# FINITE BLACK-BOX OBSERVATIONS CANNOT IDENTIFY DECOMPOSITIONS FOR EFFICIENT OPTIMIZATION

**Anonymous authors**
Paper under double-blind review

## ABSTRACT

The No-Free-Lunch theorem provides an algorithm-level statement: no decomposition strategy can be universally dominant across black-box optimization problems. We focus on a more basic, information-level question about decomposition itself rather than the strategy that implements it: *can decompositions for efficient optimization be identified from finite black-box observations?* To this end, we establish a general theoretical framework for analyzing decomposition in black-box optimization, which explicitly characterizes the decomposition space and the intrinsic complexity of optimal decomposition identification. Building on this framework, we establish a non-identifiability theory and answer the question in the negative: decompositions for efficient optimization cannot be reliably identified from finite black-box observations. Through theory-driven counterexample analysis, we reveal a fundamental objective mismatch in prevailing decomposition strategies that infer problem structure from finite information: their long-standing emphasis on structural recovery may not align with the ultimate goal of efficient optimization. We further revisit the random decomposition strategy, which is generally considered ineffective, and show that it provides a minimax-robust strategy under finite information. Together, the theoretical and empirical findings provide guidance for several promising directions in future research on decomposition strategy design.

## 1 INTRODUCTION

Divide-and-conquer is a fundamental principle for handling complex systems. Rather than handling a complex system as a whole, it decomposes the system into smaller and more manageable subsystems. This principle has been widely adopted across scientific and engineering domains (Sun et al., 2024a; Liu et al., 2024; Sun et al., 2024b). Despite substantial differences in their specific implementations, these approaches all face a fundamental choice: how should the complex system be decomposed so that the resulting subsystems facilitate, rather than hinder, the solution of the original problem?

Because practical systems differ substantially in their internal mechanisms, there is no shared way to assess the quality of a decomposition. What they share instead is a common external description: given a joint configuration of their constituent components, the system produces an overall performance value (Wei et al., 2025). Such a system can therefore be represented as a black-box function $f : \mathcal{X} \to \mathbb{R}$, where $\mathbf{x} \in \mathcal{X}$ denotes the joint configuration of the system components and $f(\mathbf{x})$ denotes the resulting system-level performance (Torres et al., 2026). Hence, in the absence of a shared internal criterion, the quality of a decomposition can be measured by the subsequent optimization performance. An effective decomposition induces tractable subproblems whose optimization provides an efficient route to solving the original problem. In contrast, an inappropriate one may leave the resulting subproblems difficult to optimize while introducing additional coordination costs, potentially making the decomposed problem even harder than the original one.

Cooperative Coevolution provides a natural framework for studying decomposition (Potter & De Jong, 1994; Ma et al., 2018; Yao & Chong, 2025). Given a system-level objective function $f$, a decomposition $\mathcal{G}$ organizes the variables in $\mathbf{x}$ into a collection of subspaces, which are subsequently optimized to improve the overall objective. In this way, decomposition across different complex systems can be studied within a unified black-box optimization framework, allowing different decomposition strategies and their effects on final optimization performance to be systematically analyzed and validated under controllable experimental settings (De Jong & Potter, 1995). Existing approaches



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

have evolved from early correlation- and covariance-based methods (Liu & Tang, 2013; Ray & Yao, 2009), to interaction-based strategies (Chen et al., 2010; Omidvar et al., 2013; Tian et al., 2024a; Qiu et al., 2025), and more recently to learning-based decomposition strategies (Tian et al., 2025; Qiu et al., 2026a;b). These developments have achieved substantial success on challenging problems such as large-scale global optimization (Omidvar et al., 2021), with significant performance improvements reported across representative benchmark suites (Li et al., 2013).

Despite their methodological differences, these decomposition strategies operate under the same fundamental information constraint: for a black-box function, only a finite set of objective evaluations is available. From these finite black-box observations, they infer latent problem structures, such as variable interactions and correlations, and use the inferred structures to construct a decomposition. Their main differences lie in how the information is acquired, interpreted, and translated into a decomposition. We refer to this general paradigm as the structure-driven decomposition strategy.

The effectiveness of this paradigm has been demonstrated primarily through empirical evaluation on benchmark suites. Such empirical evidence is essential, but any benchmark suite inevitably covers only a limited range of problem characteristics. Consequently, the advantages observed on these suites demonstrate the effectiveness of a decomposition strategy only for the tested problem classes. The No-Free-Lunch theorem implies that no decomposition strategy can be universally dominant across black-box optimization problems (Wolpert, 1996; Wolpert & Macready, 1997). However, a more basic, information-level question remains: *can decompositions for efficient optimization be identified from finite black-box observations?* Unlike the stronger notion of universal dominance, this question asks whether even a preferable decomposition can be identified for an unknown black-box optimization problem, focusing on the decomposition itself rather than the strategy used to induce it.

Complementing existing empirical studies, we develop a theoretical framework to investigate decomposition from a finite-information perspective. We establish a non-identifiability theory for efficient decompositions under finite information, thereby providing a negative answer to the above question. We then examine the implications of this result for the structure-driven and random decomposition strategy. The main contributions of this work are summarized as follows:

**First, we establish a general theoretical framework for analyzing decomposition in black-box optimization.** By establishing a rigorous theoretical framework, we show that the number of possible decompositions grows doubly exponentially with dimensionality. We then characterize the identification of an effective decomposition as a resource-constrained black-box optimization problem with an enormous search space and expensive, noisy evaluations.

**Second, within the framework, we answer three core questions about decomposition:**

**Q1.** *Can decompositions for efficient optimization be identified from finite black-box observations?* We establish a non-identifiability theory showing that decompositions for efficient optimization cannot be reliably identified from finite black-box observations, and further that no decomposition strategy induced by them can provide reliable performance guarantees.

**Q2.** *Does structural recovery in typical decomposition strategies really lead to efficient optimization?* Guided by the established theory, we construct a counterexample showing that two fundamentally different functions can produce identical finite black-box observations and thus induce the same structure-driven decomposition. However, the two functions favor different decompositions in subsequent optimization. This reveals a fundamental objective mismatch: structural recovery does not necessarily translate into optimization gains, suggesting that mainstream strategy design may be built on assumptions that do not always hold.

**Q3.** *Is the random decomposition strategy really inefficient?* In contrast to the conventional view that the random decomposition strategy becomes ineffective as dimensionality increases because the probability of grouping interacting variables into the same subspace vanishes, we theoretically show that it provides a zero-cost and minimax-robust baseline under finite black-box observations. Extensive experiments further validate the stability of its performance.

The remainder of this paper is organized as follows. Section 2 develops the theoretical framework, and Section 3 presents the main theoretical results. Section 4 discusses their implications for structure-driven and random decomposition strategies. Finally, Section 5 concludes the paper and outlines promising directions for future decomposition strategy design. All code is available in our repository.

