	0.0521	210	−0.0601
WFG9	5	Clustering	0.5103	0.2974	0.0063	210	0.0747
WFG9	10	Full objectives	2.0740	0.0850	0.0874	275	–
WFG9	10	CI linear	1.6943	0.1511	0.0921	275	−0.0435
WFG9	10	CI nonlinear	0.2131	0.5778	0.0042	275	0.0276
WFG9	10	PCA	1.7823	0.1360	0.1009	275	−0.0597
WFG9	10	mRMR	1.8625	0.1526	0.0941	275	−0.0713
WFG9	10	Clustering	0.7916	0.3387	0.0109	275	−0.0038

Table 8. Paired Wilcoxon signed-rank comparisons of CI linear with PCA, mRMR and clustering (Holm-adjusted *p*-values). Rows for DTLZ2 with m = 5 and 10 compare the CI linear subsets of the complete procedure (new runs); rows of the non-applicable configurations (DTLZ3 with m = 10, DTLZ7 and WFG9) compare historical subsets and are exploratory.

Problem	m	Baseline	HV	IGD+	Spacing	ND
DTLZ2	3	PCA	B < 0.001	CI < 0.001	NS 0.625 (6)	CI < 0.001
DTLZ2	3	mRMR	=	=	=	=
DTLZ2	3	Clustering	NS 0.760	NS 0.442	NS 0.502	=
DTLZ2	5	PCA	NS 0.073	CI < 0.001	CI < 0.001	CI < 0.001
DTLZ2	5	mRMR	B < 0.001	CI < 0.001	CI < 0.001	CI < 0.001

Table 8. *Cont.*

Problem	m	Baseline	HV	IGD+	Spacing	ND
DTLZ2	5	Clustering	NS 0.824	NS 1.000	NS 0.119	=
DTLZ2	10	PCA	B < 0.001	NS 0.109	CI < 0.001	NS 0.083
DTLZ2	10	mRMR	B < 0.001	NS 1.000	CI < 0.001	NS 0.136
DTLZ2	10	Clustering	B < 0.001	B 0.003	CI < 0.001	NS 0.317
DTLZ3	3	PCA	B < 0.001	B 0.004	NS 1.000 (2)	CI < 0.001
DTLZ3	3	mRMR	B < 0.001	B 0.004	NS 1.000 (2)	CI < 0.001
DTLZ3	3	Clustering	B < 0.001	B 0.004	NS 1.000 (2)	CI < 0.001
DTLZ3	5	PCA	CI < 0.001	CI < 0.001	CI < 0.001 (29)	CI < 0.001
DTLZ3	5	mRMR	CI < 0.001	CI < 0.001	CI < 0.001 (29)	CI < 0.001
DTLZ3	5	Clustering	B < 0.001	B < 0.001	CI < 0.001	=
DTLZ3	10	PCA	NS 1.000	B < 0.001	NS 1.000	NS 0.719
DTLZ3	10	mRMR	NS 1.000	B < 0.001	NS 0.236	NS 0.952
DTLZ3	10	Clustering	NS 1.000	B < 0.001	CI 0.037	NS 0.952
DTLZ7	3	PCA	CI < 0.001	B < 0.001	NS 0.547	=
DTLZ7	3	mRMR	B < 0.001	CI < 0.001	NA (0)	CI < 0.001
DTLZ7	3	Clustering	B < 0.001	CI < 0.001	NA (0)	CI < 0.001
DTLZ7	5	PCA	NS 1.000	NS 0.189	NS 0.547	=
DTLZ7	5	mRMR	B < 0.001	NS 0.792	CI < 0.001	CI < 0.001
DTLZ7	5	Clustering	B < 0.001	NS 0.524	CI < 0.001	CI < 0.001
DTLZ7	10	PCA	NS 1.000	NS 0.189	NS 0.552	=
DTLZ7	10	mRMR	B < 0.001	B < 0.001	CI 0.033	CI < 0.001
DTLZ7	10	Clustering	B < 0.001	B < 0.001	CI 0.033	CI < 0.001
WFG9	3	PCA	B < 0.001	B < 0.001	CI < 0.001	=
WFG9	3	mRMR	B < 0.001	B < 0.001	CI < 0.001	=
WFG9	3	Clustering	B < 0.001	B < 0.001	CI < 0.001	=
WFG9	5	PCA	NS 1.000	NS 1.000	NS 1.000	=
WFG9	5	mRMR	NS 1.000	NS 1.000	NS 1.000	=
WFG9	5	Clustering	CI < 0.001	CI < 0.001	B < 0.001	=
WFG9	10	PCA	B 0.039	B 0.012	NS 0.517	=
WFG9	10	mRMR	B < 0.001	NS 1.000	NS 1.000	=
WFG9	10	Clustering	CI < 0.001	CI < 0.001	B < 0.001	=

4.4.3. Full-Space Metrics and Validation Results

The final population of each run was evaluated in the original objective space using complementary indicators. Hypervolume (HV) measures the dominated volume in the normalized full objective space and is maximized. Exact HV was used where feasible, while the ten-objective cases used Monte Carlo HV estimation. The IGD+ indicator measures the distance to an empirical full-space non-dominated reference set obtained by the union of the compared final populations and is minimized. Spacing measures the regularity of point distribution and is also minimized. The number of non-dominated points in the original objective space was retained as ND. Before evaluation, the final populations of all methods and runs of a configuration were pooled, and each objective was min-max

normalized with the bounds of this union. HV was computed in this normalized space with the reference point 1.1 in every objective, exactly for m ≤ 5 and by Monte Carlo estimation with 100,000 uniform samples for m = 10. The IGD+ reference set is the non-dominated subset of the union, randomly subsampled to at most 20,000 points. Spacing is the standard deviation of the nearest-neighbor Euclidean distances, and the removed-objective loss is computed against the full-objective run with the same seed. Because the normalization bounds and the IGD+ reference depend on which populations are pooled, absolute values change when a formulation is added: in DTLZ2 with five and ten objectives, adding the new CI linear runs changes the IGD+ of the full formulation from 0.0060 to 0.0045 and from 0.0313 to 0.0272. The indicators therefore compare formulations within a configuration only.

In addition, a removed-objective loss indicator was computed for reduced formulations. For a subset S, let $\overline{S}$ denote the set of objectives not optimized directly. Let P_S be the final population obtained by optimizing F_S, and let P_{full} be the final population of the full-objective run with the same seed. After normalizing each objective in the full objective space, the removed-objective loss is computed as:

$$L_{rem}(S) = \frac{1}{|\overline{S}|}\sum_{j\in\overline{S}}\left[\frac{1}{|P_S|}\sum_{x\in P_S}\tilde{f}_j(x) - \frac{1}{|P_{full}|}\sum_{x\in P_{full}}\tilde{f}_j(x)\right] \quad (15)$$

where $\overline{S}$ is the set of removed objectives, $\tilde{f}_j$ is the normalized value of objective j, P_S is the final population obtained with the reduced objective subset, and P_{full} is the corresponding full-objective population.

Lower values indicate better preservation of removed objectives. Negative values mean that, on average, the reduced formulation obtained better values in the removed objectives than the corresponding full-objective reference for that comparison.

Because the full formulation directly optimizes the objective space used for evaluation, it is expected to provide the strongest HV and IGD+ reference in most cases; the informative comparison is between the reduced formulations.

Table 7 reports means over thirty runs. Because several inter-method margins are narrow, the mean differences are tested below with paired nonparametric tests rather than ranked by their means.

To assess whether the mean differences in Table 7 were systematic, CI linear was compared with PCA, mRMR-like selection and clustering by two-sided paired Wilcoxon signed-rank tests [25]. The 30 runs were matched by seed within each problem and objective dimensionality, and tests were performed separately for HV, IGD+, spacing and ND. Holm correction [26] was applied within each problem–metric family across the three dimensionalities and three baselines (nine planned comparisons). Undefined spacing values were omitted pairwise; comparisons with no valid pairs were not tested, and identical paired results were reported without a test. For DTLZ2, whose five- and ten-objective configurations were recomputed with the CI linear subsets of the complete procedure, the correction was applied within the six recomputed comparisons of each metric, and the three-objective comparisons kept their original adjustment.

Of the 144 planned comparisons, 81 were significant after Holm correction (44 favoring CI linear and 37 favoring the baselines), 42 were not significant, 19 were identical, and two had no valid spacing pairs. In the five admitted configurations (60 comparisons), 23 favored CI linear, 14 favored the baselines, 16 were not significant, and 7 were identical; in the seven exploratory configurations (84 comparisons), the counts were 21, 23, 26 and 12, together with the two untested cases. Among the significant HV and IGD+ comparisons, the baselines prevailed more often, while CI linear showed more favorable spacing and ND results. Because low spacing can coincide with a collapsed non-dominated front, these indicators must be interpreted jointly.

