KSOS-BO: IMPROVING SAMPLING IN BAYESIAN OPTIMIZATION VIA KERNEL SUM OF SQUARES

Anonymous authors
Paper under double-blind review

ABSTRACT

Bayesian Optimization (BO) is an effective framework for globally optimizing functions whose evaluations are expensive. It is particularly effective for optimizing functions defined over continuous domains and explicitly handles stochastic noise in evaluations. As a result, it is widely applied in areas such as hyperparameter tuning, robotics policy search, and scientific experiment design, where sample efficiency is essential. Its two-step procedure consists of model fitting followed by optimization of the acquisition function, which is often treated as a generic black-box problem despite its structured nature. In this work, we introduce KSOS-BO, a kernel-based framework for BO acquisition optimization. KSOS-BO integrates the recently developed KernelSOS framework to formulate the optimization of the acquisition function as a semidefinite program with kernel-induced representations, enabling a structured global search. Across a diverse set of benchmark functions with varying landscape properties, KSOS-BO achieves the best final regret on 13 out of 20 benchmarks compared with Sobol Search, Differential Evolution, CMA-ES, and multi-start L-BFGS-B. On these 13 benchmarks, it improves final regret over the second-best optimizer by 81.08% on average. Wall-clock experiments further show that, under the optimizer configurations considered in this study, KSOS-BO can reach high-quality solutions earlier on many benchmarks despite the additional semidefinite optimization step. Limitations include reduced effectiveness on functions with steep drops or plate-shaped regions.

1 INTRODUCTION

Bayesian Optimization (BO) (Shahriari et al., 2016) is a standard approach for optimizing expensive black-box functions. It builds a surrogate model from observed data and selects new evaluation points by optimizing an acquisition function. While substantial effort has been devoted to improving surrogate models and acquisition function design (Snoek et al., 2015; Wang et al., 2017), the optimization of the acquisition function itself remains a critical and actively researched component. In practice, this inner optimization problem is often non-convex and multimodal, rendering it challenging to solve as the dimensionality grows. As a result, the overall performance of BO can be strongly limited by the quality of acquisition optimization (Gan et al., 2021).

Existing approaches to acquisition optimization include gradient-based local methods when reliable derivatives are available (Wilson et al., 2018), as well as population-based derivative-free optimizers such as Differential Evolution (DE) (Das & Suganthan, 2010) and Covariance Matrix Adaptation Evolution Strategy (CMA-ES) (Hansen, 2016). Despite their broad use, these optimizers typically treat acquisition optimization as either pointwise candidate selection, local refinement, or generic black-box search, and do not explicitly exploit the smooth kernel-induced structure of acquisition landscapes (Eggensperger et al., 2015). Their effectiveness can degrade when the acquisition function is highly nonconvex, multimodal, or sensitive to initialization. Population-based methods may require many acquisition evaluations to adequately explore the landscape, while local gradient-based methods may converge to suboptimal local optima depending on their starting points.

These challenges are not unique to acquisition optimization but are inherent to a broader class of non-convex global optimization problems. In such settings, local optimization methods can become trapped in suboptimal solutions, leading to degraded performance (Nocedal & Wright, 2006). To address this issue, Sum-of-Squares (SOS) and moment-based relaxations have been extensively

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

studied as principled approaches for global optimization, providing convex surrogates that enable certifiable solutions (Lasserre, 2001; Parrilo, 2003). However, classical SOS methods are restricted to polynomial optimization problems (Lasserre, 2009), which limits their applicability in modern machine learning (Snoek et al., 2012) and data-driven settings where objective functions are often non-polynomial or only accessible through function evaluations. In addition, the computational cost of SOS methods can scale unfavorably with the degree of relaxation, making them difficult to apply in high-dimensional problems.

Kernel Sum-of-Squares (KernelSOS) (Rudi et al., 2025) extends SOS beyond polynomial representations by replacing polynomial bases with kernel functions. This formulation constructs function representations from a finite set of samples and enables global optimization using only function evaluations, thereby broadening the class of admissible objective functions. Importantly, the surrogate models commonly used in BO can induce relatively smooth acquisition functions, which align well with standard kernel choices. As a result, KernelSOS can effectively approximate and optimize these acquisition landscapes. Moreover, it replaces explicit polynomial representations with a sample-based formulation, which makes it naturally suited to data-driven settings.

In this paper, we demonstrate the potential of KernelSOS as an acquisition function optimizer in BO. We provide a systematic evaluation and application-specific considerations such as automatic kernel lengthscale determination based on the GP surrogate. In particular, KSOS-BO achieves the best final regret on 13 out of 20 benchmarks, demonstrating strong performance across a range of non-convex and multi-modal objective landscapes. We further report wall-clock results to characterize the computational trade-off introduced by the KernelSOS optimization step. Overall, these results show that improving the acquisition optimization step alone can yield measurable improvements in overall BO performance.

2 Related Work

2.1 Sum-of-Squares and Kernel Methods

SOS and moment-based hierarchies have been extensively studied for global optimization and polynomial analysis. They provide a principled framework for constructing convex relaxations with global guarantees (Parrilo, 2000) and have been widely applied in control (Tedrake et al., 2010), estimation (Yang & Carlone, 2022), and in many practical settings, where they enable certifiable global optima (Lasserre, 2015).

KernelSOS can be viewed as a generalization of polynomial SOS methods by replacing finite-dimensional polynomials with functions in more general Reproducing Kernel Hilbert Spaces (RKHS) (Paulsen & Raghupathi, 2016). Through kernel representations (Scholkopf & Smola, 2018), this formulation naturally extends to infinite-dimensional feature spaces and allows non-parametric function representations based on finite samples. Prior work has established the effectiveness of KernelSOS in optimal control, value function learning (Berthier et al., 2022), and complex robotics control and estimation tasks (Groudiev et al., 2026; Wei & Dümbgen, 2026). We extend its use to a crucial and actively researched component of BO: acquisition function optimization (Wilson et al., 2018), where the objective is typically non-convex and multimodal, but sufficiently smooth for most common choices of surrogate functions.

2.2 Acquisition Optimization in Bayesian Optimization

BO addresses black-box optimization by iteratively fitting a surrogate model and selecting query points through an acquisition function (Frazier, 2018). For a fixed set of observations, the acquisition function defines a deterministic objective over the input domain. This objective is typically non-convex and multimodal, exhibiting numerous local optima (Jones et al., 1998). In practice, acquisition optimization is often treated as a generic black-box optimization problem (Bergstra et al., 2011). Existing approaches include gradient-based methods, which exploit local derivative information when available (Wilson et al., 2018), as well as derivative-free optimizers such as DE and CMA-ES (Das & Suganthan, 2010; Hansen, 2016). In low-dimensional settings, acquisition functions can also be optimized via grid search over a discretized domain (Brochu et al., 2009). While these methods are broadly applicable and easy to implement, they are not specifically designed to exploit the structural

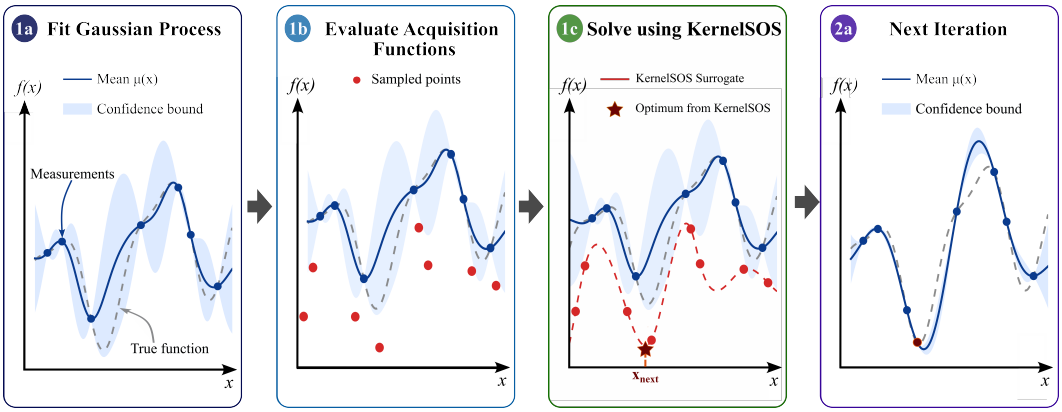


Figure 1: Overview of KSOS-BO. A Gaussian Process surrogate is fitted to the current measurement (1a), the acquisition function is evaluated at chosen points (1b), KernelSOS constructs a surrogate and minimizes it to find the next best sample (1c), and the selected point is evaluated to update the model (2a).

properties of acquisition functions, and may therefore suffer from inefficiency or suboptimal solutions in highly multimodal landscapes.

From a structural perspective, acquisition functions are induced by surrogate models and are inherently smooth (Seeger, 2004). These characteristics suggest that acquisition optimization is not an arbitrary black-box problem, but rather a structured optimization problem that can benefit from methods capable of leveraging global function representations. In contrast to purely sampling-based methods, population-based optimizers, and multi-start local refinement approaches, we formulate acquisition optimization through KernelSOS, which constructs a structured surrogate of the acquisition function using kernel-based representations. This enables global search over the acquisition landscape using only function evaluations, allowing it to better capture the global geometry of the acquisition function and provide a more effective mechanism for identifying high-quality query points in BO.

