

Igor DUDKIEWICZ
Politechnika Wrocławska

DYNAMICZNE DOSTRAJANIE PARAMETRÓW KAR W ALGORYTMIE QAOA NA POTRZEBY WAŻENIA OGRANICZEŃ

Streszczenie. W pracy zbadano integrację parametrów kar w formulacji QUBO z klasyczną pętlą optymalizacyjną algorytmu Quantum Approximate Optimization Algorithm (QAOA). Zamiast traktować współczynniki kar jako stałe hiperparametry, są one optymalizowane wspólnie z parametrami wariacyjnymi QAOA przy użyciu algorytmu COBYLA. Podejście oceniono na przykładzie wielowymiarowego problemu plecakowego z wykorzystaniem dwóch metod normalizacji.

DYNAMIC CONSTRAINT WEIGHTING IN QAOA THROUGH INTEGRATED PENALTY PARAMETER OPTIMIZATION

Summary. This paper investigates the integration of QUBO penalty parameters into the classical optimization loop of the Quantum Approximate Optimization Algorithm (QAOA). Instead of treating penalty coefficients as fixed hyperparameters, they are optimized jointly with QAOA variational parameters using COBYLA. The approach is evaluated on the Multi-Dimensional Knapsack Problem under two normalization schemes.

1. Introduction

A recurring pattern in classical computing is that computational challenges are increasingly addressed through scaling hardware resources rather than through the development of fundamentally novel algorithmic approaches. While this trajectory is understandable—given its predictability and lower risk—it raises the question of whether similar dynamics will emerge in quantum computing. In this context, there is particular value in exploring simple, principled modifications that improve the performance of existing quantum algorithms without significantly increasing their complexity.

The Quantum Approximate Optimization Algorithm (QAOA), introduced by Farhi et al. [5], is among the most prominent approaches for tackling NP-hard combinatorial optimization problems on near-term quantum devices. By alternating between a problem-dependent cost Hamiltonian and a mixing Hamiltonian, QAOA attempts to prepare a quantum state that encodes high-quality solutions. However, its performance depends critically on the selection of variational parameters, which must be optimized through a classical feedback loop. This optimization problem is itself non-trivial, often exhibiting non-convex landscapes and, on real hardware, vanishing gradients due to noise-induced barren plateaus [9, 2, 10, 3].

In QUBO-based formulations of constrained problems, an additional challenge arises in the form of the penalty parameter, which determines how constraint violations are weighted relative to the objective function. If set too low, infeasible solutions may be favored; if set too high, the optimisation landscape becomes dominated by penalty terms, obscuring the structure of the original objective. The principled selection of this parameter remains an open problem and has recently attracted increasing attention [7, 6, 8, 1].

In our previous work [4], we conducted a detailed empirical study of the Constrained Optimization by Linear Approximation (COBYLA) optimizer within the QAOA framework applied to the Multi-Dimensional Knapsack Problem (MKP). That study established a behavioural baseline for the QAOA–COBYLA loop across simulation and quantum hardware environments, providing a consistent reference point for evaluating modifications to the optimisation process.

In this paper, we investigate a simple extension of the standard QAOA–COBYLA workflow: instead of treating the penalty parameter as a fixed hyperparameter, we incorporate it directly into the classical optimisation loop as an additional variable. In this approach, COBYLA simultaneously optimises the QAOA variational parameters and the penalty parameter itself, with the aim of improving solution quality through dynamic adjustment of constraint weighting during optimisation. An additional practical advantage of this formulation is that it reduces the need to select a suitable penalty parameter *a priori*, which is often problem-dependent and difficult to determine. By incorporating P directly into the optimisation loop, the method shifts this burden from manual tuning to the classical optimiser, potentially simplifying the application of QAOA in settings where little prior knowledge about appropriate penalty scales is available.

We evaluate this integrated approach on the Multi-Dimensional Knapsack Problem using the same experimental framework as in our previous work, enabling direct comparison with the established baseline. We consider two normalisation schemes that allow results obtained under different penalty values to be meaningfully compared, and we analyse their impact on convergence behaviour and solution quality across multiple configurations.

2. Problem Formulation and Experimental Setup

In this work, we adopt the Multi-Dimensional Knapsack Problem (MKP) and its QUBO formulation as introduced in our previous study [4]. In this section, we briefly summarise only the elements of the formulation and experimental setup that are necessary to understand the integration of the penalty parameter into the optimisation loop.