Dispersion and paired effect sizes were computed for the recomputed DTLZ2 configurations. The standard deviation of HV over the 30 runs is below 0.004 for the full formulation and for CI linear, and reaches 0.17 for PCA ($m = 5$) and 0.41 for CI nonlinear ($m = 10$). The Vargha–Delaney statistic $\hat{A}_{12}$ [27], the probability that CI linear is better in a randomly chosen pair of runs, is 0.40, 0.23 and 0.45 for HV against PCA, mRMR-like selection and clustering with five objectives, and 0.93, 0.97 and 0.45 for IGD+; with ten objectives it is 0.00 for HV against all three baselines and 0.80, 0.57 and 0.37 for IGD+.

Table 8 presents Holm-adjusted p-values. CI and B identify significant differences ($p < 0.05$) favoring CI linear and the baseline, respectively; NS denotes a non-significant comparison, = identical paired results, and NA no valid pairs. For spacing, sample sizes below 30 appear in parentheses; the two NA cases concern DTLZ7 ($m = 3$) against mRMR and clustering. HV and ND are maximized, whereas IGD+ and spacing are minimized.

The number of non-dominated points in the original objective space requires separate comment, because the removed-objective loss of Equation (15) does not capture it. That indicator compares mean values of the removed objectives, so a reduced formulation can preserve those means while collapsing the front onto a small region, and this is exactly what occurs in several configurations. For DTLZ2 with three objectives, the CI nonlinear and PCA formulations attain the best removed-objective loss of the comparison, 0.3049, with only 6.77 non-dominated points out of 105; for DTLZ3 with three objectives, four of the six formulations retain 1.70 points; for DTLZ7 with three objectives, the count falls to 1.27. In these cases, the reduced problem has become effectively single-objective in the retained space, and the favorable loss value is an artifact of averaging over a degenerate set. A different form of collapse affects DTLZ2 with five objectives: the CI linear subset of the complete procedure and the clustering subset both concentrate their final populations where the retained objectives are close to zero, with identical HV and IGD+ (0.9122 and 0.1938), a spacing of about 3×10^{-4} and only about 76 distinct points among the 210 counted as non-dominated; the CI linear subsets of DTLZ2 with ten objectives behave in the same way. DTLZ2 has no redundant objectives, since its Pareto front is an $(m - 1)$-dimensional spherical segment, so every reduction removes trade-off structure, and these rows show that a large ND count can coexist with a degenerate front.

Two consequences follow, and both are adopted here. First, HV, IGD+, spacing and removed-objective loss must be read jointly with ND, and a formulation whose ND falls to single digits is not reported as competitive on any indicator. Second, the collapse is itself informative about objective selection: it occurs where the retained subset no longer spans the trade-off structure of the full problem, that is, where the reduction from m to k was too aggressive for the problem geometry, rather than where one particular selector failed. It affects the proposed method and the baselines alike, as the same rows of Table 7 show. The convergence-diversity trade-off is therefore not a side issue for objective reduction but one of its principal costs, and Section 4.7 treats the choice of k as a decision that must be justified per problem rather than fixed by convention.

4.4.4. Stochastic Stability of the Nonlinear Conditioning Selector

The active optimization validation includes both CI linear and CI nonlinear objective subsets. The present reseeding analysis focuses only on the nonlinear conditioning selector. In contrast with a Table 5-fixed comparison, the reference subset used here is the historical CI nonlinear subset, also reported in Table 5, that was used in the active optimization validation. The aim is therefore to test whether the stochastic nonlinear selector recovers its own original selected subset when the Extra Trees estimator and the cross-validation partition are independently re-seeded.

For each of the 12 benchmark cases, the full nonlinear selection pipeline was repeated 30 times. Each repetition used an independently re-seeded Extra Trees estimator and an independently re-seeded five-fold cross-validation partition. The sample of 10,000 solutions was kept fixed so that the analysis isolates the randomness introduced by the estimator and by the CV partition. The nonlinear model used 100 trees, maximum depth 8 and minimum samples per leaf 20. The adequacy threshold for OOF R^2 is the gate value of 0.50 of Section 3.4. The re-seeding study used its own sample of 10,000 solutions (sample seed 2026) and the selector configuration, whereas Table 5 reports the more flexible diagnostic configuration on the original sample, so the R^2 values of Table 5 and of the re-seeding study are not directly comparable. For the selector configuration, the gap between training and out-of-fold R^2 does not exceed 0.037 for DTLZ2, DTLZ3 and DTLZ7 and is 0.11–0.12 for WFG9.

Across the 12 benchmark cases, the mean set-overlap with the original CI nonlinear subsets is 91.9% (Table 9). When the average is weighted by the subset size k, the overlap increases to 93.5%. Exact-set recovery is obtained in 78.3% of the 360 re-seeded runs. These results show that the nonlinear conditioning selector is highly stable with respect to the estimator and cross-validation re-seeding when compared against its own original selected subsets. Because the reference subsets were obtained on a different sample and with a different configuration, these figures measure the combined effect of sample, configuration and seed. Stability with respect to the seed alone is higher: within the re-seeding study, the most frequent subset is recovered in 93.1% of the runs on average, and all 30 runs return the same subset in eight of the twelve cases.

Table 9. Exact-set recovery and overlap of the original CI nonlinear subsets under nonlinear re-seeding. Exact-set recovery compares each re-seeded objective set with the original CI nonlinear set, disregarding the order in which objectives were selected. Overlap is the fraction of selected objectives shared with the original CI nonlinear subset.

Problem	m	k	Original CI Nonlinear Subset	Runs	Exact Set	Overlap
DTLZ2	3	2	(f3, f2)	30	0.0%	50.0%
DTLZ2	5	3	(f2, f5, f4)	30	80.0%	93.3 ± 13.6%
DTLZ2	10	5	(f2, f10, f1, f9, f3)	30	100.0%	100.0%
DTLZ3	3	2	(f3, f2)	30	100.0%	100.0%
DTLZ3	5	3	(f5, f4, f3)	30	100.0%	100.0%
DTLZ3	10	5	(f1, f10, f9, f2, f8)	30	100.0%	100.0%
DTLZ7	3	2	(f2, f1)	30	100.0%	100.0%
DTLZ7	5	3	(f2, f3, f1)	30	3.3%	67.8 ± 6.1%
DTLZ7	10	5	(f5, f8, f9, f3, f4)	30	56.7%	91.3 ± 10.1%
WFG9	3	2	(f1, f2)	30	100.0%	100.0%
WFG9	5	3	(f4, f3, f2)	30	100.0%	100.0%
WFG9	10	5	(f9, f8, f7, f2, f6)	30	100.0%	100.0%

The few deviations are concentrated in DTLZ2 with three objectives and DTLZ7 with five and ten objectives. DTLZ2 with three objectives always retains f3 but replaces f2 with f1. DTLZ7 with five objectives is the least stable case relative to the original nonlinear subset: the selector usually recovers f3 and f1, but f2 is recovered only occasionally. DTLZ7 with ten objectives remains highly overlapping, but the fifth selected objective oscillates in some repetitions. Overall, these deviations are limited and structured rather than random across the objective space. The out-of-fold R^2 values and decision margins of the re-seeded nonlinear selector are reported in Table 10.

Table 10. Out-of-fold R^2, residual dependence and decision margin under nonlinear re-seeding. The decision margin is computed relative to the threshold $R^2 = 0.50$; positive values are above the threshold and negative values are below it.