3 Method

We consider the problem of minimizing an expensive black-box function $f : \mathcal{X} \to \mathbb{R}$, where $\mathcal{X} \subset \mathbb{R}^d$ is a bounded domain. The objective is to identify the global minimizer

$$x^\star = \arg\min_{x \in \mathcal{X}} f(x). \tag{1}$$

Since direct optimization of f is prohibitively expensive, BO constructs a surrogate model based on a set of noisy observations $\mathcal{D}_t = \{(x_i, y_i)\}_{i=1}^t, \quad y_i = f(x_i) + \epsilon_i$, where $\epsilon_i \sim \mathcal{N}(0, \sigma_n^2)$ denotes observation noises. BO then selects new query points by optimizing an acquisition function $\alpha(x \mid \mathcal{D}_t)$. At each iteration, BO selects the next query point by solving an acquisition optimization problem

$$x_{t+1} = \arg\min_{x \in \mathcal{X}} \alpha(x \mid \mathcal{D}_t), \tag{2}$$

where $\alpha(x \mid \mathcal{D}_t)$ denotes the acquisition function induced by the current surrogate model. For acquisition functions that are conventionally maximized, such as Expected Improvement (EI), we equivalently optimize their negated forms. For example, EI maximization is written as minimizing $\alpha_{\mathrm{EI}}(x) = -\mathrm{EI}(x)$. For fixed $\mathcal{D}_t$, the acquisition function $\alpha(x)$ is a deterministic function over $\mathcal{X}$. The corresponding optimization problem $\min_{x \in \mathcal{X}} \alpha(x)$ is generally non-convex and may exhibit multiple local minima and irregular structure induced by the surrogate model. This inner problem can therefore be viewed as its own global optimization problem over a bounded domain. The performance of BO depends critically on the ability to obtain high-quality approximate solutions to this problem at each iteration within limited time.

Algorithm 1 KSOS-BO (KernelSOS-Based Acquisition Optimization for Bayesian Optimization)

Require: Initial dataset $\mathcal{D}_0$, acquisition $\alpha(\cdot)$, iterations T, sample budget N
Ensure: Best observed solution
1: **for** $t = 1, \ldots, T$ **do** ▷ Bayesian Optimization iteration
2: Fit surrogate $\mathcal{M}_t$ on $\mathcal{D}_{t-1}$ and define $\alpha_t(x)$
3: Sample $\{x_i\}_{i=1}^N$, evaluate $\{\alpha_t(x_i)\}_{i=1}^N$, and form kernel matrix K
4: Solve KernelSOS problem equation 5 and form minimizer equation 6
5: Evaluate $y_t = f(x_t)$ and update $\mathcal{D}_t = \mathcal{D}_{t-1} \cup \{(x_t, y_t)\}$
6: **end for**
7: **return** best solution in $\mathcal{D}_T$

3.1 Global Lower-Bound Reformulation

We first reformulate acquisition minimization through a global non-negativity condition, which motivates the subsequent SOS and KernelSOS formulations. For the acquisition function α, assuming that α attains its global minimum over $\mathcal{X}$, we have

$$c^\star = \min_{x \in \mathcal{X}} \alpha(x). \tag{3}$$

Equivalently, this problem can be expressed as finding the largest global lower bound of α:

$$\begin{aligned} c^\star = \max_{c \in \mathbb{R}} \quad & c \\ \text{s.t.} \quad & \alpha(x) - c \geq 0, \qquad \forall x \in \mathcal{X}. \end{aligned} \tag{4}$$

The constraint ensures that $c \leq \alpha(x)$ for all $x \in \mathcal{X}$, so maximizing c recovers the minimum value of α. However, enforcing global non-negativity over the entire domain is generally intractable. This lower-bound formulation motivates the SOS relaxation and its kernelized extension used in KSOS-BO. Additional background on SOS optimization is provided in Appendix A.4.

3.2 KSOS Formulation

The KernelSOS formulation replaces the finite-dimensional polynomial basis $v_d(x)$ with functions in an RKHS associated with a positive definite kernel $k(\cdot, \cdot)$. Rather than explicitly parameterizing the basis functions, the surrogate is represented through kernel evaluations over a finite set of sampled points $\{x_i\}_{i=1}^N$. Let $K \in \mathbb{R}^{N \times N}$ denote the kernel matrix with entries $K_{ij} = k(x_i, x_j)$. Using the representer theorem from (Marteau-Ferey et al., 2020), the resulting finite-dimensional optimization problem can be written as

$$\begin{aligned} c^\star_{\mathrm{KSOS}} = \max_{c \in \mathbb{R},\, B \succeq 0} \quad & c - \lambda \operatorname{Tr}(B) \\ \text{s.t.} \quad & \alpha(x_i) - c = \phi_i^\top B \phi_i, \quad i = 1, \ldots, N, \end{aligned} \tag{5}$$

where $\phi_i = [k(x_i, x_1), \ldots, k(x_i, x_N)]^\top$, $\operatorname{Tr}(\cdot)$ denotes the matrix trace, and $\lambda \geq 0$ is a regularization parameter controlling the complexity of the surrogate. The optimization variables in equation 5 define a kernel-based surrogate representation of the shifted acquisition function rather than an explicit optimizer in the original input space. Following the recovery strategy used in the original KernelSOS formulation, we extract a continuous candidate point as a weighted combination of the sampled points:

$$x^\star = \sum_{i=1}^N \mu_i^\star x_i, \tag{6}$$

where $\mu_i^\star$ denotes the dual variable of the constraint associated with sample x_i in equation 5. The form of equation 6 can be derived from the Lagrangian dual as detailed in (Groudiev et al., 2026). In general, the quality of this minimizer is not guaranteed as it depends on the appropriate choice of kernel parameters and sampling point locations. The candidate may also need to be projected back to the feasible set $\mathcal{X}$, which may require specialized tools for complex geometries of $\mathcal{X}$. For the present

Table 1: Summary of KSOS-BO performance across benchmark functions. The table reports the relative ranking of KSOS-BO and its percentage improvement over the second-best optimizer when ranked first, and relative to the best optimizer otherwise.

Type	Category	Function	KSOS-BO Rank	Improvement (%)
Synthetic benchmark	Multimodal	Ackley	**1**	99.20
		Rastrigin	**1**	78.46
		Levy	**1**	59.97
		Griewank	**1**	99.44
		Schwefel	5	-19
	Steep Drops	Michalewicz	5	-99.29
	Bowl Shaped	Rotated Hyper Ellipsoid	**1**	99.86
		Sphere	3	-206.22
		Sum of Different Powers	**1**	99.42
		Trid	**1**	92.05
	Plate Shaped	Zakharov	5	-701.82
	Valley Shaped	Dixon Price	**1**	99.99
		Rosenbrock	**1**	88.79
	Other	Powell	**1**	87.53
		Styblinski Tang	2	-2.23
Engineering benchmark	–	Borehole	**1**	99.58
		Branin	2	-84.39
		Antenna Array	**1**	4.74
		Camera Calibration	**1**	45.10
		Thomson Problem	2	-0.34

application with acquisition function classes and unconstrained or box-constrained variables, this simple projection procedure and candidate-recovery strategy performed sufficiently well.

The quality of the recovered candidate depends on the choice of kernel, the number of sampled points N, and the regularization parameter λ. The kernel determines the smoothness and expressive power of the induced surrogate, where smoother kernels tend to capture broader global structure while less smooth kernels may preserve more localized variations. The number of samples N controls the coverage of the acquisition landscape, and λ balances the lower-bound objective against the complexity of the KernelSOS representation through the trace penalty on B.

3.3 KernelSOS for Acquisition Optimization

Based on the formulation above, we apply KernelSOS to the acquisition optimization problem within each BO iteration. Given the current acquisition function $\alpha_t(x)$, we first construct a finite set of candidate points $\{x_i\}_{i=1}^N \subset \mathcal{X}$ and evaluate $\alpha_t(x_i)$ at these points. We then solve equation 5 using these acquisition values and recover a candidate optimizer using equation 6. The recovered point is used as the next BO query point. Compared with purely sampling-based acquisition optimizers, this procedure does not simply select the best point among the sampled candidates. Instead, it fits a structured KernelSOS representation from the sampled acquisition values and uses this representation to recover a continuous candidate. This allows KSOS-BO to exploit the global structure of the acquisition landscape under a fixed acquisition-evaluation budget.

3.4 Discussion

The proposed method differs from standard acquisition optimizers in two key aspects. First, it constructs a structured kernel-based representation of the acquisition landscape from sampled acquisition values, rather than simply selecting among sampled candidates or relying solely on local refinement. Second, it performs a structured global search through a finite-dimensional KernelSOS relaxation, providing a principled alternative for acquisition optimization. Figure 1 and Algorithm 1 summarize the resulting KSOS-BO procedure, with implementation details provided in Appendix B.1.

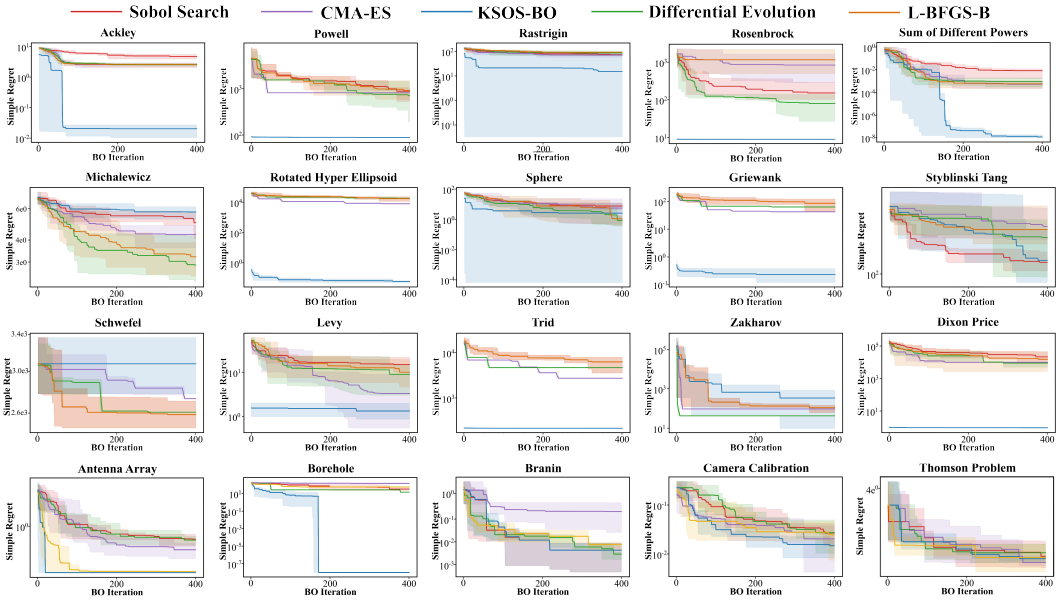


Figure 2: Convergence comparison on 20 benchmarks using EI over 400 iterations. The curves show simple regret when the global optimum is known and the best-so-far objective value otherwise, averaged over 10 random seeds on a logarithmic scale, with shaded regions indicating the 95% confidence interval.