2.1. QUBO Mapping

To apply QAOA to the MKP, the problem is mapped to a Quadratic Unconstrained Binary Optimization (QUBO) formulation. We follow the standard construction used, for example, in [7]. The cost Hamiltonian combines the objective of maximising total value with penalty terms enforcing the capacity constraints:

$$H = - \sum_{i=1}^N v_i x_i + \sum_{k=1}^D P_k \left(\sum_{i=1}^N w_{ik} x_i + \sum_{j=0}^M 2^j z_j - W_k \right)^2 \quad (1)$$

where v_i and w_{ik} denote the value and weight of item i , W_k is the capacity of dimension k , and z_j are auxiliary slack variables used to encode inequality constraints.

This formulation leads to dense quadratic interactions between variables, which translate into quantum circuits with a high number of two-qubit gates. As observed in [4], this presents a significant challenge for current NISQ devices, where gate noise and readout errors can obscure the energy differences that guide the classical optimisation loop.

2.2. Experimental Testbed

We use the same set of randomly generated MKP instances as in [4], with problem sizes defined by $N \in \{3, 4\}$ items and $D \in \{2, 3\}$ constraints. Item values and weights are drawn from bounded integer ranges, and capacities are selected to ensure non-trivial constraint satisfaction.

As in our previous work, performance is evaluated using the probability mass assigned to the optimal solution in the final measurement distribution, rather than a binary success indicator. The testbed is restricted to instances for which the QAOA ansatz is capable of producing the optimal solution under baseline conditions, allowing us to focus on the behaviour of the optimisation process itself.

2.3. Integration of the Penalty Parameter into the COBYLA Loop

The central contribution of this work is the incorporation of the QUBO penalty parameter P into the classical optimisation loop. In the standard QAOA–COBYLA workflow, P is treated as a fixed hyperparameter selected prior to optimisation. Here, we instead treat P as an additional optimisation variable, allowing COBYLA to adjust it dynamically alongside the variational parameters γ and β .

We need to consider how results obtained under different values of the penalty parameter P can be meaningfully compared. Since the penalty term directly scales contributions to the QUBO objective, larger values of P almost invariably result in larger absolute circuit energy values. As a consequence, raw energies cannot be compared across runs in a meaningful way, even when the underlying solution quality may be similar.

To address this issue, we consider two normalisation schemes. The first is a simple scaling of the form E/P . This approach does not require any knowledge of the optimal solution and can therefore be applied directly during optimisation. The motivation for this choice comes from empirical observations in our previous experiments, where the optimal solution energy was often close to zero or slightly negative. In such cases, dividing by P produces a quantity that remains informative about relative convergence behaviour while removing the dominant scaling effect introduced by the penalty term.

The second approach is an idealised normalisation of the form $(E - E_{\text{opt}})/P$, where E_{opt} denotes the true optimal energy of the problem instance. This transformation centers the objective function around the global optimum, providing a more interpretable measure of how close a given solution is to optimality after accounting for the penalty scaling. While this quantity is not available in general during optimisation, it can in principle be approximated using classical heuristics or exact solvers for small instances.

Normalisation schemes are essential for interpreting the behaviour of the integrated COBYLA loop. Without them, comparisons across different values of P would conflate improvements in solution quality with simple rescaling effects induced by the

penalty term. By introducing these transformations, we ensure that observed changes in the objective reflect meaningful differences in optimisation behaviour rather than artefacts of parameter scaling.

Finally, since both normalisation schemes involve division by P , care must be taken to avoid numerical instability when P approaches zero. Such regimes are also conceptually problematic, as a vanishing or negative penalty would fail to properly enforce constraints and could effectively invert their role in the objective function. We therefore enforce a lower bound $P \geq 1$ in all experiments to maintain both numerical stability and a consistent interpretation of the penalty term.

2.4. Optimizer Configuration

All aspects of the experimental protocol not related to the dynamic adjustment of the P parameter remain consistent with those described in [4], ensuring that the results presented here can be directly compared to the established baseline.

We evaluate the optimisation process across QAOA circuit depths $p \in \{2, 3, 10\}$, corresponding to both practical NISQ limits and higher-depth theoretical behaviour. We employ a multi-start strategy to mitigate the impact of local minima and barren plateaus, using 5 random initialisations of the optimisation parameters.

3. Results and Discussion

The central objective of this study was to investigate whether the penalty coefficient can be incorporated directly into the QAOA–COBYLA optimisation loop rather than being selected manually in advance. To evaluate this approach, we examined two normalisation schemes designed to make objective values comparable across different penalty regimes. The following subsections present the results obtained for each scheme and compare their behaviour against the fixed-penalty benchmark.