# 2 THEORETICAL FRAMEWORK AND ANALYSIS

In this section, we first formally characterize the decomposition behavior and the space in which it operates. We then distinguish between decomposition behavior itself and a decomposition strategy: the former specifies how the problem is decomposed, whereas the latter defines the mechanism for inducing such behavior. We further establish performance criteria for both. Together, these definitions constitute a general theoretical framework for analyzing decomposition.

## 2.1 ANALYTICAL SPACE AND CRITERIA

**General decomposition space.** Let $[D] = \{1, \ldots, D\}$ denote the index set of the $D$ decision variables. A decomposition $\mathcal{G}$ is a collection of nonempty subsets of $[D]$ whose union covers all variables. This leads to the following general decomposition space.

**Definition 1: General Decomposition Space**

The general decomposition space over $D$ decision variables is defined as

$$\mathfrak{D}_D = \left\{ \mathcal{G} \subseteq 2^{[D]} \setminus \{\varnothing\} : \bigcup_{G \in \mathcal{G}} G = [D] \right\}, \tag{1}$$

where $2^{[D]}$ denotes the power set of $[D]$.

This definition does not require the subsets in $\mathcal{G}$ to be disjoint. It therefore includes conventional nonoverlapping decompositions as a special case while also allowing overlapping decompositions.

**Performance of a decomposition.** A decomposition $\mathcal{G}$ serves as an intermediate step in problem solving, and its performance should ultimately be evaluated by the optimization performance achieved after decomposition (Qiu et al., 2026c). Let $\mathcal{F}$ denote the class of black-box objective functions $f : \mathcal{X} \to \mathbb{R}$, where $\mathcal{X} \subseteq \mathbb{R}^D$, and let $\mathcal{A}$ denote a fixed optimization protocol applied after decomposition. To compare different decompositions under the same available information, let $h$ denote a finite evaluation history available before decomposition. For a minimization problem $f$, let $f^\star$ denote its optimal objective value. Given a target accuracy $\epsilon > 0$, let $T_\epsilon(f, \mathcal{G}, \mathcal{A} \mid h)$ denote the number of additional function evaluations required by $\mathcal{A}$, starting from the same history $h$ and using decomposition $\mathcal{G}$, to obtain a solution $\hat{\mathbf{x}}$ satisfying $|f(\hat{\mathbf{x}}) - f^\star| \leq \epsilon$.

**Definition 2: Performance of a Decomposition**

Given a problem $f \in \mathcal{F}$, a decomposition $\mathcal{G} \in \mathfrak{D}_D$, a fixed optimization protocol $\mathcal{A}$, a finite evaluation history $h$, and a target accuracy $\epsilon > 0$, the performance of $\mathcal{G}$ is defined as

$$L_\epsilon(f, \mathcal{G}, \mathcal{A} \mid h) = \mathbb{E}[T_\epsilon(f, \mathcal{G}, \mathcal{A} \mid h)], \tag{2}$$

where the expectation is taken over the randomness of the subsequent optimization process. If the target accuracy cannot be reached, we set $T_\epsilon = +\infty$.

Based on Definition 2, for a fixed problem $f$, optimization protocol $\mathcal{A}$, finite evaluation history $h$, and target accuracy $\epsilon$, we define an optimal decomposition as

$$\mathcal{G}_f^\star(h) \in \arg\min_{\mathcal{G} \in \mathfrak{D}_D} L_\epsilon(f, \mathcal{G}, \mathcal{A} \mid h). \tag{3}$$

**Performance of a decomposition strategy.** A decomposition strategy $\Pi$ is a concrete mechanism for implementing decomposition behavior and typically consists of multiple components, including sampling, interpretation, and decomposition. Its performance should therefore be evaluated in terms of the resulting overall optimization performance. Let $q$ denote a sampling rule that generates an evaluation history consisting of $T_q$ function evaluations:

$$h = H_q(f) = \big((X_t, Y_t)\big)_{t=1}^{T_q}, \qquad Y_t = f(X_t). \tag{4}$$

Based on the resulting history $H_q(f)$, a decomposition is selected according to a conditional distribution $\pi(\mathcal{G} \mid H_q(f))$ over $\mathfrak{D}_D$. We therefore characterize a decomposition strategy $\Pi$ by the pair $(q, \pi)$, where $q$ determines how finite information is acquired and $\pi$ determines how the resulting information is used to select a decomposition.

For a decomposition strategy $\Pi = (q, \pi)$ followed by a fixed optimization protocol $\mathcal{A}$, we define its overall optimization performance as

$$C_\epsilon(f, \Pi, \mathcal{A}) = \mathbb{E}[T_q + L_\epsilon(f, \mathcal{G}, \mathcal{A} \mid H_q(f))], \qquad \mathcal{G} \sim \pi(\cdot \mid H_q(f)), \tag{5}$$

where the outer expectation is taken over the randomness of the $q$, $\pi$ and $\mathcal{A}$.

### 2.2 Intrinsic Complexity of Decomposition Identification

Eq. 3 shows that identifying an optimal decomposition itself requires searching over $\mathfrak{D}_D$ according to the resulting performance $L_\epsilon$. This search faces two fundamental challenges: the decomposition space is extremely large, and evaluating the performance is expensive and noisy.

We first quantify the general decomposition space; the proof is provided in Appendix A.1.

**Theorem 1: Doubly Exponential Growth of the General Decomposition Space**

For $D \geq 1$, the cardinality of the general decomposition space $\mathfrak{D}_D$ is

$$|\mathfrak{D}_D| = \sum_{k=0}^{D} (-1)^k \binom{D}{k} 2^{2^{D-k}-1}. \tag{6}$$

Moreover, as $D \to \infty$, $|\mathfrak{D}_D| \sim 2^{2^{D}-1}$.

The space grows explosively: $|\mathfrak{D}_3| = 10^9$, while $|\mathfrak{D}_7| \approx 1.70 \times 10^{38}$, already exceeding the estimated number of stars in the observable universe by at least 15 orders of magnitude (Manojlović, 2015). Exhaustive enumeration therefore becomes infeasible even at relatively low dimensionalities.

The second challenge lies in evaluating candidate decompositions. The quantity $L_\epsilon(f, \mathcal{G}, \mathcal{A} \mid h)$ is not directly observable from black-box feedback; estimating it requires executing the subsequent optimization protocol, which may consume substantial evaluations. Moreover, when $\mathcal{A}$ is itself stochastic due to factors such as random seeds, the realized cost $T_\epsilon$ may vary across runs. A single run therefore provides only a noisy estimate of $L_\epsilon$, and reliable comparison may require repeated evaluations. Consequently, identifying an effective decomposition can itself be viewed as a resource-constrained black-box optimization problem over $\mathfrak{D}_D$, characterized by an enormous search space and expensive, potentially noisy performance evaluations.

## 3 Fundamental Limits of Decomposition Identification

Building on the complexity characterized in Section 2.2, we now consider a more fundamental question: can finite black-box observations identify decompositions for efficient optimization? The answer is negative, and the intuition is straightforward: finite observations cannot uniquely determine the underlying objective function. Consequently, the same finite history can be consistent with multiple objective functions that favor different decompositions. We next establish this result.