4 Experimental Setup

4.1 Acquisition Optimization Setup

We adopt a standard BO framework with a Gaussian Process (GP) surrogate model. At each iteration, the next query point is selected by optimizing an acquisition function. In our experiments, we use EI as the acquisition function; its mathematical formulation is provided in Appendix A.2. All methods share the same surrogate model and acquisition function; the only difference lies in how the acquisition function is optimized.

We compare the proposed KSOS-BO optimizer with four representative acquisition optimizers. Sobol Search performs sobol sampling over the domain and selects the candidate with the lowest acquisition objective value, serving as a simple baseline. CMA-ES and DE are population-based optimizers that iteratively refine candidate solutions by stochastic search. Limited-memory Broyden–Fletcher–Goldfarb–Shanno algorithm with Bound constraints (L-BFGS-B) is a widely used gradient-based local optimizer for box-constrained optimization. It provides a strong local-refinement baseline and allows us to evaluate whether KSOS-BO offers benefits beyond standard gradient-based optimization. Details on the hyperparameter choices for all optimizers are provided in Appendix B.1.

4.2 Evaluation Protocol

We follow a standard BO setup with a GP surrogate using a Matérn kernel ($\nu = 2.5$) with additive white noise. Each run is initialized with $N_{\text{init}} = 12$ points and performed for 400 BO iterations over 10 random seeds. We use EI with $\xi = 0.01$ as the acquisition function.

For benchmarks with a known global minimum $f^\star$, we report simple regret,

$$r_t = \min_{i \le t} f(x_i) - f^\star. \tag{7}$$

For benchmarks without a known global minimum, we instead report the best-so-far objective value $\min_{i \le t} f(x_i)$. In both cases, evaluation is based on the noise-free objective values $f(x_i)$, while the GP surrogate is fitted to noisy observations $y_i = f(x_i) + \epsilon_i$. We report the mean performance

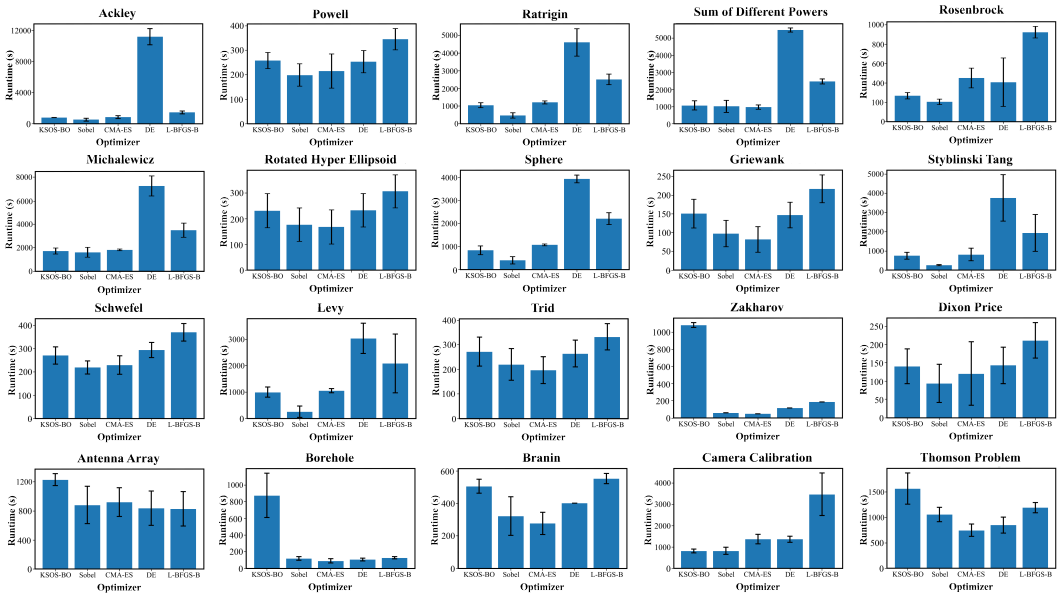


Figure 3: Average runtime over 400 iterations for five acquisition optimizers across 20 benchmark functions. Each bar shows the mean runtime over 10 random seeds, with error bars indicating one standard deviation.

with 95% confidence intervals over the 10 random seeds. We evaluate all methods on 15 synthetic and 5 engineering benchmarks. Sobol sampling is used for both Sobol Search and KSOS-BO. KSOS-BO uses an anisotropic Gaussian kernel whose bandwidths are determined from the GP Automatic Relevance Determination (ARD) length-scales. Benchmark definitions and detailed optimizer configurations are provided in Appendices A.6 and B.1, respectively.

5 Results

5.1 Overall Optimization Performance

A quantitative summary is provided in Table 1, and the convergence results are shown in Figure 2. The reported improvement is computed based on the final aggregated regret. Specifically, we define

$$\text{Improvement}(\%) = \frac{r_{\text{ref}} - r_K}{|r_{\text{ref}}|} \times 100. \tag{8}$$

r_K denotes the final regret of KSOS-BO, and the reference value r_{ref} is defined as the final regret of the second-best optimizer when KSOS-BO ranks first, and as the final regret of the best optimizer otherwise. Consequently, a positive value indicates that KSOS-BO outperforms the strongest competing method, while a negative value indicates underperformance relative to the best optimizer.

In general, KSOS-BO achieves the best performance on 13 of 20 benchmarks, consistently ranking first and yielding substantial improvements over competing methods. In particular, KSOS-BO performs strongly on multimodal functions such as Ackley, Rastrigin, Levy, and Griewank, as well as on bowl-shaped functions including Sum of Different Powers, Rotated Hyper Ellipsoid, and Trid. Moreover, KSOS-BO also performs well on heterogeneous engineering benchmarks such as Borehole and Camera Calibration. These results indicate that KSOS-BO effectively solves both highly multimodal landscapes and smooth unimodal objectives. On smooth but curved functions such as Rosenbrock and Dixon Price, KSOS-BO attains high-quality solutions early in the optimization process with negligible variance and subsequently maintains stable performance in the identified low-regret region. In contrast, KSOS-BO shows weaker performance on functions with steep drops or plate-shaped structures, such as Styblinski Tang, Michalewicz and Zakharov. These cases suggest that KSOS-BO may struggle when the objective exhibits extreme curvature or overly flat regions,

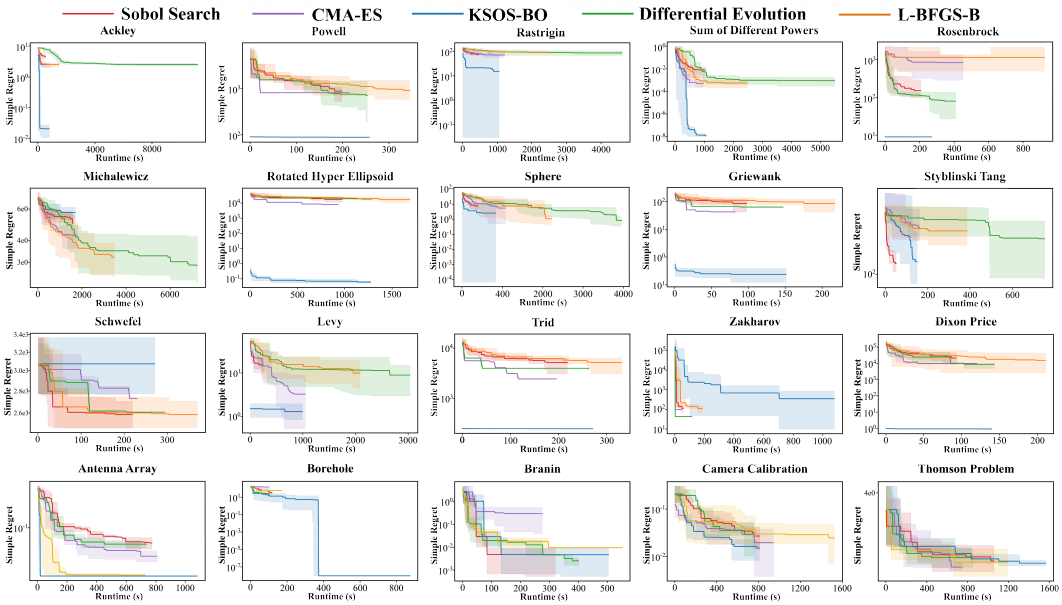


Figure 4: Convergence comparison on 20 benchmarks using EI over wall-clock time. The curves show simple regret when the global optimum is known and the best-so-far objective value otherwise, averaged over 10 random seeds on a logarithmic scale, with shaded regions indicating the 95% confidence interval.

where the kernel-based representation becomes less adequate. Overall, the results demonstrate that KSOS-BO is effective in identifying high-quality solutions for acquisition optimization.

5.2 Runtime and Computational Efficiency

We evaluate the wall-clock cost of the acquisition optimizers under the configurations specified in Appendix B.1. These configurations provide strong implementations of each baseline but do not enforce an equal number of acquisition-function evaluations across optimizers. KSOS-BO incurs matrix-based computation through the KernelSOS optimization step, while population-based and multi-start methods incur costs through repeated acquisition evaluations and local refinement. As shown in Figure 3, runtime varies across both optimizers and benchmark functions. KSOS-BO remains within the same order of wall-clock runtime as the competing methods on 19 out of 20 benchmarks. Its cost is more pronounced on highly multimodal functions such as Ackley and Rastrigin, whereas the runtime gap is smaller on smoother functions such as Powell, Rosenbrock, and Sum of Different Powers. Together with the optimization results in Figure 2, these results show that the additional computation introduced by KernelSOS is accompanied by substantial gains in final optimization quality on many benchmarks, yielding a favorable end-to-end trade-off between computational cost and BO performance.