3.1. Benchmark: Standard QAOA–COBYLA Performance

The behavioural baseline for the standard QAOA–COBYLA loop with fixed penalty parameters was established in our previous work [4], where a detailed empirical characterisation was conducted across the noiseless statevector simulator, the shots-based Aer simulator, and the IQM Garnet quantum processor. We summarise here only the findings most relevant to interpreting the integrated penalty parameter results that follow.

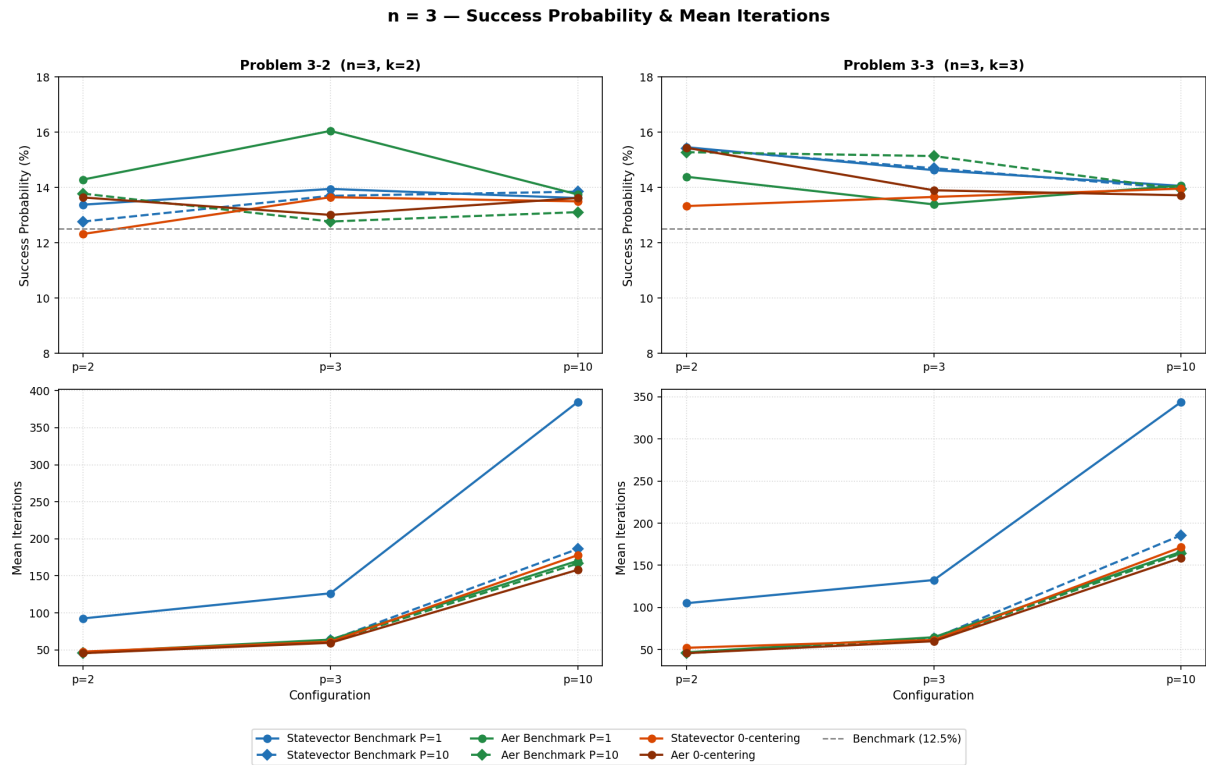
On simulated backends, success probabilities remained modestly but consistently above the random-guess benchmarks of 12.5% for $n = 3$ and 6.25% for $n = 4$ across all configurations, with no strong systematic advantage observed for either penalty regime. Convergence times were broadly stable across problem instances and circuit depths. The notable exception was the statevector simulator under low penalty ($P = 1$), where mean iteration counts were dramatically higher than in any other configuration, with a number of runs exhausting the iteration budget of 2000 entirely. This behaviour was absent from the Aer simulator and hardware results, where sampling noise effectively smooths the energy landscape and accelerates apparent convergence. On the quantum hardware, convergence times were comparable to those of the Aer simulator across all tested configurations, indicating that the COBYLA feedback loop is not significantly disrupted by hardware noise at the problem scales tested.

