A Bi-Surrogate Scheme for Black-Box Optimization with Unrevealed Constraints

Chi-En Tang*, Chen Chien*, Tian-Li Yu
Taiwan Evolutionary Intelligence Lab
Department of Electrical Engineering, National Taiwan University
Taipei, Taiwan
Email: r13921043@ntu.edu.tw, r13921095@ntu.edu.tw, tianliyu@ntu.edu.tw

***Abstract*—Surrogate-assisted evolutionary algorithms (SAEAs) are widely used on black-box optimization problems involving computationally intensive function evaluations, such as physics-based simulations that embed complex constraints. In such scenarios, the simulator usually integrates a repair mechanism that modifies the candidate solution handed by the optimizer prior to evaluation. Under these conditions, traditional SAEAs cannot accurately capture the complex constraints in the simulators since some of them are unrevealed to the EA. To address this opacity, this paper proposes a bi-surrogate architecture that explicitly models both the repair operator and the fitness evaluation. This framework characterizes the repair behavior to capture unrevealed constraint information overlooked by traditional approaches. Experiment results demonstrate that the proposed architecture significantly reduces the number of function evaluations while maintaining competitive success rates. Notably, this advantage becomes more pronounced as simulator costs increase, making the approach well-suited for computationally intensive applications.**

***Index Terms*—black-box optimization, constraint handling, genetic algorithms, surrogate-assisted evolutionary algorithms**

I. Introduction

In the realm of black-box optimization involving computationally intensive evaluations, surrogate-assisted evolutionary algorithms (SAEAs) have emerged as the preferable framework [1]. Prominent examples of such compute-intensive evaluations include finite-element analysis (FEA) in structural topology optimization [2], [3], classifier training within biomedical feature selection pipelines [4], [5], and full-wave electromagnetic (EM) simulations in metasurface design [6], [7]. These simulators incorporate complex constraints, and when they are combined with evolutionary optimization, these constraints are usually overly simplified in the fitness evaluation. This is typically because SAEAs employ statistical models, like Kriging, or machine learning models, like radial basis function neural networks, to approximate the objective function by learning a direct mapping from candidate solutions to objective values.

In practice, the simulator usually consists of two distinct mechanisms: (1) Repairing the solution candidate handed by the optimizer to restore feasibility, and (2) evaluating the repaired solution candidate. This mechanism is observed across various domains. In truss topology optimization, Richardson *et al.* [2] introduced kinematic stability repair, which modifies kinematically unstable configurations to achieve structural stability before FEA. Similar two-phase requirements occur in other domains, including repairing spurious modes in continuum topology [3], resolving non-manufacturable artifacts in metasurface design [8], and performing feature extraction to avert degraded biomedical classification [5].

In such engineering scenarios, the simulators embedding complex constraints are frequently proprietary software governed by commercial licenses, where the specific internal constraints and repair criteria are inaccessible. Consequently, traditional SAEAs learn conflated mappings that fail to capture non-smooth landscape features, resulting in search inefficiency. Although previous works [9] have predicted viability probabilities to circumvent failure-prone regions, they underutilize the structural information encoded within the repair process. To address this opacity, this paper proposes a bi-surrogate architecture that provides low-cost alternatives to the expensive evaluation pipeline, capturing information that conventional SAEAs fail to utilize. The architecture explicitly models both the repair process and the evaluation function, with genetic algorithms (GA) serving as the optimizer. Experiments are conducted on problem instances with varying evaluation costs, measuring success rates and the number of function evaluations (NFE). Results demonstrate that the bi-surrogate architecture significantly reduces the required evaluations, thereby achieving higher success rates under limited NFE compared to the approach that relies solely on the expensive evaluation pipeline without surrogate assistance.

II. Methodology

This work addresses black-box optimization problems where objective evaluations necessitate invoking computationally intensive simulators. In practical engineering scenarios, high-fidelity simulators often operate under proprietary licenses or encapsulate domain-specific repair logic, leaving certain constraints unrevealed to the optimizer. Consequently, the optimizer cannot explicitly model these constraints and must rely on the simulator to transform candidate solutions into feasible configurations.

*Chi-En Tang and Chen Chien are co-first authors.

