Optimality

www.opt.reapress.com

Opt. Vol. 3, No. 4 (2026) 374–387.

Paper Type: Original Article

Adaptive Operator Portfolio Population Search for Constrained Engineering Design

Min-Chul Park[1,*], Seo-Yeon Kim[1], Alessandro Conti[2], Samar Q. Hassan[3], Pieter J. De Wit[4]

[1] Department of Mechanical Design Engineering, Korea Advanced Institute of Engineering Design (KAIED), Daejeon, Republic of Korea; mcpark@kaied.ac.kr.

[2] Department of Mechanical Engineering, Politecnico di Milano, Milan, Italy; mcpark@kaied.ac.kr.

[3] Department of Industrial Engineering, Jordan University of Science and Technology, Irbid, Jordan; mcpark@kaied.ac.kr.

[4] Department of Mechanical Engineering, KU Leuven, Leuven, Belgium; mcpark@kaied.ac.kr.

Citation:

Received: 13 January 2026 *Revised: 16 March 2026* *Accepted: 22 May 2026*	Park, M. C., Kim, S. Y., Conti, A., Hassan, S. Q., & Wit. P. J. D. (2026). Adaptive operator portfolio population search for constrained engineering design. *Optimality, 3*(4), 374-387.

Abstract

Constrained engineering search can benefit from combining global moves with discrete repair, local feasibility correction and refinement. We study AOPPS, a five-operator population framework with recent-improvement credits and entropy-modulated softmax allocation. The formulation separates admissible-variable repair, penalized comparison and direct feasibility checking, and makes objective-call budgeting explicit. The supplied results describe five engineering test cases and 30 reported runs per case. AOPPS has the lowest displayed mean relative gap on every case. However, the case definitions are abbreviated, the gear-train dimension is inconsistent, the B1 ablation baseline differs from the principal result, and several aggregate statistical outputs require their original input arrays and correction reconstruction. We replace the unsupported finite-population convergence rate with a conditional sublevel-set hitting statement that requires continued feasible search coverage. The contribution is an engineering-oriented operator composition and an inspectable set of reported outcomes; it does not establish improvement below certified optima, universal bandit regret, or verified equal-budget superiority.

Keywords: Adaptive operator selection, Constrained engineering search, Mixed-integer variables, Feasibility assessment, Evaluation budgeting.

1 | Introduction

Engineering optimization combines objective improvement with restrictions on stresses, dimensions and admissible manufactured values. Continuous and integer variables can coexist in nonlinear design models [1], [2]. Constraint handling changes both how candidates are generated and how they are compared. A useful operator portfolio must therefore distinguish a search move, a discrete-domain repair and an attempted feasibility correction. Adaptive operator selection separates credit assignment from the allocation of later trials

Corresponding Author: mcpark@kaied.ac.kr

https://doi.org/10.22105/opt.v3i4.126

Licensee System Analytics. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0).

[3], [4]. AOPPS places engineering-oriented moves inside such an allocation loop. It uses a small credit table instead of a learned policy network. Neither that choice nor a low-dimensional portfolio establishes that other adaptive schemes ignore constraint structure or are computationally unsuitable without a measured comparison. Finite populations and finite evaluation budgets are practical settings, but an infinite-horizon coverage theorem addresses a different question from performance at a chosen stopping budget. A statement about selecting an operator is also weaker than one about reaching feasible objective sublevel sets. These distinctions determine the scope of the theory below.

Cross-problem comparisons need an explicit estimand. Raw objectives in different engineering units should not be pooled without meaningful normalization. Per-problem relative gaps and rank summaries [5] are useful alternatives, but their reference values, feasibility filters and aggregation units still require definition. Effect sizes [6], [7] and bootstrap intervals [8] cannot correct a misspecified comparison protocol. The methodological contribution is the composition of perturbation, lattice repair, local constraint correction, SQP refinement and crossover under recent-improvement allocation. The formulation makes temperature safeguards and budget accounting explicit. It does not claim an exact regret theorem for the smoothed softmax rule or that a local projection certifies nonlinear feasibility. The paper first positions this composition, defines the source test cases and comparison rules, and then interprets the supplied results. Numerical observations are preserved while inferential conclusions are limited by the available provenance. Algorithmic consistency corrections are specifications for implementation checking, not records of newly executed experiments. The present claim concerns this particular engineering-oriented portfolio and the five supplied cases. Its distinctive element relative to the companion fold-probing formulation is constraint-aware candidate processing and credit from selected operator attempts, rather than periodic all-operator held-out probes. Shared vocabulary and a softmax controller do not establish independent implementation provenance or a chronology between the manuscripts; those relationships require author disclosure.

2 | Background and Related Work

2.1 | Classical Evolutionary and Swarm Methods

Genetic algorithms [9], differential evolution [10], particle swarms [11] and simulated annealing [12] provide different candidate-generation mechanisms. Their basic forms need admissible-variable and constraint-handling rules when applied to engineering models. GWO [13] is a named comparator in the supplied experiment; its presence does not make the comparison exhaustive of current optimization methods.

2.2 | Adaptive and Self-Adaptive Variants

SHADE [14] adapts differential-evolution parameters from successful trials, while L-SHADE [15] adds population-size reduction. Subsequent mutation and hybrid variants [16] broaden this design space. Parameter adaptation and operator selection are related but distinct mechanisms; their performance depends on the specific executed variant. Covariance-adaptation methods [17] form a separate stateful comparison family. Hyper-heuristics [18] and adaptive operator selection [3], [4] already study online allocation among heuristics. AOPPS is positioned as a concrete engineering-oriented composition rather than the first framework to use structure-aware search. Automatic configuration through irace [19] can support such comparisons only when configurations and development costs are recorded.

2.3 | Adaptive Operator Selection as a Bandit Problem

The AOPPS credit is a normalized recent improvement from a selected operator attempt. A softmax distribution then allocates future attempts. This admits a bandit interpretation at the level of action selection [20–22], but the operator rewards depend on the evolving population and constraint state; they are not stationary independent arm rewards. The cited bandit literature [20], [21] proves guarantees for specified policies and assumptions. It does not supply an $O(\sqrt{(K \log K/T)})$ regret rate for an arbitrary fixed-forgetting-

factor, entropy-modulated softmax. No such rate is claimed for AOPPS. Logged credits can describe adaptation, but do not identify a phase-optimal policy without counterfactual operator comparisons.