Success probabilities on the IQM Garnet hardware fell below the random-guess benchmark across all problem instances and configurations. As argued in [4], this is not a failure of the optimiser: the QUBO energy landscapes of the tested instances contain dense clusters of near-degenerate low-energy states, and in the presence of hardware noise COBYLA converges on parameter values that distribute probability mass across this cluster rather than concentrating it on the unique global minimum. The optimiser is minimising expected energy as intended – the apparent deficit arises from a mismatch between that objective and the metric of exact optimal solution frequency.

3.2. Integrated Penalty Parameter Optimisation

Two alternative normalisation strategies were investigated. The first uses a zero-centered reference point and corresponds to the simple E/P normalisation discussed earlier, while the second centres energies relative to the optimal solution value. We begin with the zero-centered formulation.

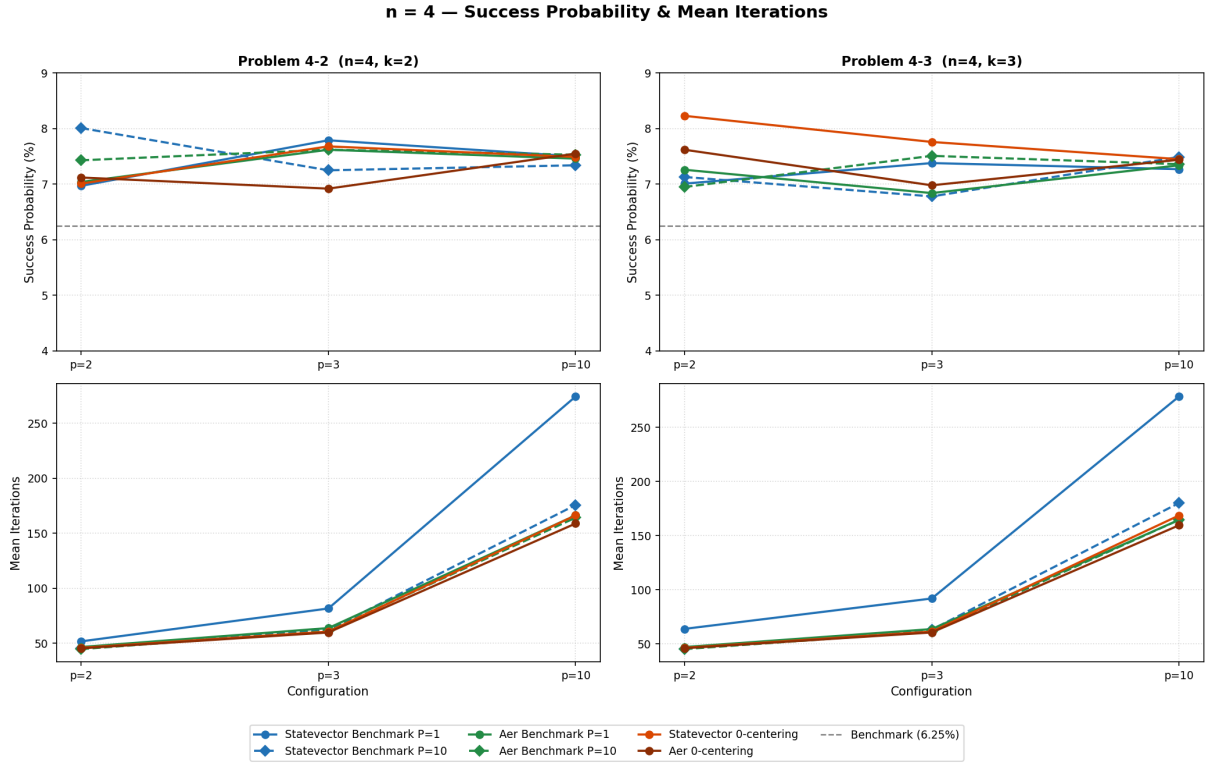
3.2.1. Zero-Centered Normalisation



Rys. 1. Results for $n = 3$ under E/P normalisation.

Given that P showed very limited movement throughout these experiments, particularly in the zero-centered tests conducted first, hardware runs were judged unlikely to yield additional insight at this stage and were omitted. We intend to revisit the integrated penalty approach on a more QAOA-amenable problem class in future work.

In initial tests using this normalisation scheme, P was initialised at a fixed value. We observed very limited movement of the penalty parameter within the COBYLA loop across these runs, with many executions leaving P entirely unchanged from its starting value. We attribute this to the combination of two factors: the barren plateaus that characterise the optimisation landscape of the Multi-Dimensional Knapsack Problem, and the



Rys. 2. Results for $n = 4$ under E/P normalisation.