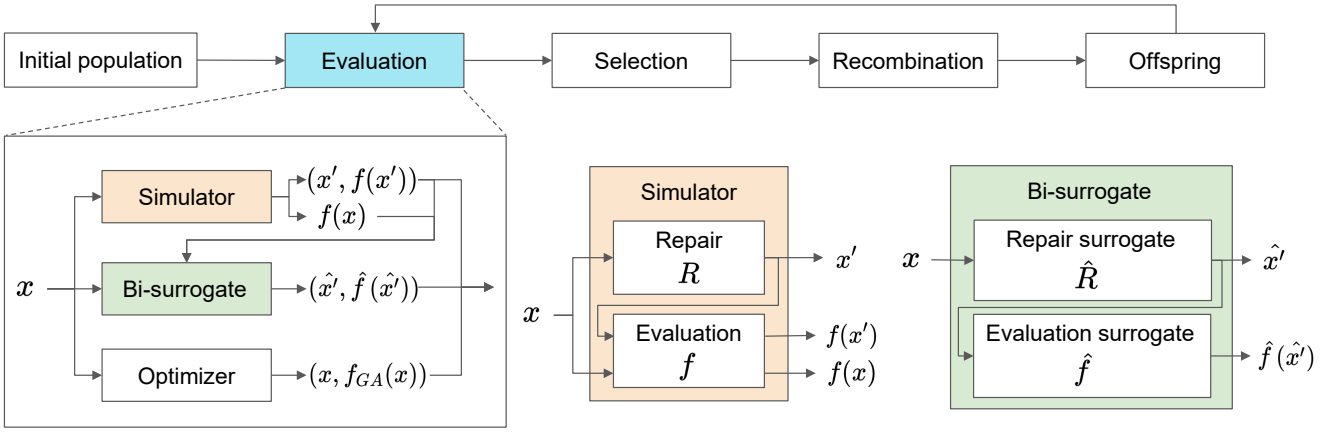

Fig. 1. Overview of the bi-surrogate framework integrated into the evolutionary optimization workflow. Top: Standard GA loop with the evaluation phase highlighted. Bottom left: The evaluation module dispatches solutions to one of three paths based on model availability. Bottom middle: The simulator performs repair (R) and evaluation (f). Bottom right: The bi-surrogate approximates repair ($\hat{R}$) and evaluation ($\hat{f}$) to mitigate the computational burden of the simulator.

To capture this dependency, the simulator is formulated as an evaluation pipeline that performs two operations upon receiving an input candidate x, as illustrated in Fig. 1:

1) Repair: The solution is refined via a local search process leveraging the simulator's internal constraint knowledge to produce a feasible configuration x'.
2) Evaluation: Both the original candidate and the repaired configuration are evaluated, yielding $f(x)$ and $f(x')$.

While this pipeline ensures feasibility, its repeated invocation during an evolutionary search is often computationally prohibitive. To mitigate this cost, a bi-surrogate framework is proposed to emulate the repair and evaluation processes through two specialized models comprising a repair surrogate $\hat{R}$ and an evaluation surrogate $\hat{f}$.

As depicted in Fig. 1, each candidate is routed through one of three paths during optimization based on model validity. Before the surrogates are sufficiently trained, the optimizer relies on a simplified fitness function f_{GA}, which excludes unrevealed constraints. Once the surrogates achieve reliability, the workflow prioritizes them for repair and evaluation tasks, reserving the high-fidelity simulator for model validation and data accumulation. This design decouples expensive simulations from the evolutionary search, enabling efficient exploration by delegating compute-intensive operations to inexpensive surrogates.

A. Components of the bi-surrogate framework

As shown in Fig. 1, the bi-surrogate architecture comprises two complementary models. The repair surrogate approximates the simulator's internal repair mechanism, learning the mapping $x \to x'$ from an initial solution to its feasible configuration. The evaluation surrogate approximates the objective function, providing performance estimates without invoking the compute-intensive simulator. By modeling both mappings, the framework captures constraint knowledge embedded in the repair process, enabling accurate fitness estimation.

1) Evaluation surrogate:

This study employs Bayesian neural networks (BNNs) as the evaluation surrogate. BNNs demonstrate efficacy in high-dimensional discrete search spaces with data sparsity [10], and their probabilistic uncertainty quantification is critical for balancing exploration and exploitation under limited training data [12]. This contrasts with deterministic alternatives such as support vector regression [11], which assume landscape continuity and lack built-in uncertainty estimates.

Conventional surrogate approaches neglect the representational discrepancy between the pre-repair input x and the post-repair output x', complicating the construction of an effective training set. Mapping x directly to $f(x')$ disregards this divergence, while training exclusively on repaired pairs $(x', f(x'))$ biases the model toward high-fitness regions and compromises discriminative capability. To address this, the proposed BNN is trained on the union of original and repaired datasets, denoted as $\{(x, f(x))\} \cup \{(x', f(x'))\}$. This strategy captures the global fitness landscape while preserving input-output correspondence, yielding the approximation $\hat{f}$.