Problem	m	Mean OOF R^2	SD	Min OOF R^2	Residual Dep.	Mean Margin	Min Margin
DTLZ2	3	0.9526	0.00014	0.9522	0.6133	0.4526	0.4522
DTLZ2	5	0.9568	0.00025	0.9564	0.4609	0.4568	0.4564
DTLZ2	10	0.8653	0.00069	0.8642	0.2249	0.3653	0.3642
DTLZ3	3	0.8353	0.00016	0.8350	0.5920	0.3353	0.3350
DTLZ3	5	0.8306	0.00029	0.8298	0.5131	0.3306	0.3298
DTLZ3	10	0.4949	0.00100	0.4928	0.4085	−0.0051	−0.0072
DTLZ7	3	0.9260	0.00030	0.9254	0.1180	0.4260	0.4254
DTLZ7	5	0.9744	0.00010	0.9742	0.0546	0.4744	0.4742
DTLZ7	10	0.9996	0.00000	0.9996	0.0332	0.4996	0.4996
WFG9	3	0.1592	0.00119	0.1568	0.4435	−0.3408	−0.3432
WFG9	5	0.1448	0.00096	0.1418	0.3040	−0.3552	−0.3582
WFG9	10	0.1193	0.00066	0.1179	0.2992	−0.3807	−0.3821

The OOF R^2 values show that re-seeding has very little effect on the fitted nonlinear model estimates: the standard deviations across seeds are very small in all cases. However, the adequacy of the nonlinear conditioning model is problem-dependent. DTLZ2, DTLZ7 and the lower-dimensional DTLZ3 cases remain above the $R^2 = 0.50$ threshold, whereas DTLZ3 with ten objectives and all WFG9 cases are below this threshold. Thus, the nonlinear selector is stable with respect to re-seeding, but its predictive adequacy is not uniform across benchmark families.

A caveat must be attached to the WFG9 rows of Table 10, because the natural reading of them is the wrong one. With a mean out-of-fold R^2 between 0.119 and 0.159 the nonlinear conditioning model explains almost none of the objective variance, so the residuals are nearly equal to the objectives and the conditioning step is close to vacuous. The very small standard deviations across seeds, and the resulting 100% exact-set recovery for all three WFG9 configurations, therefore reflect a model that fails to learn consistently rather than a selector that is robust for the right reason: a model that predicts nothing predicts nothing reproducibly, and the selection it supports is close to the marginal selection it was intended to improve upon. Under the adequacy gate of Section 3.4 all three WFG9 configurations fall below the threshold of 0.50 and are reported as cases in which conditioning-based reduction is not justified; the same applies to DTLZ3 with ten objectives, whose mean out-of-fold R^2 is 0.495. The stability results that carry meaning are those for DTLZ2, DTLZ7 and the lower-dimensional DTLZ3 cases, where the conditioning model is adequate and the recovery rates therefore describe the selector rather than the failure of its input.

4.4.5. Discussion

No reduced method dominates all problems, objective dimensionalities and metrics. Objective selection necessarily involves a trade-off between dimensionality reduction, interpretability, convergence, diversity, and preservation of objectives that are not optimized directly. PCA, mRMR, clustering, and conditioning-based selection emphasize different aspects of the objective structure, and the validation shows that these differences translate into different optimization behavior.

For DTLZ2, classical or marginal-dependence methods are often more competitive in the smoother benchmark structure. DTLZ3 exhibits more favorable behavior for

conditioning-based selection, especially in the five-objective case (the ten-objective case is non-applicable and only exploratory), where conditioning-based variants are strong in spacing and removed-objective preservation. DTLZ7 highlights the dependence of conclusions on the chosen metric: CI linear often gives very small spacing values, while CI nonlinear, clustering and mRMR obtain stronger HV and removed-objective preservation in several configurations. WFG9, for which the procedure is non-applicable, shows that conditioning-based analysis is useful as a complementary view of objective structure, but also that nonlinear conditioning does not automatically translate into better full-space optimization performance.

Overall, the proposed method should not be presented as a universally superior replacement for PCA, mRMR or clustering. Its contribution is complementary: it provides an interpretable objective-selection criterion based on residual redundancy after removing the effects of shared decision variables. The validation shows mixed results rather than a consistent advantage: the conditioned subsets tend to obtain lower spacing and, in several configurations, more non-dominated points, whereas the baselines more often obtain better HV and IGD+ and, in several cases, preserve the full objective space more effectively.

The nonlinear conditioning variant should be interpreted as a robustness diagnostic rather than as the final recommended selector. Higher modeling flexibility does not automatically lead to better reduced-objective optimization, as shown especially by WFG9. Nevertheless, the re-seeding analysis shows that the nonlinear selector is not materially unstable with respect to random seeds. Remaining differences should be interpreted as local model-sensitivity effects in a small number of benchmark cases, rather than as broad stochastic instability of the nonlinear pipeline.

4.4.6. Controlled Example and Matched Comparison

The criterion is intended to preserve the Pareto-dominance relation of the full problem as far as possible with k objectives. An objective that is a strictly increasing function of the retained ones can be removed without changing which solutions dominate which, whereas removing an objective that conflicts with the retained ones creates false dominance relations and loses non-dominated solutions. These two properties were measured on a controlled example of known structure, with seven decision variables uniform on $[0, 1]$, 10,000 samples and eight minimized objectives. The objective $f2 = \exp(2f1)$ is exactly redundant with f1; f3 and f4 conflict marginally ($r = -0.88$); f5 and f6 share x3 but conflict through x4; and f7 and f8 share the component $2(x1 + x2)$ but have opposite residual terms $\pm 0.3 \sin(2\pi x7)$, so that they are positively correlated marginally but in pure conflict after conditioning (residual correlation -1). With $k = 7$ the correct reduction removes f2 or f1. For each selector, the false-dominance rate (the fraction of dominance relations in the reduced space that do not hold in the full space, over 2000 solutions) and the recall of the full non-dominated set were computed (Table 11).

Table 11. Controlled example (eight objectives, $k = 7$): objective removed by each selector, false-dominance rate in the reduced space and recall of the full non-dominated set.

Selector	Removed Objective	Correct	False-Dominance Rate	ND Recall
CI, absolute residual correlation (standardized Equation (10))	f8	No	0.215	0.835
CI, positive-only residual correlation	f2	Yes	0.000	1.000
Same greedy score, raw |r| (matched)	f2	Yes	0.000	1.000
mRMR-like	f8	No	0.215	0.835
PCA	f1	Yes	0.000	1.000
Clustering	f2	Yes	0.000	1.000

The example exposes a limitation of the absolute form of Equation (6). Because the absolute value treats the residual conflict between f7 and f8 as redundancy, the conditioned selector removes f8, and so does the mRMR-like baseline, which penalizes absolute marginal correlation; both create false dominance in 21.5% of the reduced-space relations and lose 16.5% of the non-dominated set. The sign-aware variant of Section 3.3, which penalizes only positive residual association, removes f2 and preserves all dominance relations. Conditioning is not the cause of the failure: the f7–f8 conflict is visible only after conditioning, and it is the sign of the residual correlation that identifies it as a trade-off. The sign-aware form is therefore recommended for minimization problems (Section 3.3). On the benchmarks it returns the same subsets as the absolute form for DTLZ2 with three objectives and DTLZ3 with three and five objectives, and different ones for DTLZ2 with five (f3, f4, f1) and ten objectives (f8, f9, f2, f7, f6); its injection-molding subsets are reported in Section 4.6.

The recommendation was tested on three families of controlled scenarios, each with ten independent samples of 5000 solutions and three amplitudes $a \in \{0.1, 0.3, 0.6\}$ of the residual-conflict term: (A) the example above; (B) eleven objectives with two residual-conflict pairs and two exactly redundant pairs ($k = 9$); and (C) six objectives without residual conflict ($k = 5$). The sign-aware variant removed only redundant objectives in all 90 instances, with no false dominance and full recall of the non-dominated set. The absolute form failed in every instance of A and B, with false-dominance rates that increase with the amplitude from 0.08 to 0.40 in A and from 0.10 to 0.49 in B, and it was correct in every instance of C, where it coincides with the sign-aware variant. The mRMR-like baseline failed in all instances of A and B and in half of those of C. The matched raw selector, PCA and clustering were correct for $a \geq 0.3$ (PCA in 8 of 10 instances of A at $a = 0.3$), but all three failed at $a = 0.1$, where the marginal correlation of the conflicting pair is close to 1 and the conflict becomes visible only after conditioning. Conditioning combined with sign-aware redundancy was therefore the only selector that was correct in every scenario.

Because the active validation of Section 4.4 uses the absolute form, the sign-aware form was also validated in the two configurations in which the forms select different subsets, DTLZ2 with five (f3, f4, f1) and ten objectives (f8, f9, f2, f7, f6), with 30 NSGA-III runs each and the settings of Section 4.4.1 (Supplementary Materials S2, Tables S4 and S5). The full-space behaviour of the sign-aware subsets is indistinguishable from that of the absolute-form subsets: HV and IGD+ coincide to four decimals (0.9122 and 0.1634 for $m = 5$; 0.4463 and 0.5373 for $m = 10$, with the pooled normalization of Supplementary Materials S2), the non-dominated counts are equal, and all paired comparisons between the two forms are non-significant or identical. Against the baselines, the sign-aware subsets reproduce the mixed pattern of Table 8: better IGD+, spacing and ND than PCA and mRMR-like selection for $m = 5$, and lower HV than all three baselines for $m = 10$. In the injection-molding case, the sign-aware form keeps two of the three and four of the five selected objectives and the decisive role of the unexplained-variability term, but it is less stable with respect to the cross-validation partition (15 of 30 exact recoveries for both subset sizes, against 30 and 22 with the absolute form; Supplementary Materials S2, Tables S6–S9). The choice between the two forms therefore does not alter the conclusions of Section 4.4, while the controlled scenarios show that the sign-aware form is the safer choice when residual conflicts are present.