Given a finite evaluation history $h = \big((X_t, Y_t)\big)_{t=1}^{T}$, we define its compatible function class as

$$\mathcal{F}(h) = \{g \in \mathcal{F} : g(X_t) = Y_t, t = 1, \ldots, T\}, \tag{7}$$

which contains all black-box functions consistent with the observed history $h$.

**Lemma 1: Finite-History Function Non-Uniqueness**

Let $h = \big((X_t, Y_t)\big)_{t=1}^{T}$ be a finite evaluation history such that $\mathcal{X} \setminus \{X_1, \ldots, X_T\} \neq \varnothing$. Then,

$$|\mathcal{F}(h)| = \infty. \tag{8}$$

*Proof.* Since the finite history $h$ does not exhaust $\mathcal{X}$, there exists an unevaluated point $\bar{X} \in \mathcal{X} \setminus \{X_1, \ldots, X_T\}$. Take any $g \in \mathcal{F}(h)$. For each $c \in \mathbb{R}$, define

$$g_c(x) = \begin{cases} c, & x = \bar{X}, \\ g(x), & \text{otherwise.} \end{cases} \tag{9}$$

Because $g_c(X_t) = g(X_t) = Y_t$ for all $t = 1, \ldots, T$, we have $g_c \in \mathcal{F}(h)$ for every $c \in \mathbb{R}$. Since different values of $c$ define distinct functions, $\mathcal{F}(h)$ contains infinitely many functions. ■

Lemma 1 shows that infinitely many functions can be consistent with the same finite evaluation history, implying that such a history cannot uniquely determine the underlying objective function.

We further show that the subsequent optimization process is likewise insufficient to uniquely determine the underlying objective function. Consider a decomposition $\mathcal{G}$ and a fixed optimization protocol $\mathcal{A}$ starting from the same history $h$. Under a finite evaluation budget $B$, let $Z_1, \ldots, Z_B$ denote the subsequent query points and let $V_t = f(Z_t)$ denote the corresponding objective values. We define the resulting finite evaluation sequence as

$$\Theta_{\mathcal{G},B}(f \mid h) = (V_1, \ldots, V_B). \tag{10}$$

**Lemma 2: Finite-Evaluation Transfer across Decompositions**

Given a finite evaluation history $h$, a fixed optimization protocol $\mathcal{A}$, and any two decompositions $\mathcal{G}, \mathcal{G}' \in \mathfrak{D}_D$, for any $f \in \mathcal{F}(h)$ and any finite feasible evaluation budget $B$, there exists a function $f' \in \mathcal{F}(h)$ such that

$$\Theta_{\mathcal{G}',B}(f' \mid h) = \Theta_{\mathcal{G},B}(f \mid h). \tag{11}$$

*Proof.* Fix $h, \mathcal{A}, \mathcal{G}, \mathcal{G}', f$, and a finite feasible budget $B$. We construct a function $f' \in \mathcal{F}(h)$ recursively along the queries generated by $\mathcal{G}'$. Starting from the same history $h$, let $Z'_1$ be the first subsequent query generated by $\mathcal{A}$ under $\mathcal{G}'$. Define

$$f'(Z'_1) = V_1. \tag{12}$$

Suppose that, for some $t > 1$, $f'$ has been defined at the first $t-1$ subsequent query points such that

$$f'(Z'_s) = V_s, \qquad s = 1, \ldots, t-1. \tag{13}$$

The optimization protocol $\mathcal{A}$ then generates its next query $Z'_t$ based on $\mathcal{G}'$, the initial history $h$, and the feedback observed thus far. We may then define $f'(Z'_t) = V_t$. Repeating this construction for $t = 1, \ldots, B$ yields $f'(Z'_t) = V_t$ for all $t = 1, \ldots, B$.

Finally, set $f'(X_t) = Y_t$ for all observations in the initial history $h$ and assign arbitrary real values to all remaining points in $\mathcal{X}$. Hence $f' \in \mathcal{F}(h)$ and $\Theta_{\mathcal{G}',B}(f' \mid h) = \Theta_{\mathcal{G},B}(f \mid h)$. ■

Lemma 1 and Lemma 2 together show that, under a fixed optimization protocol $\mathcal{A}$, infinitely many function–decomposition pairs can generate the same finite history and subsequent evaluation sequence. Building on this observation, we establish the main non-identifiability theorem.

**Theorem 2: Finite-Information Non-Identifiability of Effective Decomposition**

Given a finite evaluation history $h$, a fixed optimization protocol $\mathcal{A}$, and a target accuracy $\epsilon$, consider any two decompositions $\mathcal{G}, \mathcal{G}' \in \mathfrak{D}_D$. If there exists a function $f_1 \in \mathcal{F}(h)$ such that

$$L_\epsilon(f_1, \mathcal{G}, \mathcal{A} \mid h) < L_\epsilon(f_1, \mathcal{G}', \mathcal{A} \mid h), \tag{14}$$

then there exists another function $f_2 \in \mathcal{F}(h)$ such that

$$L_\epsilon(f_2, \mathcal{G}', \mathcal{A} \mid h) < L_\epsilon(f_2, \mathcal{G}, \mathcal{A} \mid h). \tag{15}$$

*Proof.* Because only finitely many objective values have been specified by $h$ and the preceding evaluation, choose $m \in \mathbb{R}$ such that

$$V - m > \epsilon \tag{16}$$

for every objective value $V$ appearing in $h$ or in $(V_1, \ldots, V_{t-1})$.

We now construct a function $f_2 \in \mathcal{F}(h)$. By Lemma 1 and Lemma 2, we can preserve all objective values in $h$ and the preceding evaluation sequence $(V_1, \ldots, V_{t-1})$, while defining

$$f_2(Z'_t) = m, \qquad f_2(Z_t) = m + 2\epsilon + 1. \tag{17}$$

Set all remaining values above $m + \epsilon$, yielding $f_2 \in \mathcal{F}(h)$ with $f_2^\star = m$.

Under $f_2$, the $\mathcal{G}, \mathcal{G}'$ receive exactly the same history and the same feedback up to step $t-1$, and therefore generate the same respective query points $Z_t$ and $Z'_t$ at step $t$. None of the preceding evaluations satisfies the target accuracy, whereas $\mathcal{G}'$ reaches the target at $Z'_t$ and $\mathcal{G}$ does not. Hence,

$$T_\epsilon(f_2, \mathcal{G}', \mathcal{A} \mid h) = t < T_\epsilon(f_2, \mathcal{G}, \mathcal{A} \mid h). \tag{18}$$

Since infinitely many pairs $(Z_t, Z'_t)$ satisfy Eq. 17, the argument extends to stochastic $\mathcal{A}$ by conditioning on an arbitrary realization of its internal randomness. Under each realization, the construction can be carried out such that Eq. 18 holds. Consequently,

$$L_\epsilon(f_2, \mathcal{G}', \mathcal{A} \mid h) < L_\epsilon(f_2, \mathcal{G}, \mathcal{A} \mid h), \tag{19}$$

which proves the result. ■

Theorem 2 directly establishes that finite black-box observations cannot identify decompositions for efficient optimization. Fundamentally, this is because such observations cannot uniquely determine the underlying objective function, leaving infinitely many function–decomposition pairs consistent with them, including pairs in which alternative decompositions outperform any given decomposition.