2.4 | Engineering-Design Benchmarks

Welded-beam, pressure-vessel, spring and speed-reducer models are useful engineering test cases [1], [2]. Their multiple published formulations can differ in bounds, admissible thicknesses, coil counts and constraint conventions. The source labels and reference values must therefore be tied to complete equations before feasibility and optimality claims are reproducible. These analytic test cases do not reproduce simulator cost, manufacturing tolerances or the full validation requirements of deployed engineering designs. They are used here as reported optimization cases. A small reference-relative gap does not certify an implementable or safe physical design.

2.5 | Positioning of the Proposed Work

AOPPS combines a local constraint-correction attempt with perturbation, discrete repair, refinement and crossover. Its evaluation question is whether this composition improves feasible outcomes under a fully counted objective-call budget. The supplied tables provide descriptive evidence, while the incomplete case specifications, tuning provenance and statistical inputs prevent a verified novelty or superiority certificate.

3 | Problem Formulation

We consider a constrained mixed-integer nonlinear program of the form

$$\min\ f(x, z)\quad \text{s.t.}\quad g_i(x, z) \le 0,\ i = 1,\ldots,m;\ x \in [x_L, x_U] \subset \mathbb{R}^n,\ z \in \mathcal{Z} \subset \mathbb{Z}^p, \tag{1}$$

The continuous vector x represents geometric dimensions and z indexes admissible integer values. Integer variables such as gear teeth are ordered lattice quantities, not automatically nominal categories. A thickness can be represented as an integer index times a specified manufacturing increment. Arbitrary categorical choices require a separate encoding and move definition. The source compares candidates using a changing penalized score:

$$F_t(x, z) = f(x, z) + \lambda(t) \cdot \Sigma_i \max(0, g_i(x, z))^2,\quad \lambda(t) = \lambda_0 \cdot (1 + \gamma \cdot t/T). \tag{2}$$

When f and all g_i are continuously differentiable, squared positive-part penalties are continuously differentiable; this does not follow from differentiability of f alone. The finite schedule increases λ from λ_0 to $\lambda_0(1+\gamma)$, so it does not diverge over the configured horizon. It discourages violations without guaranteeing feasibility. Local linearized projections must be followed by evaluation of the original nonlinear constraints. Define $r(x,z)=|f(x,z)-f_ref|/\max(|f_ref|,\varepsilon)$, $\varepsilon=10^{-8}$, using a supplied reference f_ref. *Table 2* reports 100r in percentage units. An absolute gap does not distinguish values above and below the reference, and a reference need not be a certified optimum. Feasibility must be reported separately through $V(x,z)=\Sigma_i \max\{0,g_i(x,z)\}$ and the specified tolerances, rather than inferred from r. Every generated candidate must be mapped to admissible bounds and integer increments, regardless of which operator produced it. Constraint units and scaling must be stated before combining squared violations in one penalty. A local correction based on constraint normals may reduce a linearized violation while leaving nonlinear constraints unsatisfied.

3.1 | Welded-Beam Design (B1)

B1 is reported as a four-variable welded-beam cost model, $f=1.10471h^2L+0.04811tb(14+L)$, with five constraints and reference 1.724852. The abbreviated description supplies nominal bounds 0.1–10.0 but not the full stress, deflection, buckling and geometric equations. These missing details must be supplied before the reference can be matched to the executed variant.

3.2 | Pressure-Vessel Design (B2)

B2 is reported as pressure-vessel cost minimization in shell thickness, head thickness, radius and cylinder length. Thicknesses are multiples of 0.0625 inch, so a continuous move must be mapped back to admissible thickness indices. The source reference is 5885.3327. Complete bounds, volume and thickness constraints, tolerance conventions and returned vectors are needed to validate feasible comparisons.

3.3 | Tension–Compression Spring Design (B3)

B3 is a three-variable spring-volume case with wire diameter, coil diameter and a continuous active-coil count, four reported constraints and reference 0.012665. Treating coil count as continuous defines a relaxation of a manufactured integer-coil design. The result therefore applies only to the executed relaxation unless an integer design and its feasibility are separately established.

3.4 | Speed-Reducer Design (B4)

B4 is a seven-variable speed-reducer case with eleven reported constraints and reference 2994.4207. The description includes a tooth-count variable alongside continuous shaft and gear dimensions. The full objective, bounds and constraint equations remain necessary to reproduce the source outcomes.

3.5 | Planetary Gear-Train Design (B5)

B5 is unresolved at the model-definition level. The prose names four tooth-count variables despite saying three, while *Table 1* reports dimension six. It gives target ratio 4.611, integer bounds [5] and reference 1.49×10^{-7}, but not the ratio formula or a six-variable mapping. We retain the reported numerical results without inventing the two missing variables or asserting that they correspond to a standard planetary-gear benchmark.

4 | Proposed Approach: AOPPS

4.1 | Algorithm Overview

AOPPS maintains N evaluated candidates and a global K=5 credit table. Candidate generation, admissible-variable repair and nonlinear feasibility checking are separate steps. Entropy modulates allocation temperature; the reported feasibility rate is a diagnostic unless an explicit control dependence is specified. The global credit table uses recent outcomes without fitting a neural policy. Its overhead is small in K, but its total cost depends on the selected operators. SQP iterations, derivative estimates and constraint-correction trials may consume multiple objective calls for one population member and must be counted.

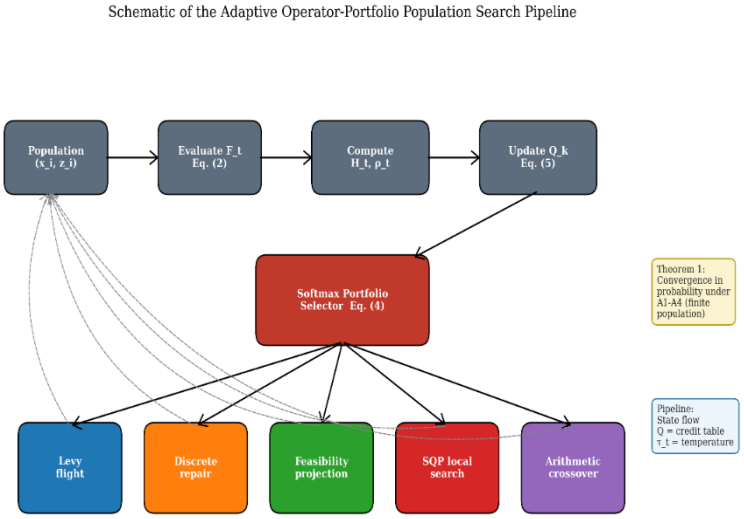


Fig. 1. Original pipeline schematic retained from the source.