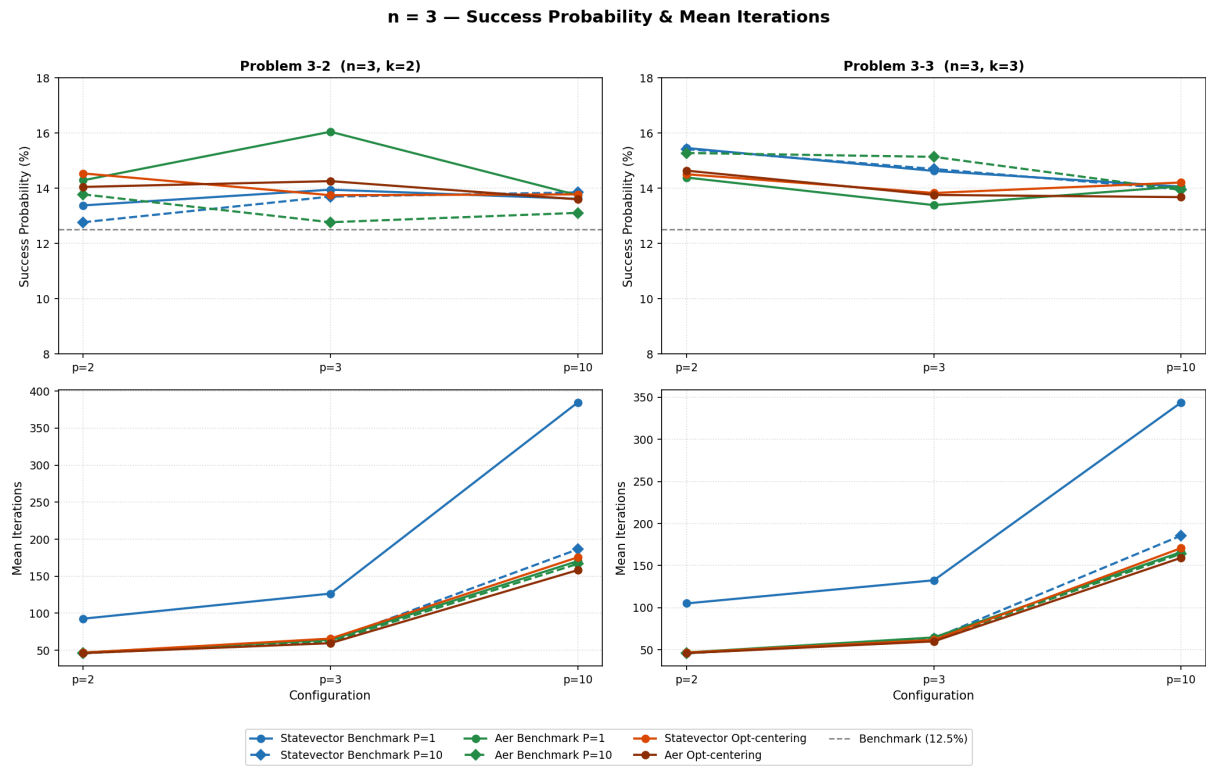
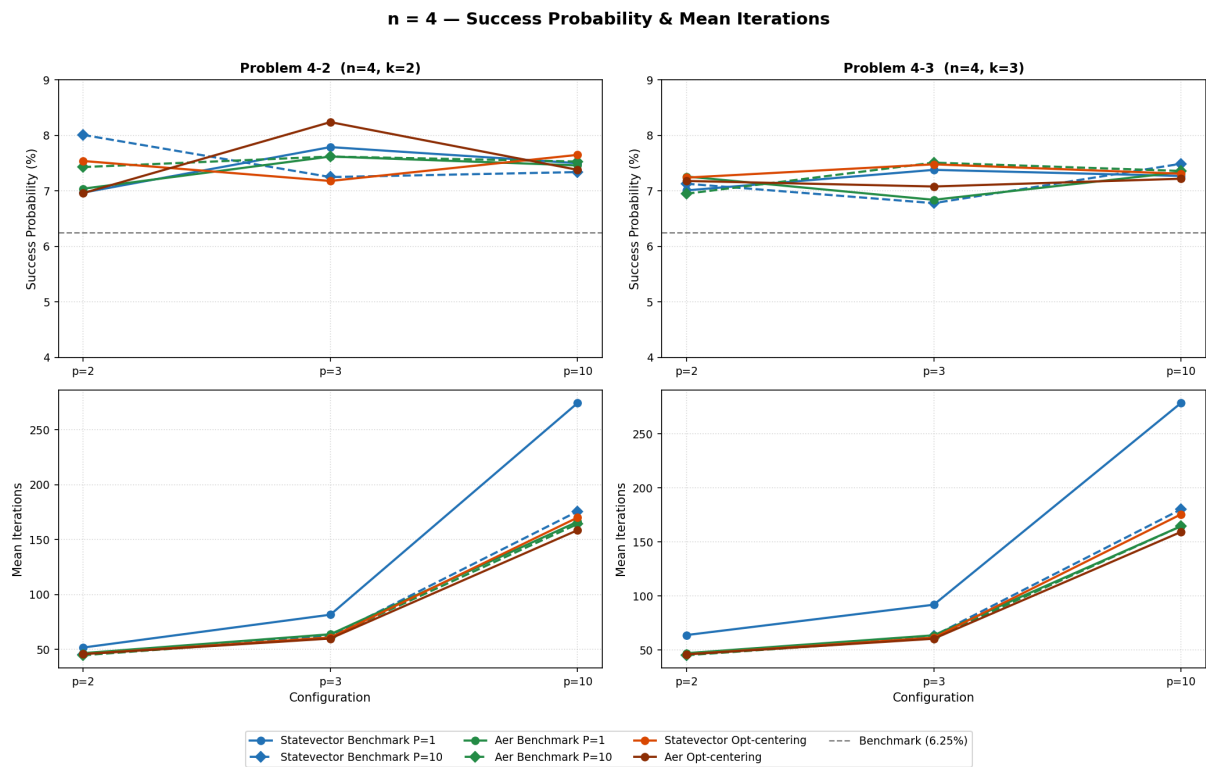
nature of the E/P normalisation itself, which combined with barren plateaus produced an optimisation surface that is nearly flat with respect to P for this problem class. Together, these effects left COBYLA with insufficient gradient signal to drive meaningful movement in the penalty parameter.