Based on this characterization of decomposition, we derive a corollary for decomposition strategies:

**No decomposition strategy based solely on finite black-box observations can provide universally reliable performance guarantees.**

# 4 Implications for Decomposition Strategy Design

## 4.1 Rethinking Structure-Driven Decomposition Strategy

To illustrate the finite-information limitation more concretely, we construct a representative failure case of the structure-driven decomposition strategy. Specifically, we show that two black-box functions can produce the same finite evaluation history and induce the same decomposition, while favoring different decompositions in subsequent optimization.

Differential Grouping (DG) (Omidvar et al., 2013) is a representative interaction-based decomposition strategy that infers pairwise variable interactions from finite-difference observations. Consider two variables $x_i$ and $x_j$, and define the forward difference of $f$ with respect to $x_i$ as

$$\Delta_{\delta,x_i} f(\ldots, x_i + \delta, \ldots) - f(\ldots, x_i, \ldots), \tag{20}$$

where $\delta \neq 0$. If there exist $a$ and $b_1 \neq b_2$ such that

$$\Delta_{\delta,x_i} f|_{x_i=a, x_j=b_1} \neq \Delta_{\delta,x_i} f|_{x_i=a, x_j=b_2}, \tag{21}$$

then $x_i$ and $x_j$ are identified as interacting variables. DG applies this criterion to infer pairwise interactions and constructs a decomposition by placing interacting variables within the same subspace.

We construct two functions, $f_1$ and $f_2$, to expose the finite-information limitation of the above strategy. To this end, we define the auxiliary functions

$$z(t) = 1 - \exp\left[-\left(\frac{t - 0.2}{0.09}\right)^2\right], \tag{22}$$

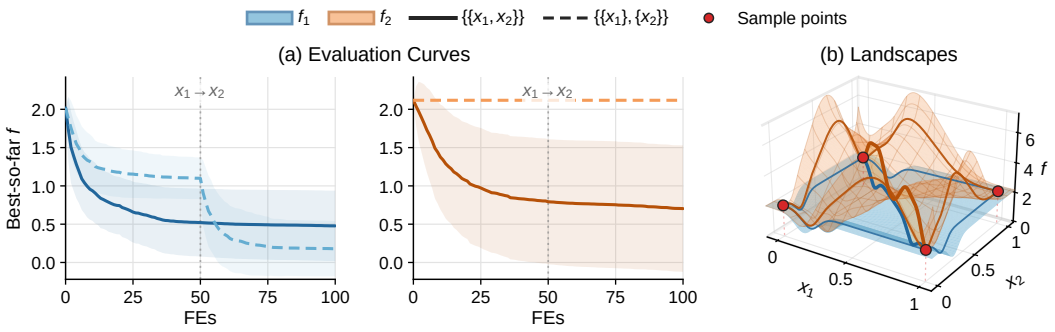


Figure 1: Functions sharing the same finite black-box observations can differ substantially elsewhere in the search space and thus favor different decompositions, yet the structure-driven decomposition strategy still assigns them the same decomposition.

and

$$w(x_1, x_2) = 8\left[x_1^2(1-x_1)^2 + x_2^2(1-x_2)^2\right]. \tag{23}$$

The two objective functions are then defined as

$$f_1(x_1, x_2) = z(x_1) + z(x_2) + 0.2(x_1 - 0.2)(x_2 - 0.2), \tag{24}$$

and

$$f_2(x_1, x_2) = f_1(x_1, x_2) + 20w(x_1, x_2)[(x_1 - x_2)^2 + 0.005]. \tag{25}$$

Consider $\delta = 1$, $a = 0$, $b_1 = 0$, and $b_2 = 1$, corresponding to the four query points $(1, 0)$, $(0, 0)$, $(1, 1)$, and $(0, 1)$. Since $w(x_1, x_2) = 0$ whenever $x_1, x_2 \in \{0, 1\}$, $f_1$ and $f_2$ coincide at all four points and therefore induce the same finite evaluation history under the DG probe. However, direct substitution shows that, for both functions, the two forward differences with respect to $x_1$ are distinct:

$$\Delta_{1,x_1} f|_{x_1=0,x_2=0} \neq \Delta_{1,x_1} f|_{x_1=0,x_2=1}. \tag{26}$$

Hence, DG identifies $x_1$ and $x_2$ as interacting and returns the same decomposition $\mathcal{G} = \{\{1, 2\}\}$.

To compare their subsequent optimization performance, we fix an optimization protocol $\mathcal{A}$ in which the subspaces are optimized sequentially using CMA-ES (Hansen & Ostermeier, 2001). A total budget of 100 function evaluations is allocated equally among the subspaces. In addition to the DG-induced decomposition $\mathcal{G} = \{\{1, 2\}\}$, we consider the alternative decomposition $\mathcal{G}' = \{\{1\}, \{2\}\}$. For each function and decomposition, we perform 200 independent runs using different seeds.

Fig. 1(a) shows that the two functions favor different decompositions. However, under the same finite black-box observations, DG returns the same decomposition for both, providing a concrete illustration of Theorem 2. Fig. 1(b) further illustrates the ambiguity characterized by Lemma 1. Although $f_1$ and $f_2$ coincide at all four queried points, they can differ substantially elsewhere in the search space. Notably, these two constructed functions represent only two members of the infinitely many functions in $\mathcal{F}(h)$ that are consistent with the same history $h$.

Beyond this finite-information limitation, the structure-driven decomposition strategy also faces a fundamental objective mismatch. Its objective is to recover a prescribed structural property, whereas the ultimate objective of decomposition is to improve downstream optimization performance. Even perfect structural recovery does not necessarily lead to more effective optimization. In other words,

**Structural recovery $\neq$ Optimization gain.**

This is a critical premise overlooked by the prevailing structure-driven decomposition strategy.

### 4.2 Rethinking Random Decomposition Strategy

Given these limitations, a natural question is which decomposition strategy should be adopted when finite information offers no reliable basis for preference. This motivates revisiting the random

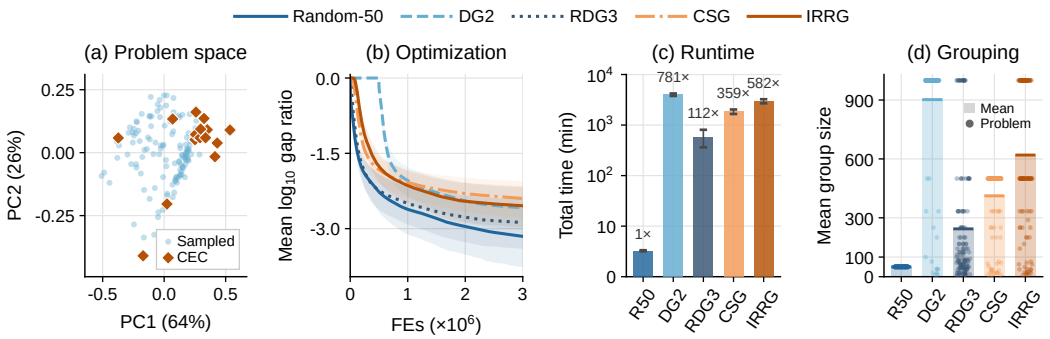


Figure 2: Across a broader problem space, random decomposition avoids the decomposition failures of structure-driven strategies and incurs no decomposition cost, making it highly competitive in both optimization performance and computational efficiency.

decomposition strategy. By imposing no explicit preference over decompositions, it naturally matches the uncertainty characterized by Theorem 2. This neutrality further yields a minimax-robust guarantee.