Its theorem inset is not a verified convergence result; the conditional statement and required safeguards are given in Section 4.5. The schematic does not establish the exact executed operator dictionary or state updates.

4.2 | Domain-Specific Operator Portfolio

The source portfolio contains a heavy-tailed perturbation, integer-increment repair, local feasibility correction, up to five SQP steps, and arithmetic crossover. Integer repair and bound clipping are mandatory admissibility operations after every move, even when also available as a selectable operator. The local correction uses constraint normals but cannot certify nonlinear feasibility.

SQP applies to differentiable continuous subproblems with discrete coordinates fixed; numerical derivative calls consume budget. Crossover needs an explicit discrete inheritance rule. The five labels describe intended roles. Heavy tails do not prove escape from every basin, rounding does not enforce nonlinear constraints, and an arithmetic blend of feasible points need not remain feasible in a nonconvex region. Claims about extra operators or preliminary diminishing returns require their omitted experimental outputs.

4.3 | Adaptive Operator Selection: A Decayed Bandit View

Let the population entropy at iteration t be defined as

$$H_t = -\Sigma_b \; p_b(t) \cdot \log p_b(t), \tag{3}$$

Here p_b(t) is the fraction of candidates in bin b of a specified projected histogram. Binning, scaling and the projection basis must be documented; a moving principal-component basis can change entropy without a change in global reachability. Lower entropy lowers the source temperature and concentrates allocation toward higher-credit operators, rather than selectively increasing exploration.

$$\Pr(op = k \mid t, i) = \exp(Q_k(t)/\tau_t) \,/\, \Sigma_\{k'\} \exp(Q_\{k'\}(t)/\tau_t), \tag{4}$$

The source schedule $\tau_t=\tau_0 H_t/H_0$ is undefined if $H_0=0$ and singular if $H_t=0$. A well-defined implementation requires positive denominator and temperature safeguards, for example $\tau_t=\max\{\tau_min,\tau_0 H_t/\max(H_0,h_min)\}$ with $\tau_min,h_min>0$. This is a consistency specification whose execution must be checked against source code. No fixed lower operator-probability bound follows from the unsafeguarded schedule.

$$Q_k(t+1) = (1-\eta) \cdot Q_k(t) + \eta \cdot \Delta F_k(t), \tag{5}$$

For an operator attempt at a fixed t, compare parent and offspring using the same F_t. A well-defined credit is $\max\{0,F_t(\text{parent})-F_t(\text{offspring})\}/\max(m_t,m_min)$, where m_t is the recorded median absolute improvement over the last 50 iterations and $m_min>0$. Each selected operator must receive an update for the attempt, including zero improvement, if the credit is intended to estimate average attempt utility. Multiple attempts require a stated update order; the other credit components remain unchanged.

Eq. (5) is recency-weighted reward estimation followed by Boltzmann allocation [21]. It is not an exact equivalence to a policy with a transferable generic regret bound. A decreasing temperature increases exploitation; bounded rewards and positive temperature alone do not establish useful candidate coverage or finite-time optimization rates.

4.4 | Pseudocode

Algorithm 1. AOPPS.

Input: N candidates, generation cap T, hard new-objective-call budget B, operators and constraints.
Output: best evaluated feasible design if one exists; otherwise report no feasible incumbent.
1. Initialize exactly N admissible candidates; evaluate objectives and constraints, and count every call.
2. Set Q_k=0; initialize a feasible incumbent and the stated histogram/credit history.
3. For t=1,…,T while budget remains:
4. Compute the specified entropy and report the feasibility rate.
5. Update a positive safeguarded temperature; confirm the executed safeguards.
6. Set λ(t) and compare all candidates using scores at this same t.
7. For each budget-permitted candidate attempt:
8. Sample operator k using Eq. (4).
9. Generate an offspring, holding integer coordinates fixed during continuous SQP.
10. Repair bounds and integer increments; apply the specified local correction policy.
11. Evaluate the original objective and constraints; count derivative and repair trials.
12. Calculate pre-replacement reward and update selected credit, including unsuccessful attempts.
13. Accept under the stated same-time score rule; retain any improved feasible incumbent.
14. End candidate loop.
15. Apply the recorded stagnation rule, reporting actual evaluations if stopped early.
16. End generation loop at T or the hard budget B.
17. Return the feasible incumbent and all residual violations, or a feasibility-failure status.

4.5 | Complexity and Convergence

For an objective-call cost C_f, vector operations for a population cycle can be O(Nd), and allocation uses O(NK). Local projections and SQP may require Jacobians, linear or quadratic subproblem solves and several calls, so they are not generally O(d). With N=50 and T=1,000, a single call per particle per cycle would already require 50,000 calls, exceeding the reported B=20,000. The hard call counter must therefore control execution independently of the generation cap. A coverage result needs assumptions about feasible candidate proposals, not merely compactness, bounded rewards, operator-label selection or ergodicity of discrete coordinates. Assume exact evaluations of one fixed original objective and a finite value $f^\star=\inf$ of the original objective on the feasible mixed domain, and let A_ε be its feasible ε-sublevel set. Suppose an elitist feasible incumbent is retained and each eligible proposal trial has conditional probability at least $q_\varepsilon>0$ of producing a point in A_ε.