The baselines differ from the proposed selector in the dependence measure, the score and the search, so the differences in Table 6 cannot be attributed to conditioning alone. A matched comparison was therefore made in which the standardized greedy score of Equation (10), the gate, the floor and all parameters are kept, and only the residual correlation is replaced by the absolute correlation of the raw objectives. On the admitted benchmark configurations, the matched raw selector returns the same subsets as the conditioned one

for DTLZ2 and DTLZ3 with three objectives, and different ones for DTLZ2 with five (f4, f1, f3) and ten objectives (f9, f2, f8, f7, f3) and for DTLZ3 with five objectives (f4, f2, f3). In the injection-molding case, the matched raw selector returns {coolef, aDTsm, sDTsmH} for k = 3, with one objective in common with the conditioned subset, and shares one of five objectives for k = 5; when the unexplained-variability term is also removed ($\beta = 0$ with raw correlation), the k = 3 subset {coolef, maxWaZ, sDTsmH} has no objective in common with it. Conditioning alone therefore changes the subset in three of the five admitted benchmark configurations and in the injection-molding case, and the separate effect of the unexplained-variability term is quantified by the ablation of Section 4.6.

4.5. Application to Injection Molding Design

The proposed conditioning-based objective selection methodology was applied to a real engineering problem involving the design of an injection molding process. The dataset, obtained from a design of experiments (DOE) with 400 samples, includes seven decision variables (in which one, gate, is categorical with values 1 or 2, and the remaining are real values) and a large set of objectives describing thermal behavior, shrinkage, warpage, stress, cooling efficiency, and energy consumption. This problem is particularly well-suited to evaluating the proposed approach due to the strong coupling between objectives and the presence of well-understood physical relationships that govern the process [18]. For details on the problem definition and the identification of decision variables and objectives, the reader is referred to [18]. Reference [18] studied the surrogate-assisted optimization pipeline for this problem, the effect of ensemble objective selection and the effect of sampling density; the present work reuses its design of experiments but introduces the conditioned redundancy criterion, the admissibility gate and the systematic benchmark validation, none of which appear in [18].

To make the case study self-contained rather than dependent on [18], the candidate objectives are defined in Table S1 of the Supplementary Materials. The naming convention is as follows: a leading a or s denotes the average or the standard deviation of a distribution evaluated over the part; DT denotes a temperature difference, with the suffixes sm, ms and ss identifying the pair of locations compared; a trailing H restricts the statistic to the 20% of locations with the largest temperature variation; Wa and Sr denote warpage and shrinkage, with T for the total and X, Y and Z for the directional components; and VMS denotes von Mises stress. The seven decision variables are the gate configuration, which is categorical with two levels, and six continuous cooling-channel geometry parameters. All objectives are expressed in minimization form before selection, so a response to be maximised is replaced by its negative; this applies in particular to the cooling-efficiency indicator.

For the pooled analysis, gate was encoded as a categorical variable. For the interaction analysis, interaction terms between gate and the continuous decision variables were added to the regression model. For the regime-specific analysis, separate models were fitted for gate = 1 and gate = 2.

The regression models used for conditioning were validated by five-fold cross-validation with shuffled folds (random_state = 0), because the rows of the design of experiments are stored in a structured order and unshuffled folds would place whole design regions in the test set. Over the full set of 39 objectives (the thirty-five candidate indicators and the four remaining energy objectives), the linear model achieved a mean training R^2 of 0.926 and a mean out-of-fold R^2 of 0.922, a gap of 0.004. The nonlinear conditioner, an Extra Trees ensemble fitted separately to each objective with 100 trees, a maximum depth of 8 and at least 20 samples per leaf, achieved 0.898 and 0.884, a gap of 0.014. The linear model therefore generalizes better and was retained as the main conditioning model. Table 12

reports the distribution of the out-of-fold R^2 over the thirty-five candidate indicators that remain after the Stage 1 energy filtering described below.

Table 12. Distribution of the per-objective out-of-fold R^2 over the thirty-five candidate indicators of the injection-molding case, five-fold cross-validation. The mean is taken over the thirty-five candidates remaining after Stage 1 filtering and is therefore not directly comparable with the aggregate quoted in the text above, which is computed before that filtering.

Statistic of the Out-Of-Fold R^2 over the 35 Candidate Indicators	Linear Conditioner	Nonlinear Conditioner
Mean	0.929	0.879
Minimum	0.777	0.699
First quartile	0.881	0.791
Median	0.959	0.879
Third quartile	0.986	0.987
Maximum	0.999	0.999
Indicators with out-of-fold $R^2 < 0.50$	0	0
Indicators with out-of-fold $R^2 < 0.80$	4	9
Indicators flagged conditioning-degenerate ($U_i < 0.01$)	7	—

Because fitting many objectives from 400 samples could raise concerns about overfitting, the per-objective distribution is reported rather than the aggregate alone. The relevant point is that model dimensionality does not grow with the number of objectives: a separate univariate-response regression is fitted for each objective on the same seven decision variables, so the sample-to-parameter ratio is approximately fifty to one for the linear conditioner regardless of how many objectives are analyzed. Table 12 reports the distribution of the out-of-fold R^2 over the thirty-five candidate indicators, together with the degeneracy diagnostic of Section 3.4. No indicator has a negative out-of-fold R^2 and none falls below 0.50, so the adequacy gate is satisfied for every objective in this case study. Four indicators lie between 0.77 and 0.80—sDTsmH, sDTmsH, maxVMS and aDTsm, all of them dispersion measures—and these are the objectives whose residual dependence should be interpreted with most caution. At the other extreme, seven indicators have U_i below the floor of 0.01 and are flagged as conditioning-degenerate: minSrT, maxSrT, maxWaY, minWaT, minWaY, minSrZ and maxSrY. These are precisely the warpage and shrinkage components discussed below, so there is an independent and purely statistical reason not to draw strong conclusions from their residual correlations.

Table 13 reports the dynamic ranges of the three components of the selection score on the same candidate set, which are the quantities referred to in Sections 3.5 and 3.6. The relevance term is discriminative but has the narrowest range of the three; the individual pairwise residual redundancies span almost the whole unit interval. This is the empirical basis for the standardization of Equation (10) introduced in Section 3.6.

Table 13. Dynamic ranges of the three components of the selection score on the injection-molding candidate set, under the linear conditioner and min-max normalized objectives.

Score Component	Minimum	Maximum	Range	SD over Objectives
$Rel_i = \mathrm{Var}(f_i)$, min-max normalized objectives	0.022	0.198	0.176	0.060
$U_i = 1 - R^2_i$	0.001	0.223	0.222	0.072
CR_{ij}, individual pairwise values	0.000	0.993	0.993	0.200
$Red_i(S)$ averaged over all other objectives	0.072	0.283	0.211	0.044

The analysis begins with the dependence structure of the full set of 39 objectives. The mean absolute pairwise dependence is 0.557 in the original objective space; after conditioning on the decision variables it decreases to 0.204, a reduction of approximately 63% (Table 14). A substantial portion of the observed dependence is therefore induced by shared decision variables, although the remaining dependence is not negligible.

Table 14. Dependence reduction in the injection-molding case: full objective set, Stage 1 and Stage 2 (linear conditioner, five-fold shuffled cross-validation).

Objective Set	Raw Dependence	Conditioned Dependence	Reduction
Full objective set (39 objectives), before Stage 1 filtering	0.557	0.204	63.4%
Stage 1: 35 objectives (28 admissible)	0.589	0.219	62.8%
Stage 2: 14 objectives selected from the 28 admissible	0.668	0.190	71.6%

These findings are consistent with the physical behavior of the injection molding process, where multiple performance indicators are influenced by the same process parameters. At the same time, there are physically expected relationships between groups of objectives. In particular, thermal behavior governs the cooling dynamics of the material, which determines shrinkage and ultimately leads to warpage. To assess whether this physically expected chain can be identified from the data, a conditioning-based incremental analysis was performed considering the sequence thermal → shrinkage → warpage. Each group (thermal, shrinkage and warpage) is summarized by the first principal component of its standardized indicators, which explains 61%, 82% and 99% of the group variance, respectively, and ΔR^2 is the increase in the in-sample R^2 of a linear model on the decision variables when the upstream composite is added.