We therefore adopted a different initialisation strategy, drawing the starting value of P at random from the range $[0.5 \times \max(r_i), 50 \times \max(r_i)]$ for each run. The results of these tests are shown in figures 1 and 2. The outcomes were not strongly positive: while convergence times showed a marginal improvement over the benchmark, success probabilities were slightly worse across many configurations, especially for $n = 3$.

A question we examined carefully during this analysis was whether the optimisation exhibited any systematic bias toward either large or small values of P – a finding that would be concerning, as it would suggest the comparative framework was fundamentally flawed. We are pleased to report that no such bias was observed. Across the full set of experiments, the run with the lowest average energy corresponded to a small value of P in some cases, a large value in others, and for the vast majority of problems it corresponded to neither extreme. This suggests that when the appropriate value of P is uncertain – or when running optimisation across large sets of similar problem instances – initialising P randomly rather than fixing it uniformly across all runs may be a promising strategy. We note, however, that this conjecture would require substantially more testing to substantiate, and we identify it as a direction for future work.

3.2.2. Optimal-Centered Normalisation

Under the optimal-centered normalisation scheme, convergence times again showed a slight improvement over the benchmark. More notably, success probabilities were no longer worse than the benchmark, though they did not show a strong systematic im-

Rys. 3. Results for $n = 3$ under optimal-centred normalisation.Rys. 4. Results for $n = 4$ under optimal-centred normalisation.

provement either. The more striking observation from these experiments lies not in the aggregate results but in the distribution of outcomes across individual problem instances: we observed a number of instances with success probabilities below 1%, the lowest being 0.04%, while a number of others approached 50% success probability – a spread of outcomes far wider than anything seen in the benchmark experiments.

We also investigated the behaviour of this scheme under fixed initialisation of P . Unlike the zero-centered case, where most runs left P entirely unchanged, here almost all runs resulted in some movement of the penalty parameter. However, these movements remained very modest: for an initial value of $P = 10$, the majority of problems converged to values within the range $[9, 11]$, with movements exceeding 1 being rare. This behaviour was consistent across different initial values of P , ruling out the interpretation that $P = 10$ simply happened to be near the optimum for these instances. We again attribute this limited mobility to the barren plateau structure of the MKP optimisation landscape, which constrains the ability of COBYLA to make large coordinated movements in any direction, including in the penalty parameter dimension.

3.2.3. Comparison of Normalisation Schemes

Across the full set of experiments, the optimal-centered normalisation consistently outperformed the zero-centered scheme in terms of success probability, with only a small number of exceptions attributable to the stochastic nature of the tests. This systematic difference further supports the conclusion that the quality of the normalisation reference point has a meaningful impact on the ability of the integrated loop to drive P toward useful values.