We consider a randomized decomposition-selection distribution $P = (p_1, \ldots, p_N)$ over $\mathfrak{D}_D = \mathcal{G}_1, \ldots, \mathcal{G}_N$, where $\sum_{j=1}^{N} p_j = 1$. We define the 0–1 decomposition-selection loss as

$$\ell(f, \mathcal{G} \mid h) = \begin{cases} 0, & \mathcal{G} \in \mathcal{G}_f^\star(h), \\ 1, & \text{otherwise.} \end{cases} \tag{27}$$

**Theorem 3: Finite-Information Minimax Robustness of Random Decomposition Strategy**

Given a finite evaluation history $h$ and a fixed optimization protocol $\mathcal{A}$, let $\mathfrak{D}_D = \{\mathcal{G}_1, \ldots, \mathcal{G}_N\}$. Suppose that, for every $\mathcal{G}_j \in \mathfrak{D}_D$, there exists $f_j \in \mathcal{F}(h)$ for which $\mathcal{G}_j$ is the unique optimal decomposition. Then

$$\min_{P} \sup_{f \in \mathcal{F}(h)} \mathbb{E}_{\mathcal{G} \sim P}[\ell(f, \mathcal{G} \mid h)] = 1 - \frac{1}{N}, \tag{28}$$

where the unique minimizer is the uniform distribution $p_j = 1/N$ for all $j$. Hence, the random decomposition strategy is uniquely minimax optimal under finite information.

The proof of Theorem 3 is provided in Appendix A.2. We further validate the performance of the random decomposition strategy empirically. To this end, we summarize existing large-scale global optimization benchmarks into a problem distribution and use Sobol sampling to generate 128 1000-dimensional problem instances as the evaluation set. We randomly partition the variables into disjoint 50-dimensional subspaces and compare this strategy against representative decomposition strategies, including DG2 (Omidvar et al., 2017), RDG3 (Sun et al., 2019), CSG (Tian et al., 2024b), and IRRG (Komarnicki et al., 2022). The optimization protocol $\mathcal{A}$ uses CMA-ES and sequentially samples one generation per subspace until the total budget of $3 \times 10^6$ function evaluations is exhausted. Each algorithm is independently run five times with different random seeds on each problem instance. Details are provided in Appendix A.3.

Fig. 2 summarizes the results. In Fig. 2(a), CEC2013LSGO (Li et al., 2013) problems occupy a relatively concentrated region, whereas our sampled instances are more broadly and evenly distributed, indicating greater diversity. Figs. 2(b) and (c) show that Random-50 is competitive in both optimization performance and, more notably, computational efficiency. Figs. 2(b) and (d) further reveal practical limitations of structure-driven decomposition beyond the objective mismatch identified above. First, decomposition can be costly: DG2 requires $O(D^2)$ function evaluations, producing the initial plateau in Fig. 2(b) and potentially becoming prohibitive under limited evaluation budgets. Second, structure-driven methods typically target specific structures, such as additive separability, and rely on sensitive thresholds, making them vulnerable to heterogeneous structures and numerical instability. As shown in Fig. 2(d), several methods consequently produce mean group sizes of

hundreds of dimensions, leading to substantially longer runtimes with limited performance gains. Detailed performance comparisons, including tables of $C_\epsilon$, are provided in Appendix A.4.

Taken together, the theoretical and empirical evidence suggests that the random decomposition strategy combines essentially zero cost with minimax robustness when reliable preference information is unavailable. This perspective revises the conventional view that random decomposition becomes ineffective as dimensionality increases because the probability of placing interacting variables in the same subspace tends to vanish (Van den Bergh & Engelbrecht, 2004).

## 5 Conclusion and Future Work

This work theoretically re-examines decomposition from an information-theoretic perspective and reveals the fundamental difficulty of identifying decompositions for efficient optimization from finite black-box observations. By establishing a general theoretical framework, we rigorously analyze decomposition in black-box optimization and explicitly characterize its search space and intrinsic difficulty. We then develop a non-identifiability theory showing that decompositions for efficient optimization cannot be reliably identified from finite black-box observations, and further that no decomposition strategy induced by such observations can provide reliable performance guarantees. Building on these theoretical results, we revisit both structure-driven and random decomposition strategies. Through theory-driven counterexamples, we show that prevailing structure-driven strategies overlook a critical premise: structural recovery does not necessarily translate into optimization gains. We further demonstrate, both theoretically and empirically, that random decomposition, often regarded as ineffective, can serve as a robust strategy under finite information.

Together, these findings suggest promising directions for future research:

1. **Decomposition is not the end; optimization is the ultimate goal.** Unlike surrogate objectives such as structural recovery, algorithm design should remain explicitly aligned with the ultimate goal of efficient optimization to avoid objective mismatch.
2. **Decomposition can be extended from a static, one-shot decision to a dynamic closed-loop process.** Rather than inferring a static decomposition from finite black-box observations, decomposition candidates can be continuously evaluated and adapted according to their actual performance during optimization. This provides direct performance information linking decompositions to optimization outcomes beyond the initial finite information.
3. **The optimization protocol deserves careful design to adapt to the chosen decomposition.** From the perspective of the overall optimization pipeline, the optimization protocol that manages subspace optimization can substantially affect final performance and should therefore be co-designed with the decomposition strategy. Given the enormous decomposition space and the difficulty of identifying effective decompositions, one practical approach is to adopt random decomposition and shift the design focus to the subsequent protocol, such as subspace optimization order and optimizer selection. Aligning the optimization protocol with the realized decomposition can likewise improve overall optimization performance.
4. **Develop broader black-box benchmark suites that incorporate diverse real-world application scenarios.** Current benchmarks often rely on a limited set of manually constructed composite problems. Their restricted scale and diversity can introduce benchmark-specific biases into decomposition strategy design, while their limited connection to real-world applications may hinder transfer to practical problems. A promising alternative is to move beyond fixed problem sets and construct simulated problem distributions grounded in real-world scenarios, from which diverse problem instances can be sampled for evaluation.

### Reproducibility Statement

Complete proofs of the theoretical results are provided in Appendices A.1- A.2. The construction of the evaluation problem distribution, experimental protocol, and detailed results are provided in Appendices A.3- A.4. All code required to reproduce the experiments is provided in our repository and the supplementary material.