5.3 Convergence Performance over Time

We further evaluate optimization progress as a function of wall-clock time. Figure 4 shows simple regret versus cumulative runtime across 20 benchmark functions, averaged over 10 random seeds and 400 BO iterations. KSOS-BO reaches low-regret regions earlier on many of the benchmarks where it achieves the best final performance. This behavior is particularly clear on Powell, Rosenbrock, Rotated Hyper Ellipsoid, Griewank, Levy, Trid, and Dixon Price, where KSOS-BO rapidly attains low regret and maintains its advantage throughout optimization. Similar behavior is observed on the multimodal Ackley and Rastrigin functions, where KSOS-BO reaches lower-regret regions earlier despite the added cost of the KernelSOS step. Overall, these results show that improved acquisition optimization with KSOS-BO translates into faster progress toward high-quality solutions in wall-

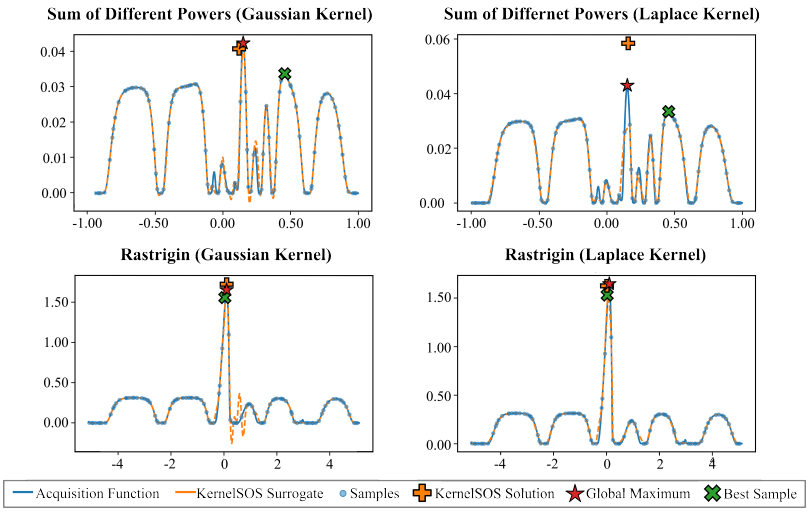


Figure 5: Visualization of KernelSOS surrogate approximation on 1D EI after 9 BO steps.

clock time across many benchmarks, demonstrating that the additional inner-loop computation can be effectively converted into improved BO convergence.

5.4 KernelSOS Surrogate Approximation of EI

To understand the behavior of KSOS-BO, we examine how its surrogate approximates the true EI acquisition landscape in 1D during BO. Figure 5 shows the acquisition functions after 9 BO steps, together with the KernelSOS surrogate fitted from 128 Sobol samples using Gaussian and Laplace kernels. With both Gaussian and Laplace kernels, the surrogate closely matches the true acquisition curves on Sum of Different Powers and Rastrigin. The overall shape of the acquisition landscape, including multiple local peaks and the dominant global maximum, is accurately captured, leading to strong alignment across the domain. Even for the highly localized and peaked structure of EI, the surrogate successfully identifies the key maxima, with only minor discrepancies in relatively flat regions. Overall, these results show that KSOS-BO can approximate complex acquisition landscapes from a limited number of evaluations. The Gaussian kernel produces smoother and more globally consistent approximations, while the Laplace kernel provides higher local resolution and better captures sharper variations. Despite these differences, both kernels reliably identify the optimal query, indicating that KSOS-BO is robust to the choice of kernel.

6 Conclusion

We presented KSOS-BO, a BO framework that uses KernelSOS as a structured kernel-based method for acquisition optimization. Across diverse benchmarks, KSOS-BO achieves the best final regret on 13 out of 20 tasks compared with Sobol Search, DE, CMA-ES, and multi-start L-BFGS-B. Wall-clock experiments further show that these optimization gains can be achieved with competitive end-to-end computational cost under the evaluated configurations. These results demonstrate the potential of structured acquisition optimization to improve overall BO performance. Future work will focus on improving the efficiency and scalability of KSOS-BO, including more efficient optimization formulations and the automatic determination of appropriate kernels. Another important direction is to study hybrid strategies that use KSOS-BO for global candidate generation followed by local refinement, and testing KSOS-BO in advanced BO pipelines, including non-differentiable acquisition functions, constrained, and multi-objective settings.

References

Yanki Aslan, Jan Puskely, Antoine Roederer, and Alexander Yarovoy. Phase-only control of peak sidelobe level and pattern nulls using iterative phase perturbations. *IEEE Antennas and Wireless*

Propagation Letters, 18(10):2081–2085, 2019.

James Bergstra, Rémi Bardenet, Yoshua Bengio, and Balázs Kégl. Algorithms for hyper-parameter optimization. *Advances in Neural Information Processing Systems*, 24, 2011.

Eloïse Berthier, Justin Carpentier, Alessandro Rudi, and Francis Bach. Infinite-dimensional sums-of-squares for optimal control. In *2022 IEEE 61st Conference on Decision and Control (CDC)*, pp. 577–582. IEEE, 2022.

E Brochu, V Cora, and N de Freitas. A Tutorial on Bayesian Optimization of Expensive Cost Functions ,with Application to Active User Modeling and Hierarchical Reinforcement Learning. 2009.

Swagatam Das and Ponnuthurai Nagaratnam Suganthan. Differential Evolution: A survey of the state-of-the-art. *IEEE Transactions on Evolutionary Computation*, 15(1):4–31, 2010.

Katharina Eggensperger, Frank Hutter, Holger Hoos, and Kevin Leyton-Brown. Efficient benchmarking of hyperparameter optimizers via surrogates. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 29, 2015.

T Erber and GM Hockney. Equilibrium configurations of N equal charges on a sphere. *Journal of Physics A: Mathematical and General*, 24(23):L1369–L1377, 1991.

Peter I Frazier. Bayesian optimization. In *Recent advances in optimization and modeling of contemporary problems*, pp. 255–278. Informs, 2018.

Weiao Gan, Ziyuan Ji, and Yongqing Liang. Acquisition functions in Bayesian optimization. In *2021 2nd International Conference on Big Data & Artificial Intelligence & Software Engineering (ICBASE)*, pp. 129–135. IEEE, 2021.

Antoine Groudiev and Frederike Dümbgen. ksos-tools: Implementation of the Kernel Sum-of-Squares optimization framework and various related solvers. URL `https://github.com/Simple-Robotics/ksos-tools`.

Antoine Groudiev, Fabian Schramm, Éloïse Berthier, Justin Carpentier, and Frederike Dümbgen. Sampling-based global optimal control and estimation via semidefinite programming. In *2026 American Control Conference (ACC)*, pp. 3882–3889. IEEE, 2026.

Nikolaus Hansen. The CMA evolution strategy: A tutorial. *arXiv preprint arXiv:1604.00772*, 2016.

Donald R Jones, Matthias Schonlau, and William J Welch. Efficient global optimization of expensive black-box functions. *Journal of Global optimization*, 13(4):455–492, 1998.

Jean B Lasserre. Global optimization with polynomials and the problem of moments. *SIAM Journal on Optimization*, 11(3):796–817, 2001.

Jean Bernard Lasserre. *Moments, positive polynomials and their applications*, volume 1. World Scientific, 2009.

Jean Bernard Lasserre. *An introduction to polynomial and semi-algebraic optimization*, volume 52. Cambridge University Press, 2015.

Ulysse Marteau-Ferey, Francis Bach, and Alessandro Rudi. Non-parametric models for non-negative functions. *Advances in Neural Information Processing Systems*, 33:12816–12826, 2020.

Jorge Nocedal and Stephen J Wright. *Numerical optimization*. Springer, 2006.

Pablo A Parrilo. *Structured Semidefinite Programs and Semialgebraic Geometry Methods in Robustness and Optimization*. PhD thesis, California Institute of Technology, 2000.

Pablo A Parrilo. Semidefinite programming relaxations for semialgebraic problems. *Mathematical Programming*, 96(2):293–320, 2003.

Vern I Paulsen and Mrinal Raghupathi. *An introduction to the theory of reproducing kernel Hilbert spaces*, volume 152. Cambridge University Press, 2016.

Alessandro Rudi, Ulysse Marteau-Ferey, and Francis Bach. Finding global minima via kernel approximations: A. Rudi et al. *Mathematical Programming*, 209(1):703–784, 2025.

Bernhard Scholkopf and Alexander J Smola. *Learning with kernels: support vector machines, regularization, optimization, and beyond*. MIT press, 2018.

Matthias Seeger. Gaussian processes for machine learning. *International Journal of Neural Systems*, 14(02):69–106, 2004.

Bobak Shahriari, Kevin Swersky, Ziyu Wang, Ryan P. Adams, and Nando de Freitas. Taking the human out of the loop: A review of bayesian optimization. *Proceedings of the IEEE*, 104(1): 148–175, 2016. doi: 10.1109/JPROC.2015.2494218.

Jasper Snoek, Hugo Larochelle, and Ryan P Adams. Practical bayesian optimization of machine learning algorithms. *Advances in Neural Information Processing Systems*, 25, 2012.

Jasper Snoek, Oren Rippel, Kevin Swersky, Ryan Kiros, Nadathur Satish, Narayanan Sundaram, Mostofa Patwary, Mr Prabhat, and Ryan Adams. Scalable bayesian optimization using deep neural networks. In *International Conference on Machine Learning*, pp. 2171–2180. PMLR, 2015.

S. Surjanovic and D. Bingham. Virtual Library of Simulation Experiments: Test Functions and Datasets. Retrieved May 6, 2026, from `http://www.sfu.ca/~ssurjano`.