The incremental contributions of thermal variables to shrinkage ($\Delta R^2 = 0.000110$) and of shrinkage to warpage ($\Delta R^2 = 0.000001$) are extremely small. Similarly, the direct contribution of thermal variables to warpage is negligible ($\Delta R^2 = 0.000003$). These increments are computed in-sample, with each block of indicators represented by its first principal component (12 thermal, 8 shrinkage and 8 warpage indicators, explaining 61%, 82% and 99% of the block variance). To assess whether they differ from zero, they were also computed out-of-fold (five folds) and tested with the nested-model F-test, and a percentile bootstrap interval (1000 resamples) was obtained for the in-sample increment. In the pooled data, the thermal → shrinkage increment is detectable but negligible in size (out-of-fold $\Delta R^2 = 1.1 \times 10^{-4}$; bootstrap interval $[2.4 \times 10^{-5}, 2.6 \times 10^{-4}]$; $p < 0.001$), whereas the shrinkage → warpage and thermal → warpage increments are not distinguishable from zero (out-of-fold $\Delta R^2 < 0$; $p = 0.73$ and 0.45). The analysis is descriptive: it shows that, once the decision variables are taken into account, the composites add almost no explanatory power for one another in the pooled data, and it neither establishes nor excludes a physical pathway. This behavior is illustrated in Figure 5, where the physically constrained dependency structure is shown together with the corresponding incremental explanatory power. The figure highlights that, although the direction of influence is physically meaningful, the strength of these relationships is dominated by the shared effect of the decision variables.

This observation has important implications for objective selection. Although shrinkage and warpage are physically meaningful outputs, their variability is largely determined by the design variables and by upstream thermal effects. In this dataset, therefore, the many shrinkage and warpage components carry limited additional independent information once thermal descriptors are included, and seven of them are conditioning-degenerate. This explains why the conditioning-based method consistently retains thermal and stress de-

scriptors, while the shrinkage and warpage family is represented by one or two components rather than by many.

To further investigate this behavior, a two-stage objective reduction strategy was adopted. In the first stage, energy-related objectives were filtered to avoid dominance effects, retaining only the total energy consumption indicator (energT) because of its relevance for sustainability. After this filtering 35 objectives remain, and conditioning reduces their mean absolute dependence from 0.589 to 0.219 (Table 14). Seven of the 35 fall below the residual-variance floor of Section 3.4, so 28 objectives are admissible candidates for the sequential selection.

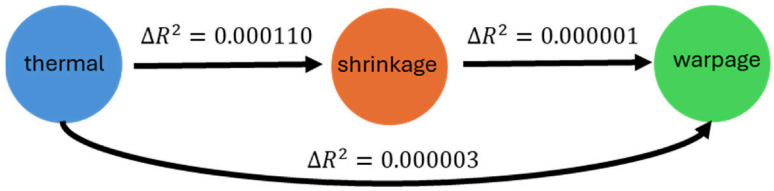


Figure 5. Physically constrained dependency structure between thermal behavior, shrinkage, and warpage. Edge labels represent the incremental explanatory contribution (ΔR^2) after conditioning on decision variables.

In the second stage, the sequential procedure was applied to the 28 admissible candidates with $k = 14$, that is, to half of them, giving {energT, aDTsm, maxVMS, maxWaZ, sDTssH, maxSrZ, minWaZ, sDTmsH, sdDens, maxWaX, minSrY, aDTms, maxWaT, aDTmsH}; its first three and first five elements are the subsets obtained with the full criterion in Section 4.6. The raw dependence of this subset (0.668) is higher than that of the full set, while its conditioned dependence (0.190) is lower (Table 14). This is the expected signature of the criterion: the selector does not penalize objectives that are marginally correlated through the shared decision variables, only those that remain correlated after conditioning.

While the selection obtained at this stage is statistically sound, it does not necessarily reflect engineering priorities. In practice, certain objectives must be retained regardless of their statistical properties. In the injection molding context, cycle time (tcycle) is a critical productivity indicator, and energy consumption (energT) is essential from a sustainability perspective. To account for this, engineering constraints were incorporated into the selection process.

When tcycle is enforced as a mandatory objective alongside the anchor energT, the procedure selects {tcycle, energT, aDTsm} for three objectives and adds the maximum von Mises stress, maxVMS, and the shrinkage component minSrY for five. The five-objective set already contains one representative of each physical category—productivity, sustainability, thermal behavior, mechanical performance and dimensional stability—without any further constraint being imposed, and it is adopted as the final objective set (Table 15).

Table 15. Final category-constrained objective set.

Category	Objective
Productivity	tcycle
Sustainability	energT
Thermal	aDTsm
Mechanical	maxVMS
Dimensional (shrinkage)	minSrY

The final set should be read together with its anchor-sensitivity range (Section 4.6). Over the 28 possible forced anchors, and excluding the run in which an objective is itself

the anchor, aDTsm is selected in 26 of 27 runs for k = 5 and maxVMS in 25, whereas minSrY is selected in 6 and energT in only 1. aDTsm and maxVMS are therefore stable members of the set, whereas energT is present because it is the default anchor of Equation (11) and the sustainability representative, tcycle because it is imposed, and minSrY because of the category constraint; these three memberships are design choices that the decision maker can revisit.

These results lead to a dataset-specific statistical observation that must be kept distinct from a physical claim: within the sampled processing window, and once the decision variables and the thermal descriptors are accounted for, the warpage and shrinkage indicators contribute little additional independent variance. Three qualifications apply. First, warpage is frequently a specification requirement; a low U_i means that it is tightly controlled by the cooling geometry within this window, which argues for tracking it as a constraint rather than removing it. Second, only the gate configuration and the cooling-channel geometry were varied, whereas melt and mold temperature, packing pressure, filling time and cooling time—the variables that dominate shrinkage and warpage—were held fixed; a wider design would probably leave more unexplained warpage variance and could reverse the conclusion. Third, seven warpage and shrinkage indicators are conditioning-degenerate, so the small incremental values reported above are partly numerical. The conditioning analysis should therefore be used to decide which of the many warpage and shrinkage components need to be optimized explicitly, not whether the family as a whole can be dropped.

The categorical variable gate, which defines the gate location, deserves separate treatment, because gate location alters the flow pattern, cooling behavior and pressure distribution in the mold and may therefore define distinct operating regimes. Models with gate-dependent interaction terms improve the fit for several objectives, and the reduction in dependence obtained by conditioning is markedly smaller within each regime than in the pooled data: from 0.386 to 0.240 (37.9%) for gate = 1 and from 0.367 to 0.264 (28.1%) for gate = 2, against 63.4% pooled. Part of the apparent simplification in the pooled analysis is therefore due to the mixing of two regimes, and the methodology applied separately to each regime returns different objective subsets.

The regime split also changes the physical pathway: within each regime the incremental contributions along the thermal–shrinkage–warpage chain become more pronounced, particularly between shrinkage and warpage (Figure 6; out-of-fold $\Delta R^2 = 0.071$ for gate = 1 and 0.063 for gate = 2, both with $p < 10^{-5}$), in contrast with the near-zero pooled values. Categorical variables that represent distinct process configurations should therefore be treated as regime-defining parameters, either by performing objective selection separately for each regime or by including interaction effects in the conditioning model.

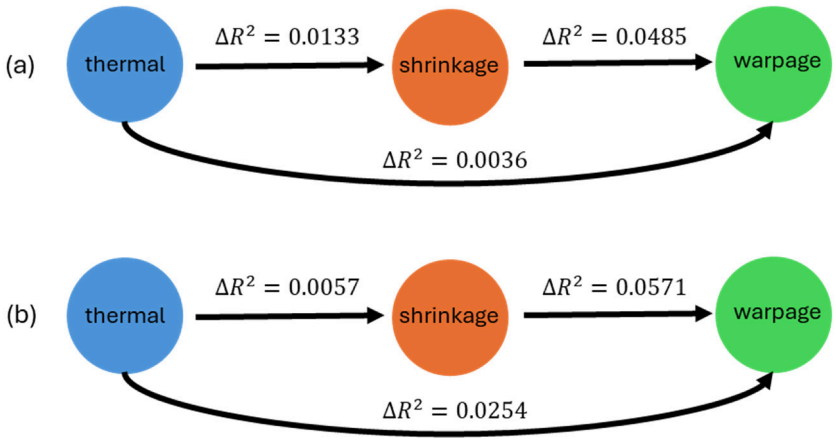


Figure 6. Thermal–shrinkage–warpage pathway by gate configuration: (**a**) gate = 1; (**b**) gate = 2.