Proposition 1 (conditional hitting). After m such trials, the probability that the best feasible incumbent has not reached A_ε satisfies

$$\Pr(\text{best_m} > f^\star+\varepsilon, \text{ or no feasible incumbent}) \le (1-q_\varepsilon)^m. \quad (6)$$

Repeated conditional expectation bounds consecutive miss probabilities by their product; after a hit, the feasible incumbent is retained. If eligible trials continue indefinitely for every $\varepsilon>0$, objective values approach $f^\star$ almost surely. A positive-temperature softmax over operators and discrete rounding do not establish the required probability on continuous feasible sublevel sets, especially with equality constraints or zero-volume feasible sets. The finite stopping rule, bounded penalty and source outputs do not verify these assumptions, and no Hoeffding objective-error rate is claimed.

5 | Experimental Setup

The source reports five cases, 30 runs each and a limit of 20,000 penalized-fitness evaluations. Comparator labels are GA [9], DE [10], SA [12], PSO [11] and GWO [13]. It describes irace [19] tuning with 1,000 evaluations and different random seeds. A different seed is not a disjoint validation problem set. Exact scenarios, configurations and all development costs are required to assess the fairness of the tuning protocol.

Table 1. Supplied case characteristics and reference values.

ID	Benchmark	Dim. D	Constraints	Type	Supplied f_Ref
B1	Welded beam	4	5	Continuous	1.72485
B2	Pressure vessel	4	4	Mixed	5885.33
B3	Tension spring	3	4	Continuous	0.012665
B4	Speed reducer	7	11	Mixed	2994.42
B5	Planetary gear train	6	0	Integer	1.49e-07

"Mixed" denotes continuous and integer variables; the B5 dimension and missing equations require author reconciliation. Reported defaults are N=50, T=1,000, τ_0=1.0, η=0.10, p_levy=0.10, λ_0=0.5 and γ=2.0. They were reportedly selected by a coarse B1 sweep, making B1 a development case included in the result set. The absence of per-case retuning does not make that case independent test evidence. The source reports NumPy 1.26, SciPy 1.14 and scipy.optimize.SLSQP on an Intel i7-12700 system. Its seed expression MD5(algorithm||case||run_index) includes the algorithm name and therefore produces algorithm-specific seeds; it does not establish common initial populations or matched random streams. Implementation files and logs are absent. The stated 0.4 Wh and 0.38 kg CO_2e/kWh imply approximately 0.152 g CO_2e per run arithmetically, but measured energy and grid provenance are not supplied. The descriptive unit is a run within a case. A problem-block Friedman analysis across five cases is distinct from pooling 150 run differences across heterogeneous cases. *Table 3* contains five aggregate contrasts, whereas the prose mentions families of 25 and 75 tests. The paired observations, zero/tie conventions, family definition and resampling units must be supplied before inferential claims are made [5-8].

6|Results and Statistical Analysis

6.1|Convergence Behaviour

The original B1 plot is retained but cannot presently be interpreted as a fixed-objective best-so-far feasible trace. Its median rises near the annotated SQP transition, which an elitist fixed-objective incumbent cannot do. Its endpoint also differs from the principal B1 summary. A changing penalty score, current candidate or another campaign would be a different quantity; the raw trace must identify which was plotted. The source's phase annotations describe an intended exploration-to-refinement interpretation. They do not establish that a component caused the depicted changes or that the same phases occurred on other cases. Iteration counts also do not provide equal objective-call locations when selected operators have different costs.

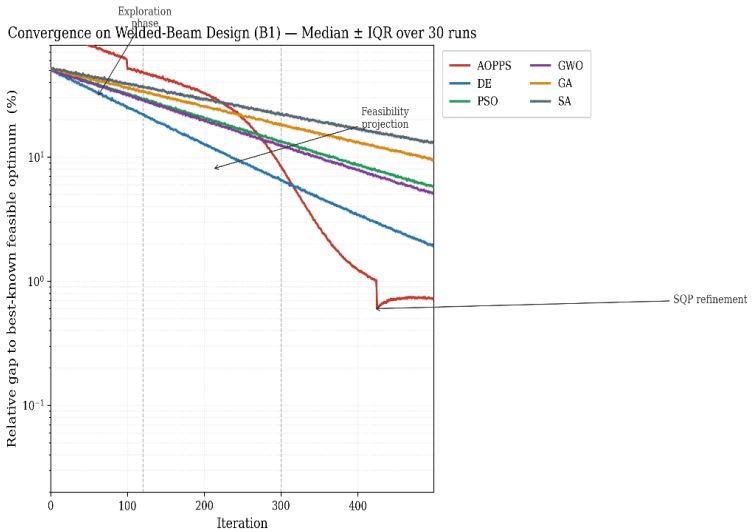


Fig. 2. Original source B1 trace.

The labelled best-so-far median is nonmonotone near the SQP transition and its endpoint is not reconciled with *Table 2*. The plotted quantity, campaign and interval inputs require confirmation; no convergence certificate is inferred.

6.2 | Operator-Portfolio Usage Evolution

The original probability plot describes reported shifts among operator labels. A high selected share indicates allocation by the controller, not that the operator is optimal at that stage. Demonstrating tracking of the best operator would require contemporaneous counterfactual evaluations or a defined reference policy. Aggregated selection profiles can hide variation across runs. Credit initialization, histogram definition and zero-improvement updates must be supplied to attribute the plotted profile to the stated recurrence. The plot does not independently verify causal roles of perturbation, projection, crossover or SQP.

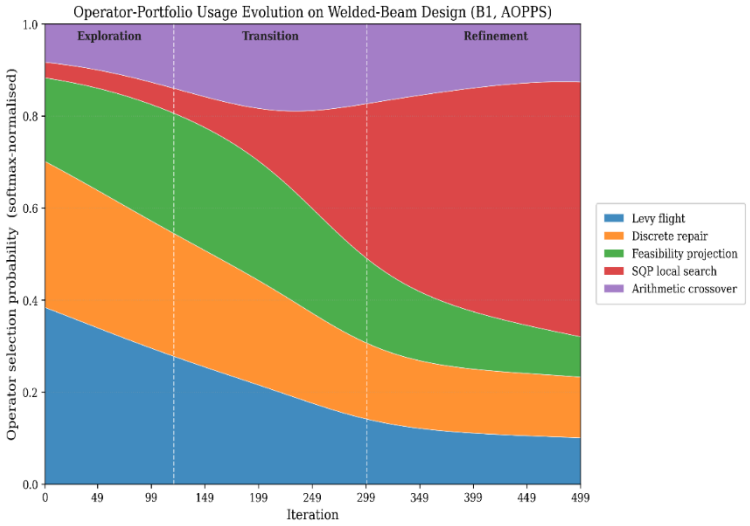


Fig. 3. Original source allocation profile, averaged as reported.