Russ Tedrake, Ian R Manchester, Mark Tobenkin, and John W Roberts. LQR-trees: Feedback motion planning via sums-of-squares verification. *The International Journal of Robotics Research*, 29(8): 1038–1052, 2010.

Hao Wang, Bas van Stein, Michael Emmerich, and Thomas Back. A new acquisition function for Bayesian optimization based on the moment-generating function. In *2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pp. 507–512. IEEE, 2017.

Zhongqi Wei and Frederike Dümbgen. Global sampling-based trajectory optimization for contact-rich manipulation via kernelsos. *arXiv preprint arXiv:2604.27175*, 2026.

Christopher Williams and Carl Rasmussen. Gaussian processes for regression. *Advances in Neural Information Processing Systems*, 8, 1995.

James Wilson, Frank Hutter, and Marc Deisenroth. Maximizing acquisition functions for Bayesian optimization. *Advances in Neural Information Processing Systems*, 31, 2018.

Heng Yang and Luca Carlone. Certifiably optimal outlier-robust geometric perception: Semidefinite relaxations and scalable global optimization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(3):2816–2834, 2022.

Zhengyou Zhang. A flexible new technique for camera calibration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(11):1330–1334, 2000.

A THEORETICAL BACKGROUND

A.1 GAUSSIAN PROCESS

A GP is a collection of random variables $\{f(x)\}_{x\in\mathcal{X}}$ such that any finite subset follows a joint Gaussian distribution. A GP is fully specified by a mean function $m(x)$ and a covariance function $k(x, x')$, and is denoted as (Williams & Rasmussen, 1995)

$$f(x) \sim \mathcal{GP}\big(m(x), k(x, x')\big). \tag{9}$$

For any inputs $x_1, \ldots, x_n$, the corresponding function values satisfy

$$\big(f(x_1), \ldots, f(x_n)\big)^\top \sim \mathcal{N}\big(m_X, K\big), \tag{10}$$

where $m_X = \big(m(x_1), \ldots, m(x_n)\big)^\top$ and $K_{ij} = k(x_i, x_j)$. Given observations $y = (y_1, \ldots, y_n)^\top$ at inputs $X = \{x_i\}_{i=1}^n$, with $y_i = f(x_i) + \epsilon_i$ and $\epsilon_i \sim \mathcal{N}(0, \sigma_n^2)$, the posterior distribution at a test point x_* remains Gaussian with mean and variance

$$\mu(x_*) = m(x_*) + k_*^\top (K + \sigma_n^2 I)^{-1}(y - m_X), \quad \sigma^2(x_*) = k(x_*, x_*) - k_*^\top (K + \sigma_n^2 I)^{-1} k_*, \tag{11}$$

where $k_* = \big(k(x_*, x_1), \ldots, k(x_*, x_n)\big)^\top$. The covariance function $k(x, x')$ is symmetric and positive semidefinite, ensuring that the kernel matrix K is positive semidefinite for any finite set of inputs.

A.2 ACQUISITION FUNCTIONS

We consider three commonly used acquisition functions in BO: EI, Probability of Improvement (PI), and Lower Bound Confidence (LCB).

For a minimization problem, EI is defined as

$$\mathrm{EI}(x) = (f_{\text{best}} - \mu(x) - \xi)\,\Phi\left(\frac{f_{\text{best}} - \mu(x) - \xi}{\sigma(x)}\right) + \sigma(x)\,\phi\left(\frac{f_{\text{best}} - \mu(x) - \xi}{\sigma(x)}\right), \tag{12}$$

where $\mu(x)$ and $\sigma(x)$ denote the posterior mean and standard deviation of the surrogate model, f_{best} is the best observed value, $\Phi(\cdot)$ and $\phi(\cdot)$ are the standard normal CDF and PDF, and $\xi \geq 0$ is the improvement threshold. Since EI is conventionally maximized, under our minimization convention we use

$$\alpha_{\mathrm{EI}}(x) = -\mathrm{EI}(x). \tag{13}$$

Similarly, PI is defined as

$$\mathrm{PI}(x) = \Phi\left(\frac{f_{\text{best}} - \mu(x) - \xi}{\sigma(x)}\right), \tag{14}$$

and under our minimization convention we use

$$\alpha_{\mathrm{PI}}(x) = -\mathrm{PI}(x). \tag{15}$$

LCB is directly minimized and is defined as

$$\alpha_{\mathrm{LCB}}(x) = \mu(x) - \beta\sigma(x), \tag{16}$$

where $\beta > 0$ controls the trade-off between exploration and exploitation.

Kernel alignment between the GP posterior and KSOS. The effectiveness of KSOS for Acquisition Function Optimization (AFO) is motivated by the structural alignment between the acquisition landscape induced by the GP posterior and the kernel representation used by KSOS.

In BO, the acquisition function is determined by the posterior mean and variance of the GP. Given observations $\mathcal{D}_t = \{(x_i, y_i)\}_{i=1}^t$, the posterior quantities $\mu_t(x)$ and $\sigma_t^2(x)$ are given by equation 11. With the Matérn-$5/2$ kernel used in our GP surrogate, these quantities vary smoothly with respect to x. Common acquisition functions are deterministic transformations of the posterior mean and

variance. For example, EI is given by equation 12, and its minimized acquisition objective is $-\mathrm{EI}(x)$. The resulting acquisition landscape therefore inherits the regularity of the GP posterior.

KSOS exploits this structure by constructing a kernel-induced representation of the acquisition objective. Given acquisition evaluations $\{(u_i, \alpha_t(u_i))\}_{i=1}^n$, we define

$$\varphi(x) = [\kappa(x, u_1), \ldots, \kappa(x, u_n)]^\top,$$

and represent the shifted acquisition objective as

$$\hat{\alpha}_t(x) = c + \varphi(x)^\top B \varphi(x), \qquad B \succeq 0, \tag{17}$$

where κ denotes the KSOS kernel. In KSOS-BO, we use an anisotropic Gaussian kernel whose dimension-wise bandwidths are determined from the ARD length-scales learned by the GP. This transfers the smoothness structure inferred by the surrogate to the KSOS representation, allowing KSOS to exploit the geometry of the acquisition landscape beyond pointwise candidate selection or local refinement.

A.3 Automatic Relevance Determination

ARD allows a GP kernel to learn a separate length-scale for each input dimension. Let $x, x' \in \mathbb{R}^d$ and let

$$\boldsymbol{\ell} = (\ell_1, \ell_2, \ldots, \ell_d)^\top \tag{18}$$

denote the ARD length-scale vector. Instead of measuring distance isotropically, ARD uses the scaled distance

$$r_{\mathrm{ARD}}(x, x') = \left(\sum_{j=1}^{d} \frac{(x_j - x'_j)^2}{\ell_j^2} \right)^{1/2}. \tag{19}$$

For the Matern kernel used in our GP surrogate, the ARD kernel is

$$k_{\mathrm{Matern}^{\mathrm{ARD}}}(x, x') = \sigma_f^2 \left(1 + \sqrt{5} r_{\mathrm{ARD}}(x, x') + \frac{5}{3} r_{\mathrm{ARD}}^2(x, x') \right) \exp\left(-\sqrt{5} r_{\mathrm{ARD}}(x, x') \right). \tag{20}$$

A smaller ℓ_j indicates that the function varies more rapidly along dimension j, while a larger ℓ_j indicates smoother variation along that dimension.

In KSOS-BO, we use the ARD length-scales learned by the GP to set the Gaussian kernel bandwidth in KSOS. After fitting the GP at BO iteration t, we extract

$$\boldsymbol{\ell}_t = (\ell_{t,1}, \ell_{t,2}, \ldots, \ell_{t,d})^\top \tag{21}$$

and define the KSOS bandwidth vector as

$$\boldsymbol{\sigma}_t^{\mathrm{KSOS}} = \rho \boldsymbol{\ell}_t, \tag{22}$$

where $\rho > 0$ is a scaling factor. This makes the KSOS kernel anisotropic and aligns its smoothness with the dimension-wise structure learned by the GP surrogate.

A.4 Sum-of-Squares Relaxations for Polynomial Optimization

We briefly review the SOS framework for unconstrained global polynomial optimization, which forms the foundation of the KernelSOS extension used in this work (Parrilo, 2000).

Global optimization via nonnegativity. Consider an unconstrained polynomial optimization problem

$$f^* = \min_{x \in \mathbb{R}^d} f(x), \tag{23}$$

where $f : \mathbb{R}^d \to \mathbb{R}$ is a polynomial. This problem can be equivalently reformulated as the search for the largest global lower bound:

$$c^* = \max_{c \in \mathbb{R}} \quad \text{s.t.} \quad f(x) - c \geq 0, \quad \forall x \in \mathbb{R}^d. \tag{24}$$

The constraint requires the shifted polynomial $f(x) - c$ to be globally nonnegative. However, verifying global nonnegativity of a multivariate polynomial is generally intractable.

Sum-of-squares relaxation. A sufficient condition for polynomial nonnegativity is that the polynomial admits a SOS decomposition. Specifically, a polynomial $p(x)$ is SOS if there exist polynomials $\{q_i(x)\}_{i=1}^m$ such that

$$p(x) = \sum_{i=1}^{m} q_i^2(x). \tag{25}$$

This guarantees $p(x) \geq 0$ for all $x \in \mathbb{R}^d$ and leads to the SOS relaxation

$$c^*_{\mathrm{SOS}} = \max_{c \in \mathbb{R}} \quad \text{s.t.} \quad f(x) - c \text{ is SOS}. \tag{26}$$

Since the SOS condition is sufficient but not necessary for polynomial nonnegativity, the relaxation provides a lower bound:

$$c^*_{\mathrm{SOS}} \leq f^*. \tag{27}$$

Semidefinite representation. A key property of SOS polynomials is that they admit a semidefinite representation. Let $v_d(x)$ denote the vector of monomials up to degree d. A polynomial $p(x)$ of degree at most $2d$ is SOS if and only if there exists a positive semidefinite matrix $Q \succeq 0$ such that

$$p(x) = v_d(x)^\top Q v_d(x). \tag{28}$$

This converts the SOS constraint into a linear matrix inequality, allowing the relaxation to be solved using semidefinite programming.