3.3. Variance Structure and Expressibility

Tabela 1

Success probability standard deviation (%) for the integrated- P experiments.

	$n = 3, k = 2$			$n = 3, k = 3$			$n = 4, k = 2$			$n = 4, k = 3$		
	$p = 2$	$p = 3$	$p = 10$	$p = 2$	$p = 3$	$p = 10$	$p = 2$	$p = 3$	$p = 10$	$p = 2$	$p = 3$	$p = 10$
<i>Aer simulator</i>												
Optimal-cent.	7.03	7.87	3.40	5.56	7.41	4.65	4.15	4.37	2.76	5.35	4.61	3.79
Zero-cent.	7.26	11.23	3.96	7.61	7.28	4.24	5.47	5.40	2.94	4.30	6.45	3.61
<i>Statevector simulator</i>												
Optimal-cent.	8.37	8.78	3.53	8.67	6.73	4.36	4.45	4.83	2.81	4.71	4.88	3.78
Zero-cent.	6.62	6.70	3.49	7.84	10.32	4.40	4.86	3.99	2.81	7.83	3.64	3.62

The standard deviation data in table 1 reveals a pattern that contrasts with the benchmark results reported in [4]. In the benchmark experiments, success probability standard deviation decreased consistently with circuit depth p across all environments and configurations. Here, that monotonic relationship frequently breaks down: across both normalisation schemes and both backends, the $p = 3$ standard deviation exceeds the $p = 2$ value in a substantial number of configurations – for instance, under zero-centered normalisation on Aer, the $n = 3, k = 2$ standard deviation rises from 7.26% at $p = 2$ to 11.23% at $p = 3$ before falling to 3.96% at $p = 10$. The $p = 10$ values do generally return to lower levels, but the intermediate elevation at $p = 3$ is a recurring feature that was absent from the benchmark data.

The aggregate standard deviation values are broadly moderate, which taken alone might suggest the integrated- P loop is simply less variable than the benchmark. We

argue, however, that this reading is incomplete. The actual experiments produced a wider range of per-instance outcomes than the benchmark: individual problem instances reached success probabilities as low as 0.04% and as high as nearly 50%, a spread not observed under fixed penalty. The moderate aggregate standard deviations arise because many runs clustered tightly near the mean for a given configuration, while a subset explored qualitatively different regions of the solution space. This pattern – moderate average variance coexisting with pronounced individual outliers – is consistent with the integrated- P loop having greater *expressibility* than the fixed-penalty baseline: the optimisation is more sensitive to the specific structure of individual problem instances, and for some problems this sensitivity translates into substantially better performance, while for others the additional freedom of a dynamic penalty does not confer a clear advantage.

This interpretation is further supported by the benchmark finding that for this problem class the COBYLA–QAOA loop can converge to regions of parameter space that do not contain the global minimum, due to the near-degenerate energy landscape of the MKP [4]. The occurrence of very low success probability outliers in the integrated experiments is consistent with the same phenomenon: when the penalty parameter converges to a value that minimises expected circuit energy without concentrating probability mass on the global minimum – a rational outcome on the near-degenerate MKP landscape – the result by the metric of exact optimal solution frequency can be worse than the fixed-penalty case.

3.3.1. Parameter Scaling Experiments

In related work applying the QAOA–COBYLA loop to the Travelling Salesman Problem (TSP), we observed that unscaled rotational parameters – which in that setting can reach values on the order of thousands due to the structure of the penalty terms – significantly degraded optimiser performance. Rescaling parameters to the range $[-1, 1]$ produced substantial improvements, with representative success probabilities increasing from approximately 2% to 30–40% in some experiments.

These observations motivated an investigation into whether a similar effect might be present in the MKP setting, despite the rotational parameters here being considerably smaller in magnitude. We therefore conducted a limited set of experiments using parameter scaling to the range $[-1, 1]$.

The results did not support the hypothesis that unscaled parameters were responsible for the modest performance observed in the MKP experiments. Success probabilities under scaled parameters were not higher than in the unscaled baseline; in the tested instances they were in fact marginally lower.

In the integrated-penalty experiments using zero-centered normalisation, parameter scaling additionally introduced a tendency toward smaller values of P , whereas this effect was not observed under optimal-centered normalisation. Although scaling increased the magnitude of P movements during optimisation, the parameter still failed to explore the full available range and remained strongly dependent on its initial value.