## AI USE STATEMENT

In this work, generative AI tools were used to aid and polish writing, provide feedback on the logical consistency of mathematical proofs, and assist with the implementation and debugging of experimental code. All AI-assisted text, proof-related feedback, and code were reviewed and verified by the authors, and all mathematical claims, proofs, and experimental results were independently checked by the authors. We take responsibility for the final content of this work, including text, claims, code, and other artifacts produced with the aid of generative AI.

## REFERENCES

Wenxiang Chen, Thomas Weise, Zhenyu Yang, and Ke Tang. Large-scale global optimization using cooperative coevolution with variable interaction learning. In *International conference on parallel problem solving from nature*, pp. 300–309. Springer, 2010.

Kenneth A De Jong and Mitchell A Potter. Evolving complex structures via cooperative coevolution. In *Evolutionary Programming*, pp. 307–317, 1995.

Nikolaus Hansen and Andreas Ostermeier. Completely derandomized self-adaptation in evolution strategies. *Evolutionary computation*, 9(2):159–195, 2001.

Marcin Michal Komarnicki, Michal Witold Przewozniczek, Halina Kwasnicka, and Krzysztof Walkowiak. Incremental recursive ranking grouping for large-scale global optimization. *IEEE Transactions on Evolutionary Computation*, 27(5):1498–1513, 2022.

Xiaodong Li, Ke Tang, Mohammad N. Omidvar, Zhenyu Yang, and Kai Qin. Benchmark functions for the CEC'2013 special session and competition on large-scale global optimization. Technical report, Evolutionary Computation and Machine Learning Group, RMIT University, Melbourne, Australia, 2013.

Jinpeng Liu and Ke Tang. Scaling up covariance matrix adaptation evolution strategy using cooperative coevolution. In *International Conference on Intelligent Data Engineering and Automated Learning*, pp. 350–357. Springer, 2013.

Ya-Feng Liu, Tsung-Hui Chang, Mingyi Hong, Zheyu Wu, Anthony Man-Cho So, Eduard A Jorswieck, and Wei Yu. A survey of recent advances in optimization methods for wireless communications. *IEEE Journal on Selected Areas in Communications*, 42(11):2992–3031, 2024.

Xiaoliang Ma, Xiaodong Li, Qingfu Zhang, Ke Tang, Zhengping Liang, Weixin Xie, and Zexuan Zhu. A survey on cooperative co-evolutionary algorithms. *IEEE Transactions on Evolutionary Computation*, 23(3):421–441, 2018.

Lazo M Manojlović. Photometry-based estimation of the total number of stars in the universe. *Applied Optics*, 54(21):6589–6591, 2015.

Mohammad Nabi Omidvar, Xiaodong Li, Yi Mei, and Xin Yao. Cooperative co-evolution with differential grouping for large scale optimization. *IEEE Transactions on evolutionary computation*, 18(3):378–393, 2013.

Mohammad Nabi Omidvar, Ming Yang, Yi Mei, Xiaodong Li, and Xin Yao. Dg2: A faster and more accurate differential grouping for large-scale black-box optimization. *IEEE Transactions on evolutionary computation*, 21(6):929–942, 2017.

Mohammad Nabi Omidvar, Xiaodong Li, and Xin Yao. A review of population-based metaheuristics for large-scale black-box global optimization—part i. *IEEE Transactions on Evolutionary Computation*, 26(5):802–822, 2021.

Mitchell A Potter and Kenneth A De Jong. A cooperative coevolutionary approach to function optimization. In *International conference on parallel problem solving from nature*, pp. 249–257. Springer, 1994.

Wenjie Qiu, Hongshu Guo, Zeyuan Ma, and Yue-Jiao Gong. A novel two-phase cooperative coevolution framework for large-scale global optimization with complex overlapping. In *Proceedings of the Genetic and Evolutionary Computation Conference Companion*, pp. 507–510, 2025.

Wenjie Qiu, Hao Guo, Zeyuan Ma, and Yue-Jiao Gong. Advancing cma-es with learning-based cooperative coevolution for scalable optimization. In *Proceedings of the Genetic and Evolutionary Computation Conference*, pp. 928–937, 2026a.

Wenjie Qiu, Zixin Wang, Hongyu Fang, Zeyuan Ma, and Yue-Jiao Gong. A learning-based cooperative coevolution framework for heterogeneous large-scale global optimization. In *Proceedings of the Genetic and Evolutionary Computation Conference*, pp. 938–947, 2026b.

Wenjie Qiu, Zhenrong Weng, Yue-Jiao Gong, Wei-Neng Chen, and Jun Zhang. Unicc: A unified coevolutionary architecture with divergence-speedup modeling for large-scale global optimization. *IEEE Transactions on Evolutionary Computation*, 2026c.

Tapabrata Ray and Xin Yao. A cooperative coevolutionary algorithm with correlation based adaptive variable partitioning. In *2009 IEEE congress on evolutionary computation*, pp. 983–989. IEEE, 2009.

Qi Sun, Xuejun Xu, and Haotian Yi. Domain decomposition learning methods for solving elliptic problems. *SIAM Journal on Scientific Computing*, 46(4):A2445–A2474, 2024a.

Qinghan Sun, Huan Ma, Tian Zhao, Yonglin Xin, and Qun Chen. Break down the decentralization-security-privacy trilemma in management of distributed energy systems. *Nature Communications*, 15(1):4508, 2024b.

Yuan Sun, Xiaodong Li, Andreas Ernst, and Mohammad Nabi Omidvar. Decomposition for large-scale optimization problems with overlapping components. In *2019 IEEE congress on evolutionary computation (CEC)*, pp. 326–333. IEEE, 2019.

Maojiang Tian, Mingke Chen, Wei Du, Yang Tang, and Yaochu Jin. An enhanced differential grouping method for large-scale overlapping problems. *IEEE Transactions on Evolutionary Computation*, 29(6):2272–2286, 2024a.

Maojiang Tian, Minyang Chen, Wei Du, Yang Tang, Yaochu Jin, and Gary G Yen. A composite decomposition method for large-scale global optimization. *IEEE Transactions on Artificial Intelligence*, 5(12):6304–6319, 2024b.

Maojiang Tian, Wei Du, Wenxuan Fang, Yang Tang, and Yaochu Jin. Learning to decompose and optimize for large-scale overlapping problems. *IEEE Transactions on Evolutionary Computation*, 2025.

Marcelo DT Torres, Yimeng Zeng, Fangping Wan, Natalie Maus, Jacob Gardner, and Cesar de la Fuente-Nunez. A generative artificial intelligence approach for peptide antibiotic optimization. *Nature Machine Intelligence*, pp. 1–16, 2026.

Frans Van den Bergh and Andries P Engelbrecht. A cooperative approach to particle swarm optimization. *IEEE transactions on evolutionary computation*, 8(3):225–239, 2004.

Ye Wei, Bo Peng, Ruiwen Xie, Yangtao Chen, Yu Qin, Peng Wen, Stefan Bauer, Po-Yen Tung, and Dierk Raabe. Deep active optimization for complex systems. *Nature Computational Science*, 5(9): 801–812, 2025.

David H Wolpert. The lack of a priori distinctions between learning algorithms. *Neural computation*, 8(7):1341–1390, 1996.