The network architecture comprises three fully connected layers with 128, 64, and 1 neurons, respectively. Following the methodology for uncertainty estimation in deep learning [12], the network employs ReLU activations with a dropout rate of 0.2 applied during both training and inference. Optimization is performed using the Adam optimizer with early stopping to prevent overfitting. As illustrated in Fig. 2, the evolutionary search is initially driven by the formulated objective f_{GA} to address cold-start data scarcity. The surrogate $\hat{f}$ assumes evaluation duties once sufficient training data ensures approximation fidelity.

2) Repair surrogate:

Unlike standard scalar regression, approximating the repair operator requires reconstructing the entire high-dimensional solution vector to generate $\hat{x}'$, which serves as an approximation of the repair mapping. To address this, a k-nearest neigh-

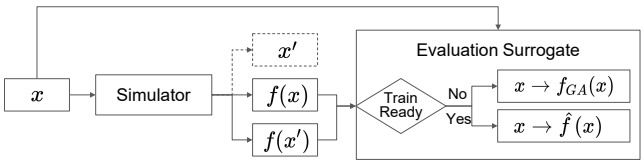

Fig. 2. Architecture of the evaluation surrogate. The surrogate $\hat{f}$ is trained on $\{(x, f(x))\} \cup \{(x', f(x'))\}$ to approximate f. Evaluation uses f_{GA} before the evaluation surrogate is ready, then transitions to $\hat{f}$.

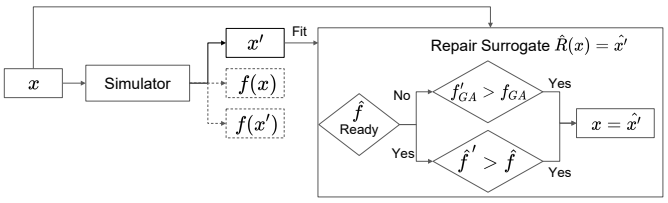

Fig. 3. Architecture of the repair surrogate. The model constructs $\hat{x}'$ via bitwise majority voting based on k-nearest neighbors. Acceptance is governed by f_{GA} before the evaluation surrogate $\hat{f}$ is ready, then transitions to $\hat{f}$.

bors (k-NN) approach using Hamming distance is employed. This methodological strategy is substantiated by the efficacy of k-NN in high-dimensional discrete surrogate modeling, as demonstrated by Chen *et al.* [13]. Preliminary experiments revealed that complex neural architectures such as multi-layer perceptrons and transformers tend to capture superficial fitness correlations rather than true repair dynamics. In contrast, k-NN simply and efficiently exploits the local correlation structure imposed by interdependent constraints.

In contrast, k-NN exploits the local correlation structure among interdependent decision variables with simplicity and computational efficiency. For a query x, the predicted repair $\hat{x}'$ is generated via bitwise majority voting among the repaired configurations of its k nearest neighbors, with ties preserving the corresponding original bit of x.

As illustrated in Fig. 3, predicted repairs $\hat{x}'$ undergo validation. If the evaluation surrogate $\hat{f}$ lacks sufficient training data, f_{GA} serves as the criterion, accepting replacements only if $f_{\text{GA}}(\hat{x}') > f_{\text{GA}}(x)$. Once reliable, validation shifts to $\hat{f}$, requiring $\hat{f}(\hat{x}') > \hat{f}(x)$.

B. Bi-surrogate DSMGA-II algorithm

To effectively address the strong variable dependency, the proposed framework is integrated with dependency structure matrix genetic algorithm II (DSMGA-II) [14]. This optimizer relies on linkage learning to identify and preserve interdependent decision variables through restricted mixing and back mixing operations [15].