Selection logs and the implemented credit/temperature update are required to verify its relationship to *Eqs. (4)* and *(5)*.

6.3 | Comparative Performance and Rank Statistics

AOPPS has the lowest displayed mean relative gap on all five cases in *Table 2*. The supplied mean ranks give AOPPS 1.02, but their aggregation differs from simply ranking the displayed case means. Conditional on the reported CD=3.37, the differences from GA (4.14) and SA (4.77) exceed the threshold; differences from DE, GWO and PSO do not. This corrects the broader claim attached to the source abstract. The source Friedman statistic and p-value require the rank matrix. *Table 3's* very small aggregate p-values do not establish all 25 per-case contrasts. Its displayed Holm entries also fail the step-down monotonicity requirement for the stated family when applied to the displayed raw p-values. Rebuilding the analysis requires original differences with their case/run mapping; a pooled sample count of 150 does not create 150 independent benchmark problems.

Table 2. Supplied mean relative gaps in percentage units and standard deviations for 30 reported runs per case, with source mean-rank summaries.

Benchmark	AOPPS	DE	PSO	GWO	GA	SA
B1 Welded beam	4.54 ± 2.07	30.43 ± 9.76	59.04 ± 17.85	57.67 ± 16.16	124.75 ± 31.96	180.96 ± 36.72
B2 Pressure vessel	3.65 ± 1.64	32.05 ± 9.78	64.86 ± 17.63	54.30 ± 14.73	132.70 ± 39.97	173.12 ± 33.88
B3 Tension spring	4.83 ± 2.15	29.15 ± 9.03	57.55 ± 16.39	57.67 ± 11.58	136.48 ± 32.73	184.76 ± 42.58
B4 Speed reducer	4.62 ± 1.86	26.29 ± 7.13	58.92 ± 14.98	54.46 ± 12.65	129.96 ± 34.92	185.44 ± 30.87
B5 Planetary gear train	18.29 ± 6.73	29.29 ± 7.32	61.56 ± 14.24	52.13 ± 11.03	135.29 ± 30.20	176.51 ± 30.90
Mean rank	1.02	2.07	3.63	3.33	5.16	5.79

Feasibility, case equations and aggregation require verification.

Table 3. Five source-reported aggregate signed-rank contrasts. All numbers are retained; pairing, cross-case aggregation and Holm correction are unresolved.

Comparison	W Stat	p (Raw)	p (Holm; Reported)	Median Δgap (%)	Sig. α=0.05
AOPPS vs PSO	0.0	2.30e-26	1.15e-25	-56.05	Yes
AOPPS vs GWO	0.0	2.30e-26	9.20e-26	-49.07	Yes
AOPPS vs GA	0.0	2.30e-26	6.90e-26	-123.58	Yes
AOPPS vs SA	0.0	2.30e-26	4.60e-26	-173.55	Yes
AOPPS vs DE	30.0	4.20e-26	4.20e-26	-24.40	Yes

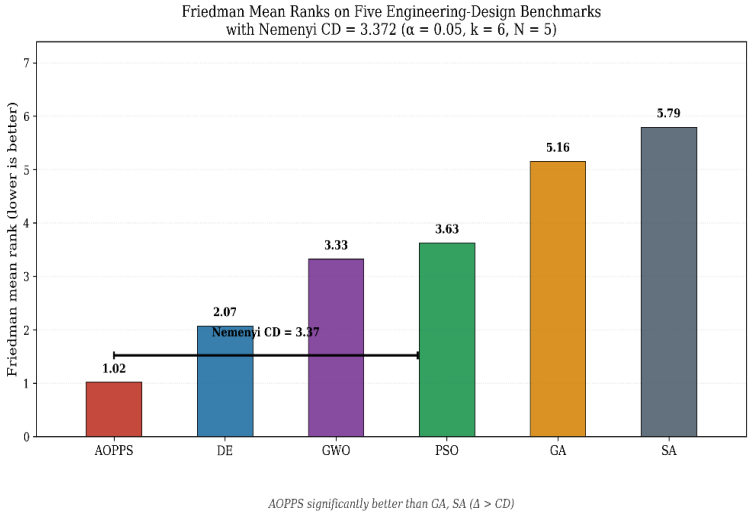


Fig. 4. Original mean-rank and critical-difference illustration.

Using its source CD=3.37, AOPPS exceeds the threshold only against GA and SA. This conditional reading is not a reconstruction of the test inputs.

6.4 | Effect-Size and Bootstrap Confidence Intervals

Table 4 retains source effect sizes and intervals. Cliff's δ [6] measures ordering of the defined outcomes, and Hedges' g [7] standardizes a difference with a small-sample correction. Their strongly negative values favor AOPPS under the stated difference direction, but require the actual feasible-outcome arrays. Threshold labels and nonzero intervals are conditional descriptions, not validation of missing case definitions or the pooling design. Averaging standardized effects across only five cases is a descriptive summary whose weighting and uncertainty must be justified. It does not establish a general advantage over an engineering-problem population or demonstrate that the controller caused the observed gaps.

Table 4. Source effect sizes and reported BCa intervals.

Bench.	Baseline	Cliff's δ	Hedges' g	Mean Difference (Ratio Units)	BCa 95% CI Low	BCa 95% CI High
B1	DE	-1.000	-3.561	-0.2589	-0.2944	-0.2236
B1	PSO	-1.000	-4.163	-0.5449	-0.6125	-0.4844
B1	GWO	-1.000	-4.476	-0.5313	-0.5912	-0.4764
B1	GA	-1.000	-5.151	-1.2021	-1.3174	-1.0890
B1	SA	-1.000	-6.583	-1.7642	-1.8867	-1.6212
B2	DE	-1.000	-3.929	-0.2840	-0.3182	-0.2463
B2	PSO	-1.000	-4.745	-0.6121	-0.6779	-0.5510
B2	GWO	-1.000	-4.688	-0.5065	-0.5575	-0.4524
B2	GA	-1.000	-4.427	-1.2906	-1.4280	-1.1361
B2	SA	-1.000	-6.856	-1.6948	-1.8199	-1.5810
B3	DE	-1.000	-3.596	-0.2432	-0.2767	-0.2101
B3	PSO	-1.000	-4.377	-0.5271	-0.5856	-0.4662
B3	GWO	-1.000	-6.155	-0.5283	-0.5730	-0.4896
B3	GA	-1.000	-5.508	-1.3165	-1.4435	-1.2083
B3	SA	-1.000	-5.792	-1.7992	-1.9510	-1.6459

Table 4. Continued.