This polynomial SOS formulation provides the foundation for the kernelized extension presented in the next subsection, where the finite-dimensional monomial representation is replaced by functions in an RKHS.

A.5 Kernel Sum-of-Squares in Reproducing Kernel Hilbert Spaces

The polynomial SOS formulation can be extended to an infinite-dimensional setting by considering functions in a reproducing kernel Hilbert space (RKHS) (Marteau-Ferey et al., 2020). This extension replaces explicit polynomial bases with kernel-induced feature maps, allowing nonnegative functions to be represented through positive operators in an RKHS.

RKHS and function representation. Let $\mathcal{H}$ be an RKHS associated with a positive definite kernel $k(\cdot, \cdot)$. By the reproducing property, any function $g \in \mathcal{H}$ satisfies

$$g(x) = \langle g, k(x, \cdot) \rangle_{\mathcal{H}}. \tag{29}$$

Given a finite set of sample points $\{x_i\}_{i=1}^N$, the representer theorem (Rudi et al., 2025) implies that solutions to regularized optimization problems over $\mathcal{H}$ can be expressed as finite kernel expansions of the form

$$g(x) = \sum_{i=1}^{N} \alpha_i k(x, x_i), \tag{30}$$

for some coefficients $\alpha \in \mathbb{R}^N$.

Kernelized sum-of-squares representation. In classical SOS optimization, a nonnegative function is represented as a sum of squared functions. KernelSOS generalizes this idea by replacing finite-dimensional polynomial bases with functions in an RKHS. Let $\phi(x)$ denote the feature map associated with the kernel k, such that

$$k(x, x') = \langle \phi(x), \phi(x') \rangle_{\mathcal{H}}. \tag{31}$$

A function $s(x)$ admits a KernelSOS representation if it can be written as

$$s(x) = \langle \phi(x), B\phi(x) \rangle_{\mathcal{H}}, \tag{32}$$

where $B \succeq 0$ is a positive semidefinite operator acting on the feature space. In the infinite-dimensional formulation, this construction ensures

$$s(x) \geq 0, \quad \forall x \in \mathcal{X}. \tag{33}$$

Thus, if a function $f(x) - c$ admits such a representation, then c is a valid lower bound of f over the domain $\mathcal{X}$.

Infinite-dimensional lower-bound formulation. Using this representation, the KernelSOS relaxation of a minimization problem can be written as

$$\begin{aligned} \max_{c\in\mathbb{R},\, B\succeq 0} \quad & c - \lambda \operatorname{Tr}(B) \\ \text{s.t.} \quad & f(x) - c = \langle \phi(x), B\phi(x)\rangle_{\mathcal{H}}, \quad \forall x \in \mathcal{X}, \end{aligned} \tag{34}$$

where $\lambda \geq 0$ is a regularization parameter. Since $B \succeq 0$, the right-hand side is nonnegative for every $x \in \mathcal{X}$, and therefore the constraint implies

$$f(x) - c \geq 0, \quad \forall x \in \mathcal{X}. \tag{35}$$

The infinite-dimensional program therefore has a lower-bound interpretation: the scalar c is optimized subject to being no larger than $f(x)$ over the entire domain. Under the assumptions studied in the original KernelSOS analysis, including sufficient smoothness of f and compatibility with the chosen RKHS, this formulation provides a principled relaxation for global minimization (Marteau-Ferey et al., 2020).

Connection to finite-dimensional approximation. The constraint over all $x \in \mathcal{X}$ is generally intractable. Following the original KernelSOS formulation, we approximate it using a finite set of sampled points $\{x_i\}_{i=1}^N$. Evaluating the feature map at these points and applying the representer theorem yields a finite-dimensional semidefinite program of the form used in the main text.

This finite-dimensional problem should be interpreted as a sampled approximation of the infinite-dimensional KernelSOS program. In particular, the practical formulation enforces the KernelSOS equality constraints only at the sampled points,

$$f(x_i) - c = \phi_i^\top B \phi_i, \quad i = 1, \ldots, N, \tag{36}$$

where $\phi_i = [k(x_i, x_1), \ldots, k(x_i, x_N)]^\top$. Therefore, the finite formulation does not by itself certify exact global nonnegativity over the entire domain. Instead, it provides a kernel-based sampled relaxation whose quality depends on the kernel, the sample distribution, the number of sampled points, and the regularization parameter.

The original KernelSOS analysis shows that, for sufficiently smooth functions and suitable kernels, the finite sampled formulation approximates the infinite-dimensional lower-bound problem, with the approximation quality controlled by the coverage of the sampled points and the smoothness of the function. In our setting, the function being optimized by KernelSOS is the acquisition objective. Common acquisition functions induced by Gaussian process surrogates, such as EI and LCB, are typically smooth functions when the posterior mean and variance are smooth. This makes the RKHS-based approximation assumption reasonable in practice. Nevertheless, our finite-dimensional implementation should be understood as a structured sampled approximation for AFO, rather than as a formal certificate of the exact global optimum of the acquisition function.

A.6 Benchmark Functions

We evaluate all optimization methods on a diverse set of 15 synthetic benchmark functions covering a range of landscape characteristics, together with 5 engineering benchmarks.

The implementations of the 15 synthetic functions, as well as the Borehole and Branin functions, are based on (Surjanovic & Bingham), with the default bounds specified for each benchmark. Detailed equations and benchmark definitions are available at: `https://www.sfu.ca/~ssurjano/index.html`. The Antenna Array, Camera Calibration, and Thomson Problem benchmarks are implemented based on (Aslan et al., 2019), (Zhang, 2000), and (Erber & Hockney, 1991), respectively.

Antenna Array

$$AF(\theta) = \sum_{j=0}^{d-1} \exp\left(\mathrm{i}\left(ksj\sin(\theta) + \phi_j\right)\right) \tag{37}$$

$$P(\theta) = \left(\frac{|AF(\theta)|}{d}\right)^2 \tag{38}$$

$$f(\phi) = \frac{1}{\alpha} \log \left(\frac{1}{|\Theta_{\text{side}}|} \sum_{\theta \in \Theta_{\text{side}}} \exp(\alpha P(\theta)) \right) + \frac{1}{2}(1 - P(\theta_0))^2 + 10^{-4} \frac{1}{d} \sum_{j=0}^{d-1} \phi_j^2 \tag{39}$$

Domain:

$$\phi_j \in [-\pi, \pi], \quad j = 0, \ldots, d-1.$$

Camera Calibration We consider an eight-dimensional camera calibration problem with parameter vector

$$x = (r_x, r_y, r_z, t_x, t_y, t_z, f, k_1),$$

where (r_x, r_y, r_z) is a Rodrigues rotation vector, (t_x, t_y, t_z) is the translation vector, f is the focal length, and k_1 is the radial distortion coefficient.

For each three-dimensional point P_i, its camera-frame coordinates are

$$q_i(x) = R(r_x, r_y, r_z)P_i + t, \tag{40}$$

where $R(\cdot)$ is the rotation matrix obtained through Rodrigues' formula. The normalized image coordinates are

$$\bar{x}_i = \frac{q_{i,1}}{\max(q_{i,3}, 10^{-6})}, \qquad \bar{y}_i = \frac{q_{i,2}}{\max(q_{i,3}, 10^{-6})}. \tag{41}$$

Using a single radial distortion coefficient, the predicted image location is

$$u_i(x) = f\left(1 + k_1 \rho_i^2\right) \begin{bmatrix} \bar{x}_i \\ \bar{y}_i \end{bmatrix}, \qquad \rho_i^2 = \bar{x}_i^2 + \bar{y}_i^2. \tag{42}$$

The objective is the mean squared normalized reprojection error:

$$\mathcal{J}(x) = \frac{1}{2N} \sum_{i=1}^{N} \left\| \frac{u_i(x) - u_i^{\text{obs}}}{100} \right\|_2^2. \tag{43}$$

The observations u_i^{obs} are generated once from a fixed reference camera configuration with fixed perturbations, making the benchmark deterministic. The parameter domain is

$$x^{\min} = [-0.6 \quad -0.6 \quad -0.6 \quad -1.0 \quad -1.0 \quad 2.0 \quad 400 \quad -0.3]^\top, \tag{44}$$

$$x^{\max} = [0.6 \quad 0.6 \quad 0.6 \quad 1.0 \quad 1.0 \quad 6.0 \quad 1200 \quad 0.3]^\top, \tag{45}$$

with $x_j \in [x_j^{\min}, x_j^{\max}], \qquad j = 1, \ldots, 8.$

Thomson Problem

$$p_i = \begin{bmatrix} \sin(\theta_i)\cos(\phi_i) \\ \sin(\theta_i)\sin(\phi_i) \\ \cos(\theta_i) \end{bmatrix} \tag{46}$$

$$f(\theta, \phi) = \sum_{1 \le i < j \le 4} \frac{1}{\|p_i - p_j\|_2} \tag{47}$$

Domain:

$$\theta_i \in [0, \pi], \quad \phi_i \in [-\pi, \pi], \quad i = 1, \ldots, 4.$$

Global minimum:

$$f(x^\star) \approx 3.6742.$$

Table 2: Experimental setup for KSOS-BO and baselines. All methods share the same BO surrogate model, acquisition function, and search domain.