4.6. Sensitivity Analysis of Selection Parameters

The sensitivity of the injection-molding selection to its control parameters—the weights, the redundancy threshold, the anchor, the cross-validation partition and the dependence measure—was examined for $k = 3$ and $k = 5$ on the 28 admissible candidates. The weights were varied by a factor of two around unity, the range within which a decision maker expresses a mild preference without overriding the other criteria, and τ over $[0.70, 0.95]$, the range over which the blocking rule is operative.

Table 16 summarizes the effect of the weighting parameters, evaluated on the twenty-eight admissible candidates without the category constraints used in Table 15. Halving any one weight changes at most one of the five selected objectives (80% overlap) and, for $k = 3$, leaves the subset unchanged except when β is halved (67%). Doubling a weight has a larger effect: doubling β leaves the three-objective subset unchanged but replaces two of the five objectives (60%); doubling κ replaces maxVMS by the warpage component minWaZ (67% and 40%); and doubling λ moves the selection towards the warpage and shrinkage components maxWaZ and maxSrZ (33% and 60%), which have the highest relevance among the admissible objectives and some of the lowest unexplained fractions. The selection is therefore robust to mild changes of emphasis but not invariant to the weights, which is the intended behavior: the weights express a genuine preference between discriminative power, information not explained by the design variables and residual redundancy, and doubling λ reproduces, in attenuated form, the mechanism isolated by the ablation below.

Table 16. Sensitivity of objective selection to weighting parameters.

k	Weights (λ, β, κ)	Selected Objectives	Overlap
3	(1, 1, 1)	(energT, aDTsm, maxVMS)	100%
3	(2, 1, 1)	(energT, maxWaZ, maxSrZ)	33%
3	(1, 2, 1)	(energT, aDTsm, maxVMS)	100%
3	(1, 1, 2)	(energT, aDTsm, minWaZ)	67%
3	(0.5, 1, 1)	(energT, aDTsm, maxVMS)	100%
3	(1, 0.5, 1)	(energT, maxWaZ, maxVMS)	67%
3	(1, 1, 0.5)	(energT, aDTsm, maxVMS)	100%
5	(1, 1, 1)	(energT, aDTsm, maxVMS, maxWaZ, sDTssH)	100%
5	(2, 1, 1)	(energT, maxWaZ, maxSrZ, maxVMS, maxWaX)	60%
5	(1, 2, 1)	(energT, aDTsm, maxVMS, sDTmsH, sDTsmH)	60%
5	(1, 1, 2)	(energT, aDTsm, minWaZ, sDTmsH, maxSrZ)	40%
5	(0.5, 1, 1)	(energT, aDTsm, maxVMS, sDTssH, maxSrX)	80%
5	(1, 0.5, 1)	(energT, maxWaZ, maxVMS, maxSrZ, aDTsm)	80%
5	(1, 1, 0.5)	(energT, aDTsm, maxVMS, maxWaZ, sDTmsH)	80%

The selected subset is identical for every tested value of $\tau \in \{0.70, 0.80, 0.90, 0.95\}$ and for $k = 3$, 4, 5 and 6: within this range the blocking rule never alters the outcome, and the subset is nested in k, each larger subset extending the smaller one by a single objective (maxWaZ, sDTssH and maxSrZ for $k = 4$, 5 and 6).

The anchor of Equation (11) is chosen by a criterion different from the subsequent steps, so its influence was also measured. Forcing each of the 28 admissible objectives in turn as the anchor yields 26 distinct subsets for $k = 3$ and 24 for $k = 5$ (mean pairwise Jaccard similarities of 0.29 and 0.37). These figures are dominated by the forced anchor itself,

which by construction belongs to every subset it initializes; the objectives selected after the anchor are considerably more stable, aDTsm being selected from 26 of the 28 anchors and maxVMS from 25 for $k = 5$. Where the anchor does change the outcome, the objective set contains several near-equivalent low-redundancy starting points, and the choice among them should be referred to relevance or engineering priorities, as in Section 4.5. This anchor sensitivity is a genuine limitation of the greedy procedure (Section 5): the subset should be reported together with its anchor-sensitivity range, here the objectives retained for most anchors (aDTsm and maxVMS), and, where several anchors are nearly equivalent, the choice among them should be stated explicitly. For k = 3, the corresponding counts are 24 of 27 for maxVMS and 15 of 27 for aDTsm.

Two further sources of variation were examined. Re-seeding the five-fold partition of the linear conditioner thirty times left the set of admissible candidates unchanged in every repetition and recovered the three-objective subset in 30 of 30 repetitions and the five-objective subset in 22 of 30; in the remaining eight, one or both of the last two objectives are replaced by related descriptors (sDTmsH, minWaZ or maxSrZ). Replacing the absolute Pearson correlation of Equation (6) by the distance correlation, as discussed in Section 3.3, changes the anchor from energT to sdDens, whose average residual dependence (0.172) is within 0.003 of that of energT (0.175); with the anchor held at energT, the distance-correlation subsets share two of three and four of five objectives with the Pearson subsets, energT, maxVMS, aDTsm and maxWaZ among them. The choice of dependence measure therefore acts on the selection mainly through the anchor, consistently with the anchor sensitivity reported above.

Three further checks were made. First, the residual-variance floor: with $U_{min} = 0.005$ or 0.01 the 28 admissible candidates and both subsets are unchanged; with 0.02 (23 candidates) and 0.05 (16 candidates) the three-objective subset is unchanged, and the five-objective subset becomes {energT, aDTsm, maxVMS, minWaZ, minSrY} and {energT, maxVMS, aDTsm, sDTssH, aDTms}; with 0.001 all 35 indicators are admitted and the subsets become {energT, maxSrY, aDTsm} and {energT, maxSrY, aDTsm, maxSrT, sDTssH}, that is, conditioning-degenerate shrinkage components enter as soon as the floor is relaxed. Second, the sign of the residual correlation: nine pairs have a residual correlation of -0.5 or less (minimum -0.721, between aDTss and coolef), and negative correlations contribute 42% of the total absolute residual dependence, but none reaches the blocking threshold. The positive-only variant of Section 3.3 selects {energT, sDTsmH, aDTsm} for k = 3 and {energT, sDTsmH, aDTsm, maxVMS, maxWaZ} for k = 5, sharing two of three and four of five objectives with the absolute form, and the signed residual correlations within the selected subsets are small ($|r| \leq 0.23$), so no strongly conflicting pair is separated. Third, replacing out-of-fold by in-sample residuals leaves both subsets unchanged.

Finally, the selected subset was checked against the original indicators. A linear regression of each of the 32 removed indicators on the three selected objectives explains on average 62% of its variance (median 70%; 56% of the indicators above 50%), and 75% with the five-objective subset. Subsets chosen by the marginal methods reconstruct the removed indicators better (87% for PCA and 84% for mRMR-like selection, k = 3), as expected, because those methods favour objectives that are strongly correlated with the others, which is precisely what the conditioned criterion penalizes. The selected subset is therefore not a compressed summary of the 35 indicators, but a set of objectives that carry complementary information after the design-variable effects are removed; the removed indicators and any process constraints must still be checked on the final designs, and the claims for this case are restricted accordingly (Section 4.7). The analysis therefore does not demonstrate that the selected subset is adequate for operational decisions; a validation of the optimized designs against the process constraints, or with process data, is required before such use.

Overall, these results show that the selected injection-molding subset is insensitive to the blocking threshold and robust to mild re-weighting, while responding in an interpretable way to stronger changes of emphasis. They should be interpreted as an application-level check rather than as proof of parameter independence across all benchmark families.

Table 17 reports the ablation of the score components: the subsets obtained with the full criterion (standardized form of Equation (10)), with the unstandardized form of Equation (10), and with each of the three terms removed in turn. Removing the unexplained-variability term reduces the criterion to a residual-redundancy analog of mRMR, so this row isolates the contribution of the term that distinguishes the proposed criterion from the standard relevance–redundancy framework. It is also the row that changes the outcome most: with $\beta = 0$, only the anchor is retained from the three-objective subset (33% overlap) and two objectives from the five-objective subset (40%). Removing the redundancy term changes one of three and two of five objectives (67% and 60%), whereas removing relevance, or using the unstandardized form of Equation (10), changes only the fifth objective (100% and 80%).

Table 17. Ablation of the components of the selection score and effect of the component standardization, injection-molding case.