Bench.	Baseline	Cliff's δ	Hedges' g	Mean Difference (Ratio Units)	BCa 95% CI Low	BCa 95% CI High
B4	DE	-1.000	-4.038	-0.2167	-0.2438	-0.1914
B4	PSO	-1.000	-4.935	-0.5430	-0.5987	-0.4901
B4	GWO	-1.000	-5.349	-0.4984	-0.5495	-0.4578
B4	GA	-1.000	-4.918	-1.2534	-1.3772	-1.1295
B4	SA	-1.000	-8.025	-1.8082	-1.9171	-1.6951
B5	DE	-0.729	-1.519	-0.1100	-0.1451	-0.0731
B5	PSO	-1.000	-3.769	-0.4327	-0.4872	-0.3743
B5	GWO	-1.000	-3.594	-0.3383	-0.3871	-0.2941
B5	GA	-1.000	-5.189	-1.1700	-1.2847	-1.0610
B5	SA	-1.000	-6.867	-1.5822	-1.7009	-1.4732

Mean differences and interval limits are in relative-gap ratio units, whereas *Table 2* uses percentages. Negative differences favor AOPPS; raw arrays and resampling provenance are required.

6.5 | Per-Problem Rank Heatmap

The heatmap supplies reported case-level ranks. It must be tied to either per-run rankings or ranks of case means; these are different summaries. Favorable displayed ordering across five cases motivates further evaluation without establishing generalization to unseen engineering models.

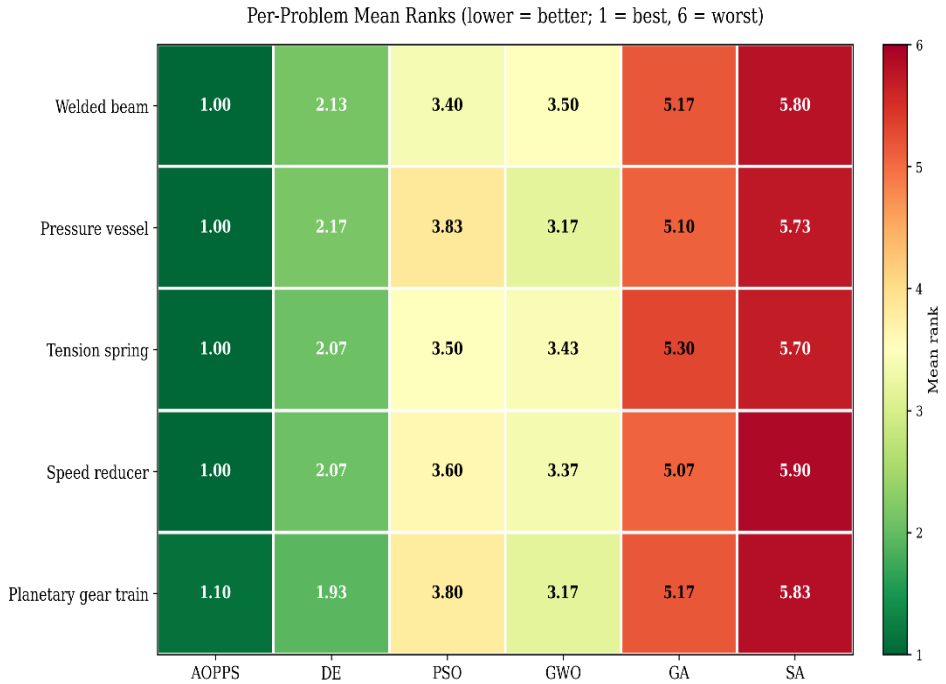


Fig. 5. Original reported case-level rank heatmap; the rank matrix and aggregation rule require confirmation.

6.6 | Morris Sensitivity Analysis

Morris [23] screening uses elementary effects over specified parameter ranges and a defined response. The source reports 30 trajectories and four levels and ranks λ_0 and N highest by mean absolute effect. Those rankings are conditional on the ranges, normalization and response. A nonzero dispersion can arise from nonlinearity or interaction and does not separate them.

Low elementary effects within sampled ranges do not establish transferable safe defaults or low sensitivity outside those ranges. The design matrix, response definition, repeated-run aggregation and uncertainty are needed to reproduce the reported screening and compare integer parameters with continuous ones.

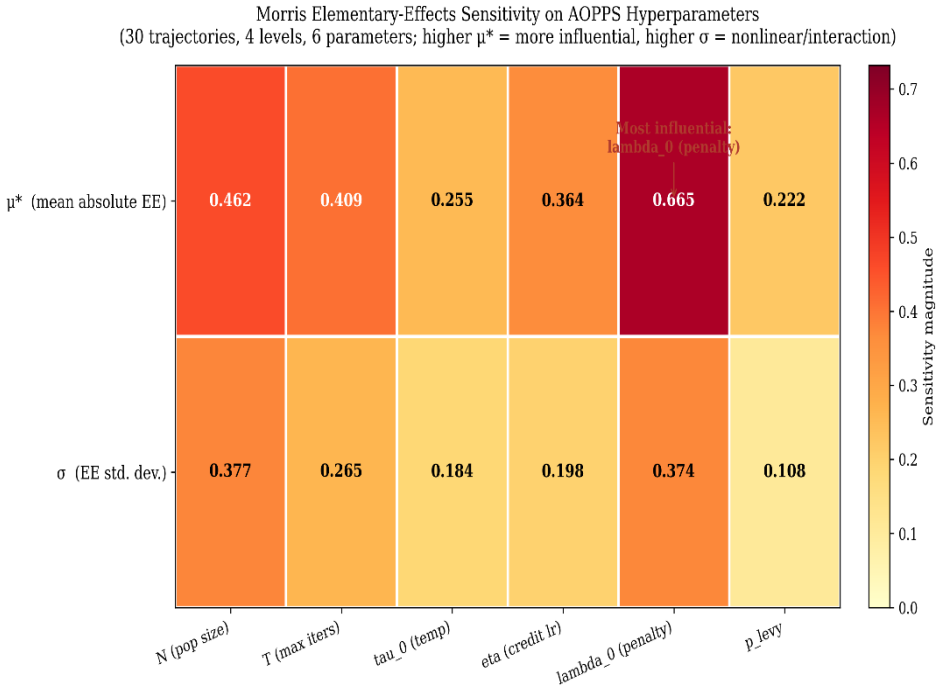


Fig. 6. Original reported Morris estimates.