COMPONENT	CONFIGURATION
Objective Functions	15 synthetic and 5 engineering benchmarks
Dimension	Synthetic: $d = 10$ except Powell ($d = 12$); Application-inspired: $d = 8$ except Branin ($d = 2$)
BO Iterations	$N_{\text{init}} = 12$; $N_{\text{iters}} = 400$
Random Seeds	10
Gaussian Process	Constant $\times$ Matérn-ARD ($\nu = 2.5$) + WhiteKernel
GP Length-scale Bounds	$\ell_j \in [10^{-2}, 10^{2}], j = 1, \ldots, d$
GP Output Normalization	Enabled (`normalize_y=True`)
GP Optimizer Restarts	2
Observation Noise	$\sigma_{\text{n}} = 0.05 \cdot \text{std}(y)$
Acquisition Function	Expected Improvement (EI), $\xi = 0.01$
Sobol Search	Evaluation samples $= 128$
CMA-ES	$x_0 = 0.5(x_{\min}+x_{\max})$; $\sigma_0 = 0.25 \cdot \text{mean}(x_{\max}-x_{\min})$; population size $= 4 + \lfloor 3 \log d \rfloor$; maximum evaluations $= 10000d$
Differential Evolution	Strategy: best1bin; SciPy population multiplier `popsize` $= 15$; Sobol initialization, giving a realized population size of $2^{\lceil \log_2(15d) \rceil}$; maximum iterations $= 1000$; tolerance $= 10^{-7}$; mutation $= (0.5, 1.0)$; recombination $= 0.7$; polish enabled
L-BFGS-B	Raw samples $= 1024$; restarts $= 20$; maximum iterations per restart $= 500$; $f_{\text{tol}} = 10^{-12}$; $g_{\text{tol}} = 10^{-8}$; maximum line-search steps $= 50$
KSOS-BO Sampling	Sobol sampling; $N = 128$
KSOS-BO Kernel	Anisotropic Gaussian kernel
KSOS-BO Bandwidth	$\boldsymbol{\sigma}_t^{\text{KSOS}} = \rho \boldsymbol{\ell}_t$, $\rho = 1$
KSOS-BO Regularization	$\lambda = 10^{-3}$
KSOS-BO Barrier	$\epsilon = 10^{-3}$
KSOS-BO Solver	Damped Newton; maximum iterations $= 100$; warm iterations $= 1$
KSOS-BO Search Region	$\boldsymbol{c} = 0.5(\boldsymbol{x}_{\min} + \boldsymbol{x}_{\max})$; $\boldsymbol{r} = 0.5(\boldsymbol{x}_{\max} - \boldsymbol{x}_{\min})$

B Experimental Details

B.1 Experimental Setup Details

Here we provide the implementation details of KSOS-BO and baselines in Table 2. N_{init} denotes the number of initial points evaluated before BO begins, while N_{iters} denotes the number of subsequent BO iterations used for sequential acquisition and evaluation. σ_{n} denotes the standard deviation of the observation noise. For acquisition functions, ξ is the improvement threshold in EI. In CMA-ES, x_0 denotes the initial sampling center and σ_0 denotes the initial sampling step size.

In L-BFGS-B, the tolerance parameters f_{tol} and g_{tol} define the stopping criteria based on objective-value improvement and gradient norm, respectively, while the maximum number of line-search steps limits the effort spent on choosing a step size within each quasi-Newton update.

In KSOS-BO, the Gaussian kernel bandwidth is adapted from the ARD length-scales learned by the GP surrogate, as described in Appendix A.3. By transferring the dimension-wise smoothness estimated by the GP to the kernel in KSOS, this ARD-based bandwidth selection better aligns the KSOS approximation with the geometry of the acquisition landscape and reduces the need for manual bandwidth tuning.

B.2 Hardware and Operating System

All experiments are conducted on a machine equipped with an AMD Ryzen 7 5800H CPU (16 threads) and an NVIDIA GeForce RTX 3060 GPU. The system runs Ubuntu 20.04.6 LTS with Linux kernel version 5.15.0-139-generic on an x86_64 architecture. All experiments are implemented in

Table 3: **Ablation study on different acquisition functions.** Final regret after 400 BO iterations on five 10-dimensional benchmark functions. Results are reported as mean ± standard deviation over 10 seeds. Lower is better.

Acquisition	Method	Ackley	Rastrigin	Rosenbrock	Michalewicz	Sum of Different Powers
PI ($\xi = 0.01$)	KSOS-BO	**0.052 ± 0.041**	**21.36 ± 16.32**	**9.01 ± 0.10**	5.33 ± 0.17	**0.00066 ± 0.00028**
	Sobol Search	4.82 ± 0.74	91.47 ± 6.69	109.04 ± 6.28	4.88 ± 0.19	0.0069 ± 0.0055
	DE	2.58 ± 0.21	50.25 ± 20.13	73.76 ± 31.28	4.16 ± 0.91	0.0011 ± 0.00092
	CMA-ES	2.55 ± 0.27	33.57 ± 7.03	56.13 ± 29.86	3.59 ± 0.68	0.0010 ± 0.00030
	L-BFGS-B	2.10 ± 0.15	46.73 ± 12.08	68.49 ± 43.24	**3.46 ± 1.03**	0.0012 ± 0.0011
LCB ($\beta = 2$)	KSOS-BO	**5.02 ± 3.22**	**53.16 ± 8.80**	**10.46 ± 0.57**	6.12 ± 0.37	**0.000094 ± 0.000057**
	Sobol Search	7.32 ± 1.10	72.99 ± 8.46	168.88 ± 38.77	7.28 ± 0.71	0.0051 ± 0.0039
	DE	7.31 ± 0.76	57.07 ± 9.14	126.03 ± 16.78	4.10 ± 0.88	0.00021 ± 0.00011
	CMA-ES	7.99 ± 0.34	59.20 ± 19.46	102.01 ± 24.09	4.30 ± 0.51	0.00029 ± 0.00014
	L-BFGS-B	6.52 ± 3.17	56.28 ± 8.28	97.13 ± 11.28	**3.57 ± 0.64**	0.00014 ± 0.000094

Table 4: **Ablation study on different initial design sizes.** Final regret after 400 BO iterations using EI on five 10-dimensional benchmark functions. Results are reported as mean ± standard deviation over 5 seeds. Lower is better.

N_{init}	Method	Ackley	Rastrigin	Rosenbrock	Michalewicz	Sum of Different Powers
32	KSOS-BO	**0.019 ± 0.001**	**31.16 ± 34.51**	**8.95 ± 0.01**	5.56 ± 0.43	$\mathbf{7.12 \times 10^{-9} \pm 2.11 \times 10^{-9}}$
	Sobol Search	4.98 ± 0.50	77.39 ± 3.28	175.24 ± 96.38	4.79 ± 0.64	0.00968 ± 0.00144
	DE	4.10 ± 0.30	65.65 ± 5.68	135.32 ± 62.19	4.68 ± 0.30	0.00367 ± 0.00087
	CMA-ES	4.65 ± 0.46	62.16 ± 4.41	126.37 ± 17.60	5.09 ± 0.29	0.00292 ± 0.00138
	L-BFGS-B	2.87 ± 0.41	54.72 ± 6.31	92.48 ± 34.17	**3.92 ± 0.55**	0.00165 ± 0.00072
64	KSOS-BO	**0.012 ± 0.001**	**27.02 ± 38.14**	**8.96 ± 0.01**	4.81 ± 0.45	$\mathbf{9.04 \times 10^{-9} \pm 1.06 \times 10^{-8}}$
	Sobol Search	4.54 ± 0.71	67.36 ± 9.55	125.07 ± 25.71	4.92 ± 0.14	0.00993 ± 0.00230
	DE	3.92 ± 0.12	53.43 ± 8.14	114.75 ± 51.39	4.42 ± 0.26	0.00298 ± 0.00083
	CMA-ES	2.68 ± 0.46	47.00 ± 5.17	228.06 ± 21.24	4.99 ± 0.58	0.000933 ± 0.000127
	L-BFGS-B	1.94 ± 0.33	43.85 ± 5.72	78.62 ± 21.45	**3.71 ± 0.49**	0.00061 ± 0.00019

Python 3.12.12, ensuring a consistent computational environment across all benchmarks. As for the environment, we use NumPy, SciPy, scikit-learn, and CMA-ES (via the cma library (**?**)) as the underlying optimization and modeling frameworks for all benchmark algorithms. DE and L-BFGS-B baseline are implemented using SciPy, and KernelSOS is implemented using (Groudiev & Dümbgen).

C Additional Studies

C.1 Ablation across Acquisition Functions

We report additional results using PI ($\xi = 0.01$), and LCB ($\beta = 2$), as shown in Table 3. Except for the acquisition function, all other experimental settings are identical to those in Figure 2. Across all three acquisition functions, KSOS-BO achieves the lowest final regret on 4 out of 5 benchmarks, consistently outperforming the baselines on Ackley, Rastrigin, Rosenbrock, and Sum of Different Powers. This indicates that the advantage of KSOS-BO is robust to the choice of acquisition function.

C.2 Ablation Study on N_{init}

We study the influence of the initial design size N_{init} on the performance of different acquisition optimizers. In BO, N_{init} denotes the number of initial observations collected before sequential optimization begins and affects the quality of the initial surrogate model. In addition to $N_{\text{init}} = 12$ used in the main text, we evaluate $N_{\text{init}} = 32$ and $N_{\text{init}} = 64$ over 400 BO iterations. Except for the initial design size N_{init}, all other experimental settings are identical to those in Figure 2.

As shown in Table 4, KSOS-BO achieves the lowest final regret on 4 out of 5 benchmarks for both $N_{\text{init}} = 32$ and $N_{\text{init}} = 64$, consistently performing best on Ackley, Rastrigin, Rosenbrock, and Sum of Different Powers. Increasing N_{init} from 32 to 64 further reduces the regret on Ackley from 0.019 to 0.012 and on Michalewicz from 5.56 to 4.81, although the improvement is not uniform across all benchmarks, as reflected by the higher variability on Rastrigin. Overall, the relative performance of KSOS-BO remains robust across different initial design sizes.

Table 5: **Ablation study on different problem dimensions.** Final regret after 400 BO iterations using EI on five benchmark functions. Results are reported as mean ± standard deviation over 10 seeds. Lower is better.