Variant of the Criterion	k = 3 Subset	k = 5 Subset	Overlap with Standardized Equation (10)
Standardized Equation (10), full criterion ($\lambda = \beta = \kappa = 1$)	(energT, aDTsm, maxVMS)	(energT, aDTsm, maxVMS, maxWaZ, sDTssH)	100%/100%
Equation (10), unstandardized components	(energT, aDTsm, maxVMS)	(energT, aDTsm, maxVMS, sDTssH, minSrY)	100%/80%
$\beta = 0$, unexplained-variability term removed	(energT, maxWaZ, maxSrZ)	(energT, maxWaZ, maxSrZ, maxWaX, sdDens)	33%/40%
$\lambda = 0$, relevance term removed	(energT, aDTsm, maxVMS)	(energT, aDTsm, maxVMS, sDTssH, maxSrX)	100%/80%
$\kappa = 0$, residual-redundancy term removed	(energT, maxVMS, sDTsmH)	(energT, maxVMS, sDTsmH, aDTsm, sDTmsH)	67%/60%

The mechanism behind the $\beta = 0$ row is directly visible in the unexplained fractions. Ranked by U_i among the twenty-eight admissible candidates, the full criterion selects aDTsm (rank 25), maxVMS (rank 26) and sDTssH (rank 22), that is, objectives carrying some of the largest shares of variance not explained by the design variables. Without the unexplained-variability term, the selector takes instead maxWaZ, maxSrZ, maxWaX and sdDens, ranked 5, 4, 1 and 3—objectives lying just above the residual-variance floor, with U_i between 0.012 and 0.018 against a median of 0.061. Removing the term therefore drives the selection towards objectives that are almost determined by the decision variables, which is precisely the behavior that the term exists to prevent. Decomposing the winning scores of the full criterion confirms the ordering: for $k = 5$ the unexplained-variability term contributes between 67% and 87% of the score at three of the four greedy steps, and the order of influence of the three terms in this case is $U_i > Red_i(S) > Rel_i$.

4.7. When Conditioning-Based Selection Should Be Preferred

The validation of Section 4.4 establishes no dominance, so the useful question is not which method is better in general but when the additional cost of conditioning is justified. The results support four operational conditions, each of which can be checked before any optimization run is launched. Satisfying them makes the criterion applicable; it does not by itself establish that it should be preferred to a marginal method, which depends on whether the decision maker needs the decomposition that it provides.

The first is availability. Conditioning requires the decision variables of the sampled solutions and not only their objective values. Where only an objective archive exists—for instance when reusing published Pareto approximations—the criterion cannot be applied at all, and marginal-dependence methods are the only option.

The second is conditioner adequacy in the sense of Section 3.4. Where the best conditioning model reaches a mean out-of-fold R^2 above 0.50, the residuals carry information that the marginal correlation does not, and the resulting subsets differ from the marginal ones in four of the five admitted benchmark configurations of Table 6. Where it does not, WFG9 at every dimensionality in this study, and DTLZ3 with ten objectives, conditioning adds computational cost without adding information, and PCA, mRMR or clustering should be preferred outright. A mean out-of-fold R^2 above 0.50 is therefore a necessary but not sufficient admissibility condition: the residual-variance floor must subsequently be checked. DTLZ7 illustrates the complementary failure mode, in which conditioning is highly predictive but the residual variance becomes degenerate. This is the single most useful decision rule the paper provides, because it is cheap to evaluate and it is checked on the same sample that the selection will use.

The third is the structure of the objective set. The criterion pays off where many objectives are known to be driven by the same small set of design parameters, which is the normal situation in engineering problems containing families of related indicators: the thirty-five injection-molding indicators reduce to a five-objective set with an interpretable member in each physical category, this subset carries complementary rather than representative information (Section 4.6), so it does not replace a check of the removed indicators and of the process constraints on the final designs. Where the objectives are already few and physically distinct, marginal screening is sufficient and cheaper.

The fourth is whether interpretability is required. The proposed criterion returns objectives in their original units together with a decomposition of why each was retained—a high unexplained fraction, a low residual redundancy, a high relevance—and that decomposition is auditable by a domain expert and can be overridden by engineering constraints, as in Table 15. This is the property that motivated the method, and it is the reason to use it; the optimization results alone do not provide one. Where interpretability is not required and only the quality of the final front matters, the results of Section 4.4 give no reason to prefer it to mRMR or clustering.

The analysis also identifies a case in which conditioning-based selection should be avoided even when the conditioning model is accurate. Where a large fraction of the objectives falls below the residual-variance floor, as in DTLZ7, the residual redundancy matrix is dominated by uninterpretable entries, and the selection should be driven by relevance or by decision-maker preference instead. Finally, the diversity collapse documented in Section 4.4.3 implies that the subset size k must be justified per problem: a value of k that reduces the non-dominated count to single digits has removed trade-off structure rather than redundancy, and this failure is independent of which selector produced the subset.

5. Conclusions

This work proposed a conditioning-based criterion for objective selection in many-objective optimization, in which redundancy is evaluated on the residuals that remain after each objective is regressed on the decision variables. The estimator is classical partial-correlation analysis; the contribution lies in its use as a selection criterion ahead of an evolutionary optimizer, in a greedy score that adds an unexplained-variability term to the relevance–redundancy framework, and in an admissibility gate that reports when the criterion does not apply. The method is conditional on this diagnostic: in the benchmarks it

applies to five of the twelve configurations, and it is offered as a criterion for problems in which an adequate conditioning model exists, not as a general solution to objective selection.

On the DTLZ and WFG benchmarks, the complete procedure admitted five of the twelve configurations; in four of them the conditioned subset differed from those of all three marginal baselines, and in three it also differed from a matched selector using raw correlations, and the active NSGA-III validation showed that the reduced subsets can be used directly inside the optimizer, with mixed full-space results (more favourable spacing and non-dominated counts, but more often worse HV and IGD+ than the baselines) and no consistent statistical superiority over PCA, mRMR or clustering, as shown by paired Wilcoxon tests. The practical contribution is therefore the specification of when the criterion should be used (Section 4.7): the decision variables must be available, the conditioning model must be adequate, the objective set should contain families driven by shared design parameters, and the retained objectives must remain interpretable. Where the gate fails, as for WFG9, cheaper marginal methods are the correct choice.

In the injection-molding case, the ablation showed that the unexplained-variability term is the most influential component of the score, and that conditioning-degenerate objectives must be excluded from the candidate set rather than merely exempted from the redundancy penalty. The final five-objective set contains one interpretable descriptor per physical category, the many shrinkage and warpage components collapse onto one or two representatives within the sampled processing window, and the gate location defines distinct process regimes that should be analyzed separately. These findings describe the statistical structure of the sampled indicators; they do not demonstrate that the selected subset is adequate for operational decisions, which requires validation against process constraints and process data.

The main limitations are the dependence on an adequate conditioning model, the cost in front diversity that accompanies aggressive reduction, and a design of experiments that held the thermal processing variables fixed. Four further limitations should be stated. The greedy subset depends on the anchor (Section 4.6), so the subset should be reported together with its anchor-sensitivity range. The absolute residual correlation treats residual conflict as redundancy, and the controlled example of Section 4.4.6 shows that this can remove an essential objective; the sign-aware variant should be used when such conflicts are plausible. The active validation used a single optimizer, NSGA-III, with one budget per dimensionality; because reference-direction algorithms are particularly sensitive to the number of objectives, the differences between full and reduced formulations may be different for other optimizers, and the conclusions are limited to this setting. Finally, the full-space indicators depend on a normalization computed from the pooled populations, so they compare formulations within a configuration only. Future work will extend the validation to other optimizers and objective-reduction methods, integrate the selection mechanism into adaptive evolutionary workflows, address multi-regime and time-dependent problems, and validate the injection-molding subset against process constraints and process data.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/math14214051/s1, Supplementary Material S1: Table S1: Definition of the candidate objectives of the injection-molding case study. Supplementary Material S2: Objective Selection with Sign-Aware Residual Redundancy; Table S2: Controlled scenarios: correct selections and false-dominance rate for each family and amplitude of the residual-conflict term; Table S3: Subsets returned by the complete procedure with the absolute and the sign-aware residual redundancy (benchmark configurations); Table S4: Active optimization validation for DTLZ2 with the sign-aware subsets (mean ± standard deviation over 30 NSGA-III runs); Table S5: Paired Wilcoxon signed-rank comparisons of the sign-aware subset (Holm-adjusted p-values and Vargha–Delaney effect sizes); Table S6: Objective subsets selected in the injection-molding case with the absolute and the sign-

aware residual redundancy; Table S7: Sensitivity of the injection-molding selection to the weighting parameters with both forms; Table S8: Ablation of the score components with both forms; Table S9: Robustness checks of the injection-molding selection with both forms.