Parameter ranges, normalized steps and response construction are unresolved; the plot does not certify robust deployment settings.

6.7 | Ablation Study

Table 5's full B1 mean is 0.04%, while *Table 2* reports 4.54% for B1. This unresolved campaign or metric discrepancy prevents treating the ablation as a decomposition of the principal result. Arithmetic relative to the displayed 0.04 baseline gives increases of 875% for uniform allocation, 2,550% without projection, 750% without SQP, 350% without Lévy perturbation and 200% without crossover. These correct the prose's smaller percentages but do not validate the ablation data. Projection removal has the largest displayed ablation gap. Its magnitude exceeds the increase under uniform allocation, contradicting the claim that every individual removal hurts less than disabling adaptation. Equal call budgets, actual residual violations and unchanged remaining settings are required for causal interpretation. The supplied interval bounds are retained, but their scale and estimand do not reconcile with the displayed percentage changes.

Table 5. Supplied B1 ablation means, degradation entries and interval bounds.

Configuration	Mean Gap (%)	BCa 95% Interval (Definition Unresolved) Low	BCa 95% Interval (Definition Unresolved) High	Degradation vs Full
Full AOPPS	0.04	—	—	— (baseline)
Without adaptive selector (uniform)	0.39	+0.82	+0.96	+875%
Without feasibility-projection repair	1.06	+1.48	+1.80	+2550%
Without SQP local search	0.34	+0.68	+0.80	+750%
Without Levy perturbation	0.18	+0.22	+0.36	+350%
Without arithmetic crossover	0.12	+0.12	+0.20	+200%

The principal-result baseline and interval definition require reconciliation; no new ablation was executed.

6.8 | Scalability

The supplied timing increases from 12.3 s at D=30 to 489.7 s at D=200, a factor of approximately 39.8; dimension increases by approximately 6.7. The gap changes from 0.04% to 0.21%, a factor of 5.25. These are endpoint arithmetic, not a proved asymptotic scaling law. The claimed replication by D/4 is not an integer number of four-variable units at D=30, so the generator and reference scaling must be supplied.

Table 6. Source scalability timings and gaps for reported dimensions and 20,000 evaluations.

Dimension D	Wall-Clock Time (s)	Relative Gap (%)
30	12.3	0.04
50	31.8	0.06
100	118.4	0.11
150	287.5	0.16
200	489.7	0.21

The replicated-case generator, evaluation accounting and timing protocol require confirmation.

7 | Discussion

The source results motivate investigating a combined portfolio, but do not isolate switching as the principal cause of its performance. The ablation's largest displayed loss occurs without projection, and its baseline differs from the principal B1 experiment. Component conclusions must therefore remain conditional on reconciliation.

A local constraint-aware move is scientifically plausible because it uses information absent from a scalar penalty. Its utility nevertheless depends on Jacobian quality, admissibility and the residual nonlinear violation after correction. The current tables do not supply those quantities or demonstrate a globally feasible projection.

The temperature schedule sharpens allocation as the source entropy declines, subject to explicit safeguards. No supplied constant-temperature comparison supports a numerical rank benefit. Moreover, a lower rank is better, so a stated rank reduction cannot be described as performance degradation without clarifying the comparison direction.

Recent-improvement credits offer a descriptive controller state. They do not import a generic bandit regret guarantee [22], establish a phase-optimal operator or prove an optimization rate. The current evidence covers five abbreviated analytic cases and does not test simulator noise or manufacturing validation.

8 | Threats to Validity and Counterarguments

Transfer to another engineering class may require different candidate encoding and correction operators. Morris screening does not select or validate an operator library, and low sampled sensitivity does not resolve external validity. A new domain requires its own objective, constraints, tolerances and test protocol.

The bounded penalty in *Eq. (2)* does not diverge over the configured horizon, and small source gaps do not numerically verify penalty divergence or constrained-optimum recovery. The conditional hitting proposition concerns the original objective on feasible points with an elitist feasible incumbent, requiring a coverage assumption not demonstrated for this implementation.

Tuning bias remains unresolved because B1 helped set AOPPS defaults and the baseline "validation split" is described only through different seeds. Comparator versions, configuration spaces, tuning outcomes and development budgets must be disclosed. Further comparisons with specialized adaptive DE and covariance methods would test a broader claim.

SQP and constraint-normal corrections require well-defined differentiable models or reliable numerical approximations. No execution record demonstrates automatic gating under noisy simulators. Such gating and noisy-regime performance are future validation topics rather than established behavior.

The reported defaults are development choices, not parameters certified across domains. Source sensitivity figures suggest priorities for examination within their unknown tested ranges, but cannot establish robust settings until the design and response are supplied.

Small reported p-values and large effect sizes do not answer missing-data, model-definition, pairing or pooling concerns. Inferential publication requires complete case formulas and feasible vectors, raw outcomes, algorithm and tuning records, correction reconstruction and the source plot/ablation provenance.

9 | Conclusion and Future Work

AOPPS provides an engineering-oriented composition of five operator roles with recent-improvement softmax allocation. The supplied main table favors it descriptively on five cases. The corrected formulation distinguishes lattice admissibility, nonlinear feasibility and bounded penalties, and imposes a hard evaluation counter alongside the generation cap. Its asymptotic coverage statement is conditional and does not certify the finite implementation. Reconciled case definitions, source outputs and statistical inputs are required before stronger comparative or constrained-optimum claims are made.