At each generation, the population is partitioned by a repair ratio r. A subset is evaluated by the full simulator, aggregating tuples into archive $\mathcal{A}$ to train surrogates $\hat{R}$ and $\hat{f}$. For the remaining candidates, $\hat{R}$ proposes a repaired candidate $\hat{x}'$, which is accepted based on the conditional validation

Algorithm 1 Bi-surrogate DSMGA-II

Item: P: population, $\hat{R}$: repair surrogate, $\hat{f}$: evaluation surrogate, S: simulator, r: repair ratio, p: top-p ratio, A: archive
Input: N: population size
Output: best solution found x^*

$P \leftarrow$ InitializePopulation(N)
$A \leftarrow \emptyset$ ▷ Repaired dataset
while stopping criterion not met **do**
  $P_{\text{eval}} \leftarrow$ SelectForRealEvaluation(P, r)
  for each $x \in P_{\text{eval}}$ **do**
    $(f(x), x', f(x')) \leftarrow S(x)$
    $A \leftarrow A \cup \{(x, x', f(x), f(x'))\}$
  ready $\leftarrow$ TrainSurrogates($\hat{R}, \hat{f}, A$)
  for each $x \in P \setminus P_{\text{eval}}$ **do**
    $\hat{x}' \leftarrow \hat{R}(x)$ ▷ Predict candidate
    $\tilde{f} \leftarrow$ *ready* ? $\hat{f} : f_{\text{GA}}$
    if $\tilde{f}(\hat{x}') > \tilde{f}(x)$ **then**
      $x \leftarrow \hat{x}'$
    $\hat{f}(x) \leftarrow$ *ready* ? $\hat{f}(x) : f_{\text{GA}}(x)$ ▷ Predict fitness
  $T \leftarrow$ TournamentSelection(P)
  $D \leftarrow$ UpdateDSM(T)
  for $i = 1$ to $|P|$ **do**
    $P_i \leftarrow$ RestrictedMixing($P_i, D, \hat{f}$)
    $P \leftarrow$ BackMixing($P_i, \hat{f}$)
$P_{\text{top}} \leftarrow$ SelectTopBySurrogate($P, \hat{f}, p$)
for each $x \in P_{\text{top}}$ **do**
  $(f(x), \cdot, \cdot) \leftarrow S(x)$
  $A \leftarrow A \cup \{(x, \cdot, f(x), \cdot)\}$
return $arg\,max_{(x,x',f(x),f(x'))\in A} max(f(x), f(x'))$

mechanism. The candidate's fitness is then evaluated by either $\hat{f}$ or f_{GA}, depending on model readiness. Following this, standard DSMGA-II operations proceed. Upon termination, the top-p solutions undergo high-fidelity simulator evaluation to guarantee validity, returning the best solution found.

III. Experiment results[1]

This section presents a comprehensive evaluation of the proposed bi-surrogate framework on a complex optimization problem. The analysis outlines the testbench specification, maps the spin glass model to the target optimization scenario, and presents the criteria for selecting representative problem instances. Subsequently, the experiment setup is described, covering the definition of NFE, surrogate configurations, and baseline methods. The section concludes with an examination of performance under varying computational regimes of the simulator, highlighting the trade-off between solution quality and computational cost.

A. Testbench specification

The problem is modeled via a two-dimensional spin glass, where a configuration $x \in \{-1, +1\}^n$ represents n spins on

[1]The source code is available at https://github.com/jj-jane/DSMGA-II-Surrogate.

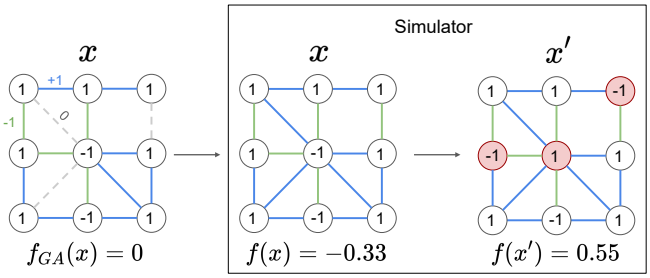

Fig. 4. Optimizer versus simulator view on a 3×3 spin glass. Blue and green edges denote ferromagnetic and antiferromagnetic couplings, respectively ($J_{ij} \in \{-1, +1\}$); gray dashes indicate unrevealed interactions ($J_{ij} = 0$). The left panel shows solution x evaluated under f_{GA}, yielding fitness 0. The center panel shows the same solution evaluated under f, revealing real fitness -0.33. The right panel shows the repaired solution x' after the simulator flips highlighted spins, achieving improved fitness 0.55.

TABLE I
PARAMETER SETTINGS FOR THE FOUR SCHEMES, WITH r AND p DENOTING THE REPAIR RATIO AND TOP-p VALIDATION RATE.