Author Contributions: Conceptualization, A.G.-C.; methodology, A.G.-C.; software, A.G.-C. and T.M.; validation, T.M.; formal analysis, A.G.-C. and T.M.; investigation, T.M.; writing—original draft preparation, A.G.-C.; writing—review and editing, A.G.-C., T.M. and A.J.P. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by FEDER funds through the COMPETE 2020 Program and by National Funds through FCT (Portuguese Foundation for Science and Technology) under the projects UID-B/05256/2020 and UID-P/05256/2020, and by "Agendas para a Inovação Empresarial" (Project no. 49, acronym "INOV.AM", reference PRR/49/INOV.AM/EE, operation code 02/C05-i01.01/2022.PC644865234-00000004), supported by the RRP—Recovery and Resilience Plan and by the European Funds Next Generation EU.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data and code supporting the findings of this study, including the injection-molding dataset, the implementation of the selection procedure, the scripts that regenerate the benchmark samples and analyses, and the final populations and run-level results of the NSGA-III validation, are openly available in DataRepositoriUM, the research data repository of the University of Minho, at https://doi.org/10.34622/datarepositorium/AZOFUA.

Acknowledgments: During preparation of this manuscript, the authors used ChatGPT (GPT-6) for bibliographic searches, Claude (Opus 5.5) for editing, restructuring, and text review, and Grammarly (v.1.2) for final text checking. The authors reviewed all output and take full responsibility for the content of this publication.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Deb, K.; Jain, H. An Evolutionary Many-Objective Optimization Algorithm Using Reference-Point-Based Nondominated Sorting Approach, Part I: Solving Problems with Box Constraints. *IEEE Trans. Evol. Comput.* **2014**, *18*, 577–601. [CrossRef]
2. Li, B.; Li, J.; Tang, K.; Yao, X. Many-Objective Evolutionary Algorithms: A Survey. *ACM Comput. Surv.* **2015**, *48*, 13. [CrossRef] [PubMed]
3. Yuan, Y.; Xu, H.; Wang, B.; Yao, X. A New Dominance Relation-Based Evolutionary Algorithm for Many-Objective Optimization. *IEEE Trans. Evol. Comput.* **2016**, *20*, 16–37. [CrossRef]
4. Jolliffe, I.T. *Principal Component Analysis*, 2nd ed.; Springer: New York, NY, USA, 2002. [CrossRef]
5. Peng, H.; Long, F.; Ding, C. Feature Selection Based on Mutual Information: Criteria of Max-Dependency, Max-Relevance, and Min-Redundancy. *IEEE Trans. Pattern Anal. Mach. Intell.* **2005**, *27*, 1226–1238. [CrossRef] [PubMed]
6. Saxena, D.K.; Deb, K. Dimensionality Reduction of Objectives and Constraints in Multi-Objective Optimization Problems: A System Design Perspective. In Proceedings of the 2008 IEEE Congress on Evolutionary Computation, Hong Kong, China, 1–6 June 2008; pp. 3204–3211. [CrossRef]
7. Gaspar-Cunha, A.; Costa, P.; Monaco, F.; Delbem, A. Many-Objectives Optimization: A Machine Learning Approach for Reducing the Number of Objectives. *Math. Comput. Appl.* **2023**, *28*, 17. [CrossRef]
8. Peters, J.; Janzing, D.; Schölkopf, B. *Elements of Causal Inference: Foundations and Learning Algorithms*; MIT Press: Cambridge, MA, USA, 2017.
9. Pearl, J. *Causality: Models, Reasoning, and Inference*, 2nd ed.; Cambridge University Press: Cambridge, UK, 2009.
10. Schölkopf, B.; Locatello, F.; Bauer, S.; Ke, N.R.; Kalchbrenner, N.; Goyal, A.; Bengio, Y. Toward Causal Representation Learning. *Proc. IEEE* **2021**, *109*, 612–634. [CrossRef]
11. Deb, K.; Thiele, L.; Laumanns, M.; Zitzler, E. Scalable Multi-Objective Optimization Test Problems. In Proceedings of the 2002 Congress on Evolutionary Computation, Honolulu, HI, USA, 12–17 May 2002; pp. 825–830. [CrossRef]
12. Huband, S.; Hingston, P.; Barone, L.; While, L. A Review of Multiobjective Test Problems and a Scalable Test Problem Toolkit. *IEEE Trans. Evol. Comput.* **2006**, *10*, 477–506. [CrossRef]

13. Kraskov, A.; Stögbauer, H.; Grassberger, P. Estimating Mutual Information. *Phys. Rev. E* **2004**, *69*, 066138. Erratum in *Phys. Rev. E* **2011**, *83*, 019903. [CrossRef] [PubMed]
14. Yang, N.; Liu, H.-L.; Xiao, J. An Objective Reduction Algorithm Based on Population Decomposition and Hyperplane Approximation. *Swarm Evol. Comput.* **2024**, *86*, 101495. [CrossRef]
15. Saxena, D.K.; Duro, J.A.; Tiwari, A.; Deb, K.; Zhang, Q. Objective Reduction in Many-Objective Optimization: Linear and Nonlinear Algorithms. *IEEE Trans. Evol. Comput.* **2013**, *17*, 77–99. [CrossRef]
16. Pal, M.; Saha, S.; Bandyopadhyay, S. DECOR: Differential Evolution Using Clustering Based Objective Reduction for Many-Objective Optimization. *Inf. Sci.* **2018**, *423*, 200–218. [CrossRef]
17. He, X.; Cai, D.; Niyogi, P. Laplacian Score for Feature Selection. In *Advances in Neural Information Processing Systems 18*; Weiss, Y., Schölkopf, B., Platt, J., Eds.; MIT Press: Cambridge, MA, USA, 2005; pp. 507–514.
18. Marques, T.; Melo, J.B.; Pontes, A.J.; Gaspar-Cunha, A. Surrogate-Assisted Many-Objective Optimization of Injection Molding: Effects of Objective Selection and Sampling Density. *Appl. Sci.* **2026**, *16*, 1578. [CrossRef]
19. Spirtes, P.; Glymour, C.; Scheines, R. *Causation, Prediction, and Search*, 2nd ed.; MIT Press: Cambridge, MA, USA, 2001.
20. Gu, X.; Guo, J.; Xiao, L.; Li, C. Conditional Mutual Information-Based Feature Selection Algorithm for Maximal Relevance Minimal Redundancy. *Appl. Intell.* **2022**, *52*, 1436–1447. [CrossRef]
21. Chugh, T.; Gaspar-Cunha, A.; Deutz, A.H.; Duro, J.A.; Oara, D.C.; Rahat, A. Identifying Correlations in Understanding and Solving Many-Objective Optimisation Problems. In *Many-Criteria Optimization and Decision Analysis*; Brockhoff, D., Emmerich, M., Naujoks, B., Purshouse, R., Eds.; Springer: Cham, Switzerland, 2023; pp. 241–267. [CrossRef]
22. Bhatija, S.; Zuercher, P.-D.J.; Thumm, J.; Bohné, T. Multi-Objective Causal Bayesian Optimization. In Proceedings of the 42nd International Conference on Machine Learning, Vancouver, BC, Canada, 13–19 July 2025; Volume 267, pp. 4172–4195.
23. Gretton, A.; Bousquet, O.; Smola, A.; Schölkopf, B. Measuring Statistical Dependence with Hilbert–Schmidt Norms. In *Algorithmic Learning Theory*; Jain, S., Simon, H.U., Tomita, E., Eds.; Springer: Berlin/Heidelberg, Germany, 2005; pp. 63–77. [CrossRef]
24. Székely, G.J.; Rizzo, M.L. Brownian Distance Covariance. *Ann. Appl. Stat.* **2009**, *3*, 1236–1265. [CrossRef] [PubMed]
25. Wilcoxon, F. Individual Comparisons by Ranking Methods. *Biom. Bull.* **1945**, *1*, 80–83. [CrossRef]
26. Holm, S. A Simple Sequentially Rejective Multiple Test Procedure. *Scand. J. Stat.* **1979**, *6*, 65–70.
27. Vargha, A.; Delaney, H.D. A Critique and Improvement of the CL Common Language Effect Size Statistics of McGraw and Wong. *J. Educ. Behav. Stat.* **2000**, *25*, 101–132. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.