Taken together, these observations indicate that parameter scaling does not provide a meaningful performance improvement for the MKP instances considered here. In particular, the absence of any increase in success probability under scaled parameters suggests that the limited performance of both the baseline and integrated-penalty approaches arises primarily from the structure of the optimisation landscape itself rather than from COBYLA struggling with large rotational parameters.

4. Conclusion

Building on the behavioural baseline established in our previous work [4], this paper investigated a simple extension of the standard QAOA–COBYLA workflow in which the QUBO penalty parameter P is incorporated directly into the classical optimisation loop as an additional variable. We evaluated this approach on the Multi-Dimensional Knapsack Problem under two normalisation schemes, using the same experimental framework as the benchmark study to enable direct comparison.

The results present a mixed but interpretable picture. Neither normalisation scheme produced a decisive improvement in aggregate success probability over the benchmark, and we attribute a significant part of this to the choice of testbed: the MKP is poorly suited to QAOA, its optimisation landscape characterised by barren plateaus and near-degenerate energy clusters that limit the performance of the variational ansatz regardless of penalty configuration. In this setting, distinguishing the effect of the integrated penalty approach from the inherent limitations of the problem class is difficult.

Nevertheless, several meaningful observations emerge. The absence of any systematic bias toward extreme values of P under random initialisation lends validity to the comparative framework itself, and suggests that random initialisation of P across large sets of similar instances may be a worthwhile strategy when a good penalty value is not known in advance. The optimal-centered normalisation consistently outperformed the zero-centered scheme across the full set of experiments, with only isolated exceptions attributable to the stochastic nature of the tests, indicating that even a rough classical approximation of the optimal energy – cheaply obtainable for small instances – is sufficient to meaningfully improve the behaviour of the integrated loop. Most notably, the integrated approach produced a substantially wider spread of per-instance outcomes than the fixed-penalty baseline, with success probabilities ranging from near zero to nearly 50%, alongside a disruption of the monotonic decrease in outcome variance with circuit depth observed in the benchmark. We interpret this as evidence of greater expressibility in the QAOA–COBYLA system when P is treated as a free variable: the optimisation becomes more sensitive to the structure of individual problem instances, which on a well-suited problem class could translate into meaningful performance gains.

The primary direction for future work is therefore clear: the integrated penalty parameter approach should be evaluated on a problem class for which the COBYLA–QAOA loop performs reliably under baseline conditions, providing a reference point strong enough to draw conclusive findings about the benefit of dynamic penalty tuning. We also identify the random initialisation of P as a promising idea worth further investigation, particularly for large problem sets where identifying a good penalty value in advance is non-trivial.

REFERENCES

1. Ayodele M.: . Penalty weights in qubo formulations: Permutation problems 2022.
2. Cerezo M., Arrasmith A., Babbush R., Benjamin S. C., Endo S., Fujii K., McClean J. R., Mitarai K., Yuan X., Cincio L., Coles P. J.: Variational quantum algorithms. *Nature Reviews Physics*, 2021, Vol.3, p. 625–644.

3. Cerezo M., Sone A., Volkoff T., Cincio L., Coles P. J.: Cost function dependent barren plateaus in shallow parametrized quantum circuits. *Nature Communications*, 2021, Vol.12(1), p. 1791.
4. Dudkiewicz I.: Quantum circuit depth and penalty parameters in qaoa: A cobyla convergence study on the multi-dimensional knapsack problem. *Proceedings of the International Conference on Soft Computing Models in Industrial and Environmental Applications (SOCO 2025)*, 2026,. To appear.
5. Farhi E., Goldstone J., Gutmann S.: . A quantum approximate optimization algorithm 2014.
6. Glover F., Kochenberger G., Du Y.: . A tutorial on formulating and using qubo models 2019.
7. Güney E., Ehrental J., Hanne T.: Qubo formulations and characterization of penalty parameters for the multi-knapsack problem. *IEEE Access*, 2025, Vol.13, p. 47086–47098.
8. Kochenberger G., Hao J.-K., Glover F., Lewis M., Lü Z., Wang H., Wang Y.: The unconstrained binary quadratic programming problem: A survey. *Journal of Combinatorial Optimization*, 2014, Vol.28(1), p. 58–81.
9. McClean J. R., Boixo S., Smelyanskiy V. N., Babbush R., Neven H.: Barren plateaus in quantum neural network training landscapes. *Nature Communications*, 2018, Vol.9(1), p. 4812.
10. Wang S., Fontana E., Cerezo M., Sharma K., Sone A., Cincio L., Coles P. J.: Noise-induced barren plateaus in variational quantum algorithms. *Nature Communications*, 2021, Vol.12(1), p. 6961.