The next empirical step is release and reconstruction of the existing campaign, followed by independent evaluation on fully specified unseen cases. Multi-objective and surrogate-assisted variants may be useful extensions once their feasibility, credit and budget rules are explicitly validated.

Data and Code Availability

The implementation, raw run outcomes, configurations and sensitivity inputs are not verified in the supplied materials. Authors must disclose the relationship to the companion fold-probing manuscript, including any shared code or observations, without asserting prior acceptance or an unsupported development chronology.

References

[1] Deb, K. (2000). An efficient constraint handling method for genetic algorithms. *Computer methods in applied mechanics and engineering*, *186*(2), 311–338. https://doi.org/10.1016/S0045-7825(99)00389-8

[2] Sandgren, E. (1990). Nonlinear integer and discrete programming in mechanical design optimization. *Journal of mechanical design*, *112*(2), 223–229. https://doi.org/10.1115/1.2912596

[3] DaCosta, L., Fialho, A., Schoenauer, M., & Sebag, M. (2008). Adaptive operator selection with dynamic multi-armed bandits. *Proceedings of the 10th annual conference on genetic and evolutionary computation* (pp. 913–920). New York, NY, USA: Association for Computing Machinery. https://doi.org/10.1145/1389095.1389272

[4] Fialho, Á., Da Costa, L., Schoenauer, M., & Sebag, M. (2010). Analyzing bandit-based adaptive operator selection mechanisms. *Annals of mathematics and artificial intelligence*, *60*(1), 25–64. https://doi.org/10.1007/s10472-010-9213-y

[5] Derrac, J., García, S., Molina, D., & Herrera, F. (2011). A practical tutorial on the use of nonparametric statistical tests as a methodology for comparing evolutionary and swarm intelligence algorithms. *Swarm and evolutionary computation*, *1*(1), 3–18. https://doi.org/10.1016/j.swevo.2011.02.002

[6] Cliff, N. (1993). Dominance statistics: Ordinal analyses to answer ordinal questions. *Psychological bulletin*, *114*(3), 494–509. https://doi.org/10.1037/0033-2909.114.3.494

[7] Hedges, L. V. (1981). Distribution theory for Glass's estimator of effect size and related estimators. *Journal of educational statistics*, *6*(2), 107–128. https://doi.org/10.3102/10769986006002107

[8] Efron, B., & Tibshirani, R. J. (1994). *An introduction to the bootstrap*. Chapman and Hall/CRC. An introduction to the bootstrap

[9] Holland, J. H. (1992). *Adaptation in natural and artificial systems: An introductory analysis with applications to biology, control, and artificial intelligence*. MIT press. https://doi.org/10.7551/mitpress/1090.001.0001

[10] Storn, R., & Price, K. (1997). Differential evolution – a simple and efficient heuristic for global optimization over continuous spaces. *Journal of global optimization*, *11*(4), 341–359. https://doi.org/10.1023/A:1008202821328

[11] Eberhart, R., & Kennedy, J. (1995). Particle swarm optimization. *Proceedings of the IEEE international conference on neural networks* (Vol. 4, pp. 1942–1948). IEEE. https://doi.org/10.1109/ICNN.1995.488968

[12] Kirkpatrick, S., Gelatt, C. D., & Vecchi, M. P. (1983). Optimization by simulated annealing. *Science*, *220*(4598), 671–680. https://doi.org/10.1126/science.220.4598.671

[13] Mirjalili, S., Mirjalili, S. M., & Lewis, A. (2014). Grey wolf optimizer. *Advances in engineering software*, *69*, 46–61. https://doi.org/10.1016/j.advengsoft.2013.12.007

[14] Tanabe, R., & Fukunaga, A. (2013). Success-history based parameter adaptation for differential evolution. *2013 IEEE congress on evolutionary computation* (pp. 71–78). IEEE. https://doi.org/10.1109/CEC.2013.6557555

[15] Tanabe, R., & Fukunaga, A. S. (2014). Improving the search performance of shade using linear population size reduction. *2014 IEEE congress on evolutionary computation (CEC)* (pp. 1658–1665). IEEE. https://doi.org/10.1109/CEC.2014.6900380

[16] Sun, X., Jiang, L., Shen, Y., Kang, H., & Chen, Q. (2020). Success history-based adaptive differential evolution using turning-based mutation. *Mathematics*, *8*(9), 1565. https://doi.org/10.3390/math8091565

[17] Hansen, N. (2006). The CMA evolution strategy: A comparing review. In *Towards a new evolutionary computation: Advances in the estimation of distribution algorithms* (pp. 75–102). Berlin, Heidelberg: Springer Berlin Heidelberg. https://doi.org/10.1007/3-540-32494-1_4

[18] Burke, E. K., Gendreau, M., Hyde, M., Kendall, G., Ochoa, G., Özcan, E., & Qu, R. (2013). Hyper-heuristics: A survey of the state of the art. *Journal of the operational research society*, *64*(12), 1695–1724. https://doi.org/10.1057/jors.2013.71

[19] López-Ibáñez, M., Dubois-Lacoste, J., Pérez Cáceres, L., Birattari, M., & Stützle, T. (2016). The irace package: Iterated racing for automatic algorithm configuration. *Operations research perspectives*, *3*, 43–58. https://doi.org/10.1016/j.orp.2016.09.002

[20] Auer, P., Cesa-Bianchi, N., & Fischer, P. (2002). Finite-time analysis of the multiarmed bandit problem. *Machine learning*, *47*(2), 235–256. https://doi.org/10.1023/A:1013689704352

[21] Sutton, R. S., Barto, A. G., & Barto, A. (1998). *Reinforcement learning: An introduction*. MIT press Cambridge. 10.1017/S0263574799271172

[22] Bubeck, S., & Cesa-Bianchi, N. (2012). Regret analysis of stochastic and nonstochastic multi-armed bandit problems. *Foundations and trends in machine learning*, *5*(1), 1–122. https://doi.org/10.1561/2200000024

[23] Morris, M. D. (1991). Factorial sampling plans for preliminary computational experiments. *Technometrics*, *33*(2), 161–174. https://doi.org/10.1080/00401706.1991.10484804