David H Wolpert and William G Macready. No free lunch theorems for optimization. *IEEE transactions on evolutionary computation*, 1(1):67–82, 1997.

Xin Yao and Siang Yew Chong. *Coevolutionary Computation and Its Applications*. Springer, 2025.

# A Appendix

## A.1 Proof of Theorem 1

*Proof.* Let $\mathcal{U}$ denote the set of all collections of nonempty subsets of $[D]$, without requiring full coverage. Since $[D]$ has $2^D - 1$ nonempty subsets, each of which can be either included or excluded,

$$|\mathcal{U}| = 2^{\,2^D - 1}. \tag{29}$$

For each $i \in [D]$, define

$$A_i = \left\{ \mathcal{G} \in \mathcal{U} : i \notin \bigcup_{G \in \mathcal{G}} G \right\}, \tag{30}$$

where $A_i$ contains the collections that do not cover variable $i$. Hence,

$$\mathfrak{D}_D = \mathcal{U} \setminus \bigcup_{i=1}^{D} A_i. \tag{31}$$

For any $S \subseteq [D]$ with $|S| = k$, every collection in $\bigcap_{i \in S} A_i$ can contain only nonempty subsets of the remaining $D - k$ variables. Since there are $2^{D-k} - 1$ such subsets,

$$\left| \bigcap_{i \in S} A_i \right| = 2^{\,2^{D-k} - 1}. \tag{32}$$

Applying the inclusion–exclusion principle gives

$$|\mathfrak{D}_D| = \sum_{k=0}^{D} (-1)^k \binom{D}{k} 2^{\,2^{D-k} - 1}. \tag{33}$$

For the asymptotic result, write

$$|\mathfrak{D}_D| = 2^{\,2^D - 1} + R_D, \tag{34}$$

where

$$|R_D| \le \sum_{k=1}^{D} \binom{D}{k} 2^{\,2^{D-k} - 1}. \tag{35}$$

Since $2^{D-k} \le 2^{D-1}$ for every $k \ge 1$,

$$|R_D| \le (2^D - 1)\, 2^{\,2^{D-1} - 1}. \tag{36}$$

Therefore,

$$\frac{|R_D|}{2^{\,2^D - 1}} \le (2^D - 1) 2^{-2^{D-1}} \to 0, \qquad D \to \infty. \tag{37}$$

Hence,

$$|\mathfrak{D}_D| \sim 2^{\,2^D - 1}. \tag{38}$$

■

## A.2 Proof of Theorem 3

*Proof.* By assumption, for each $\mathcal{G}_j \in \mathfrak{D}_D$, there exists a function $f_j \in \mathcal{F}(h)$ for which $\mathcal{G}_j$ is the unique optimal decomposition. Under a randomized decomposition-selection distribution $P = (p_1, \ldots, p_N)$, the expected loss on $f_j$ is therefore

$$\mathbb{E}_{\mathcal{G} \sim P} \left[ \ell(f_j, \mathcal{G} \mid h) \right] = 1 - p_j. \tag{39}$$

Hence,

$$\sup_{f \in \mathcal{F}(h)} \mathbb{E}_{\mathcal{G} \sim P} \left[ \ell(f, \mathcal{G} \mid h) \right] \ge \max_{1 \le j \le N} (1 - p_j) = 1 - \min_{1 \le j \le N} p_j. \tag{40}$$

Since $\sum_{j=1}^{N} p_j = 1$,

$$\min_{1 \le j \le N} p_j \le \frac{1}{N}, \tag{41}$$

which gives

$$\sup_{f \in \mathcal{F}(h)} \mathbb{E}_{\mathcal{G} \sim P} \left[ \ell(f, \mathcal{G} \mid h) \right] \ge 1 - \frac{1}{N}. \tag{42}$$

Now consider the uniform distribution, $p_j = 1/N$ for all $j$. For any $f \in \mathcal{F}(h)$, at least one decomposition belongs to $\mathcal{G}_f^\star(h)$. Therefore,

$$\mathbb{E}_{\mathcal{G} \sim P} \left[ \ell(f, \mathcal{G} \mid h) \right] = 1 - \sum_{\mathcal{G}_j \in \mathcal{G}_f^\star(h)} p_j \le 1 - \frac{1}{N}. \tag{43}$$

Thus, the lower bound $1 - 1/N$ is attained.

Finally, attaining this bound requires $\min_j p_j = 1/N$. Together with $\sum_{j=1}^{N} p_j = 1$, this is possible only if

$$p_1 = \cdots = p_N = \frac{1}{N}. \tag{44}$$

Hence, the uniform randomized decomposition-selection distribution is the unique minimax-optimal distribution under the defined loss. ■

### A.3 Details of the Evaluation Problem Distribution

**Distribution.** By summarizing and analyzing representative benchmark suites such as CEC2013LSGO (Li et al., 2013) and BNS (Tian et al., 2024b), we construct a problem distribution that generates objective functions on $[-100, 100]^D$. A profile $\mathbf{p} = (p_g, p_o, p_\kappa, p_r, p_c, p_h, p_b) \in [0, 1]^7$ parameterizes subspace granularity, overlap, conditioning, ruggedness, overlap incidence strength, between-subspace heterogeneity, and contribution imbalance, respectively. Given a profile, a random partition covering all coordinates is augmented with overlapping memberships to form generative hyperedges $S_1, \ldots, S_K$. These subspaces are used only for problem generation and are not assumed to represent true interaction groups. The resulting objective is

$$f(\mathbf{x}) = f^\star + 10^\alpha \sum_{k=1}^{K} w_k \, \ell_k(\mathbf{a}_k \odot \mathbf{z}_{S_k}), \qquad \mathbf{z} = \frac{\mathbf{x} - \mathbf{x}^\star}{100}. \tag{45}$$

Here, $\mathbf{x}^\star$ is sampled uniformly from $[-80, 80]^D$, $\alpha$ from $[-1, 1]$, and $f^\star$ from $[-1000, 1000]$. The incidence vector $\mathbf{a}_k$ assigns weight one to primary memberships and $p_c$ to added memberships, while $\odot$ denotes elementwise multiplication. The positive weights $w_k$ are normalized lognormal draws summing to one. Each $\ell_k$ applies positive diagonal scaling and an orthogonal butterfly rotation, followed by a Dirichlet-weighted mixture of six origin-centered primitives: quadratic, different powers, Rosenbrock, Rastrigin, Ackley, and Griewank. Primitive and mixture scales are normalized using sample means over 32 common Sobol reference points within the feasible box to reduce differences in raw magnitude. Complete primary coverage and the unique zero of each nonnegative local landscape ensure that $\mathbf{x}^\star$ is the unique global minimizer with objective value $f^\star$.