Scheme	r	p	Configuration
Baseline (Simu-based)	1.0	1.00	Simulator only
Evaluation surrogate	0.5	0.05	BNN ($R^2 \geq 0.9$, epochs ≈ 50)
Repair surrogate	0.5	0.05	k-NN (max sample $= 2.5rN$)
Bi-surrogate	0.5	0.05	Repair + Evaluation surrogate

TABLE II
WEIGHT CONFIGURATIONS FOR COMPUTATIONALLY INEXPENSIVE AND EXPENSIVE SIMULATION REGIMES.

Regime	w_1 (N_{GA})	w_2 (N_{rep})	w_3 (N_{eval})	w_4 (N^s_{rep})	w_5 (N^s_{eval})
Inexpensive	1	8	4	2	2
Expensive	1	800	400	2	2

a lattice. The system constraints map to pairwise interactions with coupling strengths $J_{ij} \in \{-1, +1\}$.

In real-world engineering scenarios, intrinsic performance is evaluated through high-fidelity simulators that operate as black-box functions. In the spin glass formulation, the simulator function $f(x)$ computes the objective value by summing the coupling strengths $J_{ij}x_ix_j$ over the complete interaction set E, in accordance with the Ising Hamiltonian [16].

Since exhaustively characterizing the underlying problem is infeasible, the evolutionary search relies on an objective function derived from explicitly known interactions. Consequently, the optimizer is confined to a subset $E_{\text{known}} \subset E$, and constructs its objective $f_{\text{GA}}(x)$ over this restricted set. This exclusion effectively assigns zero coupling strength to unrevealed pairs.

Fig. 4 illustrates the discrepancy between the optimizer's partial view and the simulator's complete evaluation. A solution optimized under f_{GA} may yield a different fitness when evaluated under f, as the latter incorporates unrevealed interactions. The figure also demonstrates the simulator's repair operation, which relaxes unrevealed conflicts, producing a repaired configuration x' with improved target fitness.

B. Selection of problem instances

The experiment analysis utilizes spin glass instances characterized by varying system sizes $\ell \in \{100, 196, 400\}$ and the number of unrevealed interactions, denoted by u, which govern the optimization difficulty. To generate these instances, the specific interactions to be unrevealed are randomly selected from the complete interaction set.

Instance selection is governed by two criteria to ensure a rigorous assessment of the bi-surrogate framework:

- Standard DSMGA-II: The instances must exhibit sufficient complexity to render the standard DSMGA-II ineffective in the absence of the repair mechanism.
- Simulator-based DSMGA-II: The instances necessitate the simulator to achieve high success rates, yet demand a substantial number of generations to locate the global optimum, thereby imposing a significant computational burden.

Three instances satisfying both criteria are selected: $(\ell, u, \text{ID}) \in \{(196, 18, 8), (196, 20, 93), (400, 20, 53)\}$. Across these instances, standard DSMGA-II achieves success rates of 43.0%, 76.3%, and 44.2%, averaging over 360 generations. In contrast, the simulator-based scheme achieves significantly higher success rates of 95.2%, 99.6%, and 86.2%, averaging approximately 120 generations. This gap underscores the necessity of the simulator on these challenging instances. Nevertheless, the simulator-based scheme still imposes a prohibitive computational burden given the high expense of simulator evaluations. Thus, these instances constitute an ideal benchmark for demonstrating the proposed framework's capacity to minimize computational overhead while maintaining optimization quality.

C. Experiment setup

The bi-surrogate framework is benchmarked against three comparative schemes, comprising baselines and ablation variants to validate its effectiveness. All schemes utilize DSMGA-II as the backbone, sharing a standardized configuration: Population size $N = 1000$, a cap of 1000 generations, and 1000 independent runs per instance. Table I summarizes the specific hyperparameters for each scheme, which were empirically tuned via preliminary experiments.

The **baseline** relies exclusively on the simulator for both repair and evaluation. The **evaluation surrogate** employs BNNs to approximate fitness evaluations, utilizing a coefficient of determination (R^2) threshold of 0.9 to determine model validity. The **repair surrogate** utilizes a k-NN model with a training buffer limited to the number of repaired solutions generated over 2.5 generations. This constraint prevents outdated historical data from misleading the evolutionary search. The **bi-surrogate** integrates both surrogate modules.

In addition to parameter settings, computational cost constitutes a critical performance metric. Each scheme is quantified by the weighted total NFE, defined as $\text{TotalNFE} = w_1N_{\text{GA}} + w_2N_{\text{rep}} + w_3N_{\text{eval}} + w_4N^s_{\text{rep}} + w_5N^s_{\text{eval}}$, where N_{GA} denotes

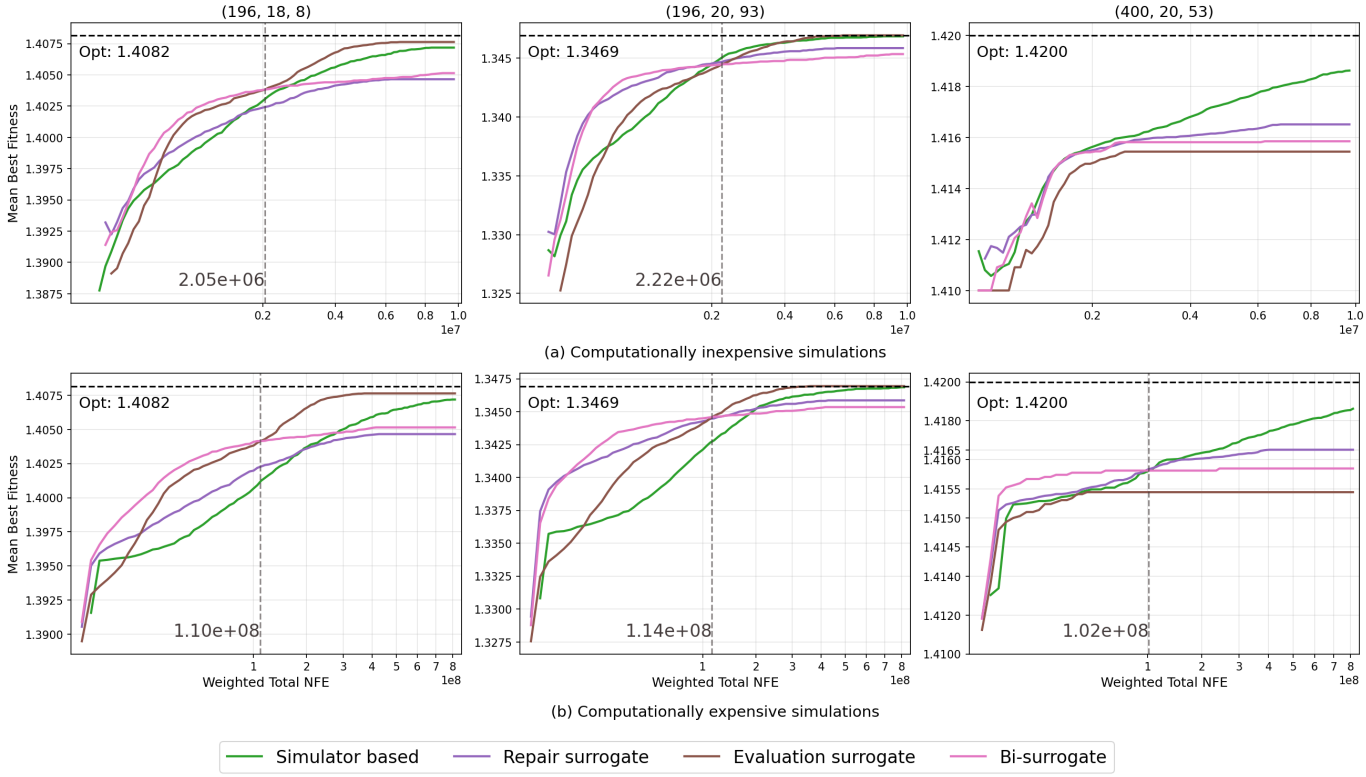

Fig. 5. Mean best fitness versus weighted total NFE under (a) computationally inexpensive and (b) computationally expensive simulation regimes. Each column corresponds to a specific problem instance (ℓ, u, ID). The horizontal dashed line indicates the global optimum, while the vertical dashed line marks the critical point beyond which direct simulation becomes more advantageous than surrogate assistance.

the NFE consumed by the GA evaluation process. The terms N_{rep}, N_{eval}, N^s_{rep}, and N^s_{eval} represent the NFE consumed by simulator repair, simulator evaluation, surrogate repair, and surrogate evaluation, respectively.

The coefficients w_1 through w_5 denote the relative computational weights of each operation. Given the significant cost disparity between surrogate inference and high-fidelity simulation, these weights are scenario-specific. Specifically, simulator repair incurs a higher computational penalty than evaluation ($w_2 > w_3$) because restoring feasibility typically demands an iterative internal local search, whereas evaluation requires only a single forward computation. The exact assignment of these values is detailed in the subsequent section.