**PCA visualization.** We compare the two benchmarks through differences in their observable fitness landscapes, characterized directly from function evaluations under a common black-box probing protocol. Probe locations and perturbation magnitudes are defined relative to each problem's native search box, enabling comparable observations without modifying the original problem definitions. From the resulting evaluations, we construct six landscape descriptors. Along sampled directions, ruggedness is measured by the loss of rank correlation between adjacent evaluations, while multimodality is quantified by the frequency of sampled local minima. Anisotropy is characterized by the dispersion of symmetric directional responses. Single-variable perturbations measure contribution imbalance across variables, whereas joint perturbations quantify interaction through deviations from additive effects. Finally, heterogeneity is measured by the variation in local response magnitudes

across reference locations. These descriptors provide a common empirical characterization of both benchmarks based solely on observable function responses.

To make descriptors with different numerical ranges comparable, each is mapped to $[0, 1]$ using fixed transformations. Ruggedness is divided by two, while the nonnegative and unbounded anisotropy and heterogeneity descriptors are transformed by $q/(1 + q)$; the remaining three descriptors already lie in $[0, 1]$.

We then use PCA to summarize the six-dimensional descriptor space in two dimensions. The PCA center and directions are fitted using only the 128 instances, and the 15 CEC2013LSGO functions are projected using the same transformation.

**Mean $\log_{10}$ gap ratio.** For algorithm $a$, problem $i$, and run $r$, let $b_{air}(t)$ be the optimization incumbent after $t$ charged function evaluations, and let $b_{ir}^{(0)}$ be the best of the 32 initial points shared across algorithms. Using the known optimum value $f_i^\star$ only for reporting, define

$$g_{air}(t) = \frac{b_{air}(t) - f_i^\star}{b_{ir}^{(0)} - f_i^\star}, \qquad s_{air}(t) = \log_{10}\big(\max\{g_{air}(t), 10^{-12}\}\big). \tag{46}$$

The curve averages $s_{air}(t)$ over actual runs within each problem, then equally over all 128 problems. Lower values indicate better progress; zero denotes the initialization reference.

**Mean group size.** For each run, let $K$ denote the number of variable groups used for optimization and $d_k$ the number of variables in group $k$. The mean group size is $\bar{d} = K^{-1} \sum_{k=1}^{K} d_k$. For each problem, we average $\bar{d}$ over its actual runs. Each dot represents this problem-level average, and each bar is the equal-weight mean of these values over the 128 problems.

### A.4 Details of Experimental Results

Table 1: Comparison of $L_\epsilon$ across different decompositions on $f_1$.

| Target accuracy $\epsilon$ | $\{\{x_1, x_2\}\}$ | $\{\{x_1\}, \{x_2\}\}$ |
|---|---|---|
| 1.0 | **10** | **10** |
| 0.9 | **15** | 20 |
| 0.8 | **20** | **20** |
| 0.7 | 25 | **20** |
| 0.6 | 35 | **20** |
| 0.5 | 65 | **30** |
| 0.4 | $\infty$ | **40** |
| 0.3 | $\infty$ | **50** |
| 0.2 | $\infty$ | **100** |

For $f_1$, the optimal objective value is $f^\star = 0$.

Table 2: Comparison of $L_\epsilon$ across different decompositions on $f_2$.

| Target accuracy $\epsilon$ | $\{\{x_1, x_2\}\}$ | $\{\{x_1\}, \{x_2\}\}$ |
|---|---|---|
| 1.0 | **21** | $\infty$ |
| 0.9 | **28** | $\infty$ |
| 0.8 | **38** | $\infty$ |
| 0.7 | **82** | $\infty$ |

For $f_2$, the optimal objective value is $f^\star = 0$.

Table 3: Comparison of Evaluation Costs Required by Different Decomposition Strategies to Reach Target Mean $\log_{10}$ Gap Ratios on the Sampled Problems.

| Gap ratio | Random-50 | DG2 | RDG3 | CSG | IRRG |
|---|---|---|---|---|---|
| $-0.50$ | **0.012** | 0.539 | 0.038 | 0.105 | 0.150 |
| $-0.75$ | **0.043** | 0.563 | 0.060 | 0.135 | 0.188 |
| $-1.00$ | **0.061** | 0.585 | 0.086 | 0.165 | 0.225 |
| $-1.25$ | **0.090** | 0.615 | 0.113 | 0.210 | 0.285 |
| $-1.50$ | **0.128** | 0.653 | 0.150 | 0.278 | 0.368 |
| $-1.75$ | **0.188** | 0.735 | 0.216 | 0.428 | 0.495 |
| $-2.00$ | **0.308** | 0.938 | **0.308** | 0.780 | 0.743 |
| $-2.25$ | **0.488** | 1.418 | 0.518 | 1.692 | 1.275 |
| $-2.50$ | **0.780** | 2.228 | 0.998 | $\infty$ | 2.423 |
| $-2.75$ | **1.395** | $\infty$ | 1.875 | $\infty$ | $\infty$ |
| $-3.00$ | **2.190** | $\infty$ | $\infty$ | $\infty$ | $\infty$ |

Evaluation costs are reported in millions of function evaluations.

Table 4: Comparison of $C_\epsilon$ across Different Decomposition Strategies on the Sampled Problems.

| Target accuracy $\epsilon$ | Random-50 | DG2 | RDG3 | CSG | IRRG |
|---|---|---|---|---|---|
| 1.0 | **0.0113** | 0.5336 | 0.0351 | 0.1091 | 0.2132 |
| 0.9 | **0.0132** | 0.5376 | 0.0376 | 0.1160 | 0.2213 |
| 0.8 | **0.0138** | 0.5423 | 0.0408 | 0.1233 | 0.2292 |
| 0.7 | **0.0174** | 0.5478 | 0.0471 | 0.1319 | 0.2382 |
| 0.6 | **0.0184** | 0.5568 | 0.0458 | 0.1518 | 0.2494 |
| 0.5 | **0.0213** | 0.5584 | 0.0484 | 0.1481 | 0.2691 |
| 0.4 | **0.0284** | 0.5603 | 0.0554 | 0.1454 | 0.2333 |
| 0.3 | **0.0329** | 0.5870 | 0.0671 | 0.1804 | 0.2480 |
| 0.2 | **0.0373** | 0.5967 | 0.0761 | 0.1878 | 0.2833 |
| 0.1 | **0.0464** | 0.6199 | 0.0925 | 0.2540 | 0.3185 |

The optimal objective value is $f^\star = 0$ for all sampled problems, and $C_\epsilon$ is reported in millions of function evaluations.

Table 5: Runtime comparison of decomposition strategies on the sampled problems at $3 \times 10^6$ FEs.

| Strategy | Mean total time (min) | Relative time |
|---|---|---|
| Random-50 | **5.13** | **1.00**× |
| DG2 | 4,005.42 | 781.08× |
| RDG3 | 575.12 | 112.15× |
| CSG | 1,839.86 | 358.78× |
| IRRG | 2,983.81 | 581.86× |

Table 6: Comparison of Mean Group Size and Across-Problem Standard Deviation across Decomposition Strategies on Sampled Problems

| Strategy | Mean group size | Standard Deviation |
|---|---|---|
| Random-50 | **50.00** | **0.00** |
| DG2 | 902.10 | 276.12 |
| RDG3 | 243.72 | 307.74 |
| CSG | 411.93 | 168.31 |
| IRRG | 619.74 | 363.56 |