D. Results

This section benchmarks the proposed bi-surrogate framework against the baseline and ablation variants under varying resource constraints. For computationally constrained scenarios, performance is quantified by the mean best fitness achieved within a specified NFE limit.

1) Computationally unconstrained simulations:

The initial analysis evaluates algorithmic performance in the absence of weighted NFE constraints. While the evaluation surrogate achieves performance comparable to the simulator-based scheme on moderately difficult instances ($\ell = 196$, 97.0% vs. 95.2% success rate), its performance deteriorates significantly on more complex instances. For $\ell = 400$, the simulator-based scheme retains a robust success rate of 86.2%, whereas the evaluation surrogate drops to 52.5%, and the proposed bi-surrogate achieves 58.5%. This highlights that navigating rugged search spaces requires the exact repair and precise evaluation inherently unattainable by surrogates alone.

However, achieving such high success rates via direct simulation inevitably incurs a prohibitive NFE overhead. In practice, real-world optimization applications rarely mandate locating the exact global optimum; rather, the primary objective is to rapidly secure highly competitive, near-optimal solutions under strict resource limitations. This operational reality motivates the constrained analyses detailed below.

2) Computationally inexpensive simulations:

In the computationally inexpensive regime, as detailed in Table II, the relative cost of simulator operations is moderate compared to surrogate inference. Fig. 5(a) illustrates the convergence trajectories for all schemes across varying problem instances.

For instances with $\ell = 196$, the bi-surrogate scheme exhibits competitive performance in the early stages, specifically when the weighted NFE is below 2×10^6. Beyond this point, however, the simulator-based scheme overtakes the bi-surrogate and eventually outperforms all surrogate-assisted variants. This trend indicates that given sufficient NFE, direct simulator evaluations offer superior precision, avoiding the accumulation

of approximation errors.

For the more challenging $\ell = 400$ instance, all schemes exhibit similar convergence behavior in the initial phase. However, as the cumulative NFE increases, the simulator-based scheme establishes a clear dominance, achieving significantly higher fitness than the surrogate-assisted variants. This performance gap underscores the necessity of precise simulator feedback for tackling the complex problem instances.

3) Computationally expensive simulations:

In the computationally expensive regime, as detailed in Table II, the high relative cost of simulator operations magnifies the importance of sample efficiency. Fig. 5(b) depicts the convergence trajectories under these conditions.

Unlike in the inexpensive regime, the vertical dashed line here indicates that the critical crossover point has shifted significantly to the right, reaching approximately 1.1×10^8 weighted NFE. Consequently, the bi-surrogate scheme maintains a substantial fitness lead across nearly the entire range. By achieving superior fitness with fewer expensive simulator calls, the proposed framework effectively balances exploration and exploitation under strict computational constraints.

While other surrogate-assisted schemes occasionally achieve comparable performance in narrow intervals, the bi-surrogate consistently matches or exceeds all baselines across the broader spectrum. This sustained dominance highlights its robustness and suitability for scenarios where every simulator evaluation incurs a significant cost.

IV. Conclusion

This paper proposed a novel bi-surrogate framework designed to address black-box optimization problems characterized by computationally intensive simulators and internal repair mechanisms arising from unrevealed constraints. In contrast to traditional approaches that treat the simulator as a unified black box, the proposed methodology explicitly models the internal mechanics by strategically employing k-NN to approximate the repair process and BNNs for evaluation function estimation. Experiment results on the spin glass testbench demonstrate the high predictive fidelity of these models, confirming this configuration as a robust foundation for SAEAs in binary optimization domains. Integration of the bi-surrogate architecture with DSMGA-II successfully mitigates the high computational burden associated with frequent simulator calls. Crucially, in the computationally expensive simulation regime, the proposed framework exhibits clear dominance over single-surrogate and baseline schemes. This approach maximizes sample efficiency, achieving superior solution quality while significantly reducing reliance on high-fidelity evaluations. Future work aims to extend the framework's applicability to binary problems with more intricate unrevealed constraints. To ensure robust performance across such diverse landscapes, the integration of multi-armed bandit mechanisms for adaptive hyperparameter tuning is identified as a promising avenue for maximizing optimization efficiency.

Acknowledgment

The authors would like to express thanks for the support provided by the National Science and Technology Council in Taiwan under grant No. 114-2221-E-002-103. Additionally, the authors appreciate the comments and the advice received from the reviewers.

References

[1] C. He, Y. Zhang, D. Gong, and X. Ji, "A review of surrogate-assisted evolutionary algorithms for expensive optimization problems," *Expert Systems with Applications*, vol.217, Art. no. 119495, May 2023.

[2] J. N. Richardson, S. Adriaenssens, P. Bouillard, and R. Filomeno Coelho, "Multiobjective topology optimization of truss structures with kinematic stability repair," *Structural and multidisciplinary optimization*, vol. 46, no. 4, pp. 513-532, Feb. 2012.

[3] F. Ferrari and O. Sigmund, "Towards solving large-scale topology optimization problems with buckling constraints at the cost of linear analyses," *Computer Methods in Applied Mechanics and Engineering*, vol. 363, Art. no. 112911, Feb. 2020.

[4] P. Hu, J. S. Pan, S. C. Chu, and C. Sun, "Multi-surrogate assisted binary particle swarm optimization algorithm and its application for feature selection," *Applied Soft Computing*, vol. 121, Art. no. 108736, May 2022.

[5] N. R. Eluri, G. R. Kancharla, S. Dara, and V. Dondeti, "Cancer data classification by quantum-inspired immune clone optimization-based optimal feature selection using gene expression data: deep learning approach," *Data Technologies and Applications*, vol. 56, no. 2, pp. 247–282, Mar. 2022.

[6] M. M. Elsawy, S. Lanteri, R. Duvigneau, J. A. Fan, and P. Genevet, "Numerical optimization methods for metasurfaces," *Laser & Photonics Reviews*, vol. 14, no. 10, Art. no. 1900445, Aug. 2020.

[7] H. Cai, S. Srinivasan, D. A. Czaplewski, A. B. F. Martinson, D. J. Gosztola, L. Stan, T. Loeffler, S. K. R. S. Sankaranarayanan, and D. López, "Inverse design of metasurfaces with non-local interactions," *Npj Computational Materials*, vol. 6, no. 1, pp. 1–8, Aug. 2020.

[8] P. Dainese, L. Marra, D. Cassara, A. Portes, J. Oh, J. Yang, A. Palmieri, J. Rodrigues, A. Dorrah, and F. Capasso, "Shape optimization for high efficiency metasurfaces: theory and implementation," *Light: Science & Applications*, vol. 13, no. 1, Oct. 2024, Art. no. 300.

[9] J. H. Bussemaker, P. Saves, N. Bartoli, T. Lefebvre, and B. Nagel, "Surrogate-based optimization of system architectures subject to hidden constraints," in *AIAA AVIATION Forum And ASCEND 2024*, Jul. 2024, Art. no. 4401.

[10] J. M. Hernández-Lobato, J. Requeima, E. O. Pyzer-Knapp, and A. Aspuru-Guzik, "Parallel and Distributed Thompson Sampling for Large-scale Accelerated Exploration of Chemical Space," in *Proceedings of the 34th International Conference on Machine Learning*, PMLR, pp. 1470–1479, Jul. 2017.

[11] A. J. Smola, B. Schölkopf. "A Tutorial on Support Vector Regression," in *Statistics and Computing*, vol. 14, no. 3, pp. 199-222, Aug. 2004.

[12] Y. Gal, Z. Ghahramani. "Dropout as a Bayesian Approximation: Representing Model Uncertainty in Deep Learning." in *Proceedings of The 33rd International Conference on Machine Learning*, PMLR, pp. 1050-1059, Jun. 2016.

[13] K. Chen, B. Xue, M. Zhang and F. Zhou, "Correlation-Guided Updating Strategy for Feature Selection in Classification With Surrogate-Assisted Particle Swarm Optimization," *IEEE Transactions on Evolutionary Computation*, vol. 26, no. 5, pp. 1015-1029, Oct. 2022.

[14] S. H. Hsu and T. L. Yu, "Optimization by pairwise linkage detection, incremental linkage set, and restricted/back mixing: DSMGA-II," in *Proceedings of the 2015 Annual Conference on Genetic and Evolutionary Computation*, pp. 519–526, Jul. 2015.

[15] T. L. Yu, D. E. Goldberg, K. Sastry, C. F. Lima, and M. Pelikan, "Dependency structure matrix, genetic algorithms, and effective recombination," *Evolutionary Computation*, vol. 17, no. 4, pp. 595–626, Dec. 2009.

[16] S. F. Edwards and P. W. Anderson, "Theory of spin glasses," *Journal of Physics F: Metal Physics*, vol. 5, pp. 965–974, May 1975.