Dimension	Method	Ackley	Rastrigin	Rosenbrock	Michalewicz	Sum of Different Powers
20D	KSOS-BO	**0.050 ± 0.0083**	**56.71 ± 15.90**	**18.96 ± 4.63**	6.91 ± 0.18	0.0074 ± 0.0013
	Sobol Search	7.04 ± 0.19	214.62 ± 10.69	921.91 ± 379.61	5.64 ± 0.37	0.054 ± 0.045
	DE	2.56 ± 0.037	181.52 ± 8.58	478.54 ± 47.16	4.26 ± 1.45	0.0063 ± 0.0029
	CMA-ES	3.75 ± 0.28	171.79 ± 29.55	1031.77 ± 93.86	4.96 ± 0.74	0.0079 ± 0.00084
	L-BFGS-B	3.54 ± 0.47	203.03 ± 27.58	1500.68 ± 292.51	**3.45 ± 1.43**	**0.0017 ± 0.0013**
50D	KSOS-BO	**0.087 ± 0.0067**	**234.91 ± 30.76**	**48.96 ± 6.30**	25.74 ± 3.68	0.35 ± 0.12
	Sobol Search	18.51 ± 0.40	700.62 ± 51.51	4261.86 ± 479.56	23.09 ± 5.81	2.05 ± 0.60
	DE	7.79 ± 0.16	718.28 ± 57.03	1159.94 ± 40.76	9.53 ± 1.09	0.71 ± 0.14
	CMA-ES	7.08 ± 0.53	570.88 ± 32.78	2573.64 ± 466.40	9.41 ± 1.75	0.45 ± 0.67
	L-BFGS-B	8.39 ± 0.14	524.05 ± 16.66	2080.57 ± 66.05	**7.16 ± 0.99**	**0.28 ± 0.098**

Table 6: **Ablation study on sampling methods.** Final regret after 400 iterations using EI on five 10-dimensional benchmark functions with $N_{\text{init}} = 12$. Results are reported as mean ± standard deviation over 10 seeds. Lower is better.

Method	Ackley	Rastrigin	Rosenbrock	Michalewicz	Sum of Different Powers
KSOS-BO (Sobol)	**0.020 ± 0.0073**	**14.97 ± 33.31**	**8.95 ± 0.009**	5.75 ± 0.46	$6.60 \times 10^{-9} \pm 7.27 \times 10^{-9}$
KSOS-BO (Uniform)	0.888 ± 0.096	34.06 ± 23.42	10.24 ± 0.41	**4.98 ± 0.13**	$6.56 \times 10^{-6} \pm 3.58 \times 10^{-6}$
Sobol Search	4.63 ± 0.67	72.38 ± 3.38	152.23 ± 89.28	5.00 ± 0.69	0.0080 ± 0.0042
Uniform Search	5.15 ± 0.36	71.47 ± 8.32	168.88 ± 38.77	5.11 ± 0.34	0.0093 ± 0.0059

C.3 Ablation Study on Dimensionality

We further analyze the scalability of all acquisition optimizers in higher-dimensional settings using EI. Specifically, we evaluate the methods on five benchmark functions at 20 and 50 dimensions. Except for the dimensionality, all other experimental settings are identical to those in Figure 2.

As shown in Table 5, KSOS-BO consistently performs best on Ackley, Rastrigin, and Rosenbrock at both 20D and 50D, while L-BFGS-B remains stronger on Michalewicz and Sum of Different Powers. More importantly, the relative performance pattern remains stable as the dimensionality increases, indicating that KSOS-BO maintains its optimization effectiveness in higher-dimensional search spaces. Overall, these results demonstrate the robustness of KSOS-BO to increasing problem dimensionality, particularly on challenging multimodal benchmarks.

C.4 Ablation Study on Sampling Methods

Here, we study the influence of sampling strategies on AFO. We compare two simple sampling-based baselines, Sobol Search and Uniform Search, both of which select the best candidate from a set of sampled points. We also evaluate the effect of replacing Sobol sampling with Uniform sampling within KSOS-BO. We exclude CMA-ES, DE, and L-BFGS-B because their search mechanisms are intrinsic to the algorithms and are not directly comparable to these sampling strategies. Except for the sampling strategies, all other experimental settings are identical to those in Figure 2.

As shown in Table 6, Sobol Search outperforms Uniform Search on 4 out of 5 benchmarks, although their overall performance remains relatively similar. Within KSOS-BO, Sobol sampling also achieves lower final regret on 4 out of 5 benchmarks, while Uniform sampling performs better on Michalewicz.

Importantly, even when Uniform sampling is used within KSOS-BO, it still performs better than standalone Sobol Search and Uniform Search across all five benchmarks. This shows that the gains of KSOS-BO cannot be attributed solely to Sobol sampling. Rather, the KernelSOS optimization procedure provides a substantial performance improvement, while Sobol sampling further enhances its performance through better sample coverage.

Table 7: **Ablation study on the kernel choice in KSOS-BO.** Final regret after 400 iterations using EI on five 10-dimensional benchmark functions. Results are reported as mean ± standard deviation over 10 seeds. Lower is better.

Method	Ackley	Rastrigin	Rosenbrock	Michalewicz	Sum of Different Powers
KSOS-BO (Gaussian)	0.0207 ± 0.0096	$\mathbf{24.89 \pm 43.00}$	$\mathbf{8.952 \pm 0.0078}$	$\mathbf{5.56 \pm 0.52}$	$1.06 \times 10^{-8} \pm 6.86 \times 10^{-9}$
KSOS-BO (Laplace)	$\mathbf{0.0122 \pm 0.0149}$	27.51 ± 47.65	8.968 ± 0.0159	5.77 ± 0.26	$\mathbf{4.44 \times 10^{-9} \pm 3.91 \times 10^{-9}}$

Table 8: **Ablation results for different EI exploration parameters ξ.** Final regret is reported as mean ± standard deviation over 10 seeds. Lower is better.

ξ	Method	Ackley	Rastrigin	Rosenbrock	Michalewicz	Sum of Different Powers
1	KSOS-BO	$\mathbf{0.0054 \pm 0.0093}$	$\mathbf{19.27 \pm 16.80}$	$\mathbf{32.86 \pm 17.06}$	6.05 ± 0.82	$\mathbf{3.16 \times 10^{-8} \pm 1.54 \times 10^{-9}}$
	Sobol Search	4.69 ± 0.37	71.10 ± 9.72	109.04 ± 6.28	5.41 ± 0.37	0.0076 ± 0.0044
	DE	4.19 ± 0.28	67.19 ± 9.18	73.91 ± 8.13	3.91 ± 0.75	0.00054 ± 0.00017
	CMA-ES	4.24 ± 0.46	53.52 ± 19.89	72.89 ± 17.72	4.32 ± 0.55	0.0012 ± 0.00077
	L-BFGS-B	4.10 ± 0.34	36.58 ± 20.19	58.32 ± 17.15	$\mathbf{3.76 \pm 0.49}$	0.00066 ± 0.00056
0.1	KSOS-BO	$\mathbf{0.016 \pm 0.028}$	$\mathbf{24.66 \pm 32.71}$	$\mathbf{24.60 \pm 9.19}$	5.78 ± 0.67	$\mathbf{5.79 \times 10^{-8} \pm 1.94 \times 10^{-9}}$
	Sobol Search	4.19 ± 0.37	72.76 ± 9.25	109.05 ± 6.28	5.46 ± 0.54	0.0081 ± 0.0037
	DE	2.94 ± 0.21	55.89 ± 4.29	46.21 ± 5.74	3.58 ± 0.56	0.00071 ± 0.00026
	CMA-ES	2.46 ± 0.47	47.25 ± 4.41	63.05 ± 20.82	3.95 ± 0.12	0.00091 ± 0.000064
	L-BFGS-B	2.07 ± 0.34	43.87 ± 3.06	39.02 ± 9.21	$\mathbf{3.29 \pm 0.19}$	0.00049 ± 0.00027

C.5 Ablation Study on KSOS-BO with Laplace Kernel

We study the effect of the kernel choice in KSOS-BO by replacing the Gaussian kernel with a Laplace kernel. Since the kernel choice only affects KSOS-BO, we directly compare the two KSOS variants while keeping all other experimental settings identical to those in Figure 2.

As shown in Table 7, replacing the Gaussian kernel with a Laplace kernel yields broadly comparable performance across the five benchmarks. The Laplace kernel achieves lower final regret on Ackley and Sum of Different Powers, while the Gaussian kernel performs slightly better on Rastrigin, Rosenbrock, and Michalewicz. Notably, the Laplace kernel also reduces the variance on Michalewicz and Sum of Different Powers. Overall, the two kernels produce similar optimization performance, suggesting that KSOS-BO is not strongly dependent on a specific kernel choice.

C.6 Ablation Study on ξ in EI

We report additional results using $\xi = 0.1$ and $\xi = 1$ for EI, where ξ controls the exploration–exploitation trade-off by setting the improvement threshold in the acquisition function. Except for the value of ξ, all experimental settings are identical to those reported in Figure 2. As shown in Table 8, the overall performance of KSOS-BO remains stable across different values of the EI exploration parameter ξ. For both $\xi = 0.1$ and $\xi = 1$, KSOS-BO achieves the lowest final regret on 4 out of 5 benchmarks, including Ackley, Rastrigin, Rosenbrock, and Sum of Different Powers, while L-BFGS-B performs best on Michalewicz. Although the absolute performance varies with ξ, the overall ranking remains unchanged. These results indicate that KSOS-BO is not strongly sensitive to moderate variations in the EI exploration parameter and maintains consistent optimization performance across different ξ settings.

C.7 Analysis on Inner Acquisition Function Optimization

We report a thorough analysis of the bare-bones acquisition function optimization performance. Because different acquisition optimizers induce different BO evolutions, their GP posteriors and acquisition functions diverge after the iteration; hence, raw acquisition values from their independent evolutions are not directly comparable. We therefore compare all optimizers on shared, fixed acquisition landscapes. Specifically, we evaluate all acquisition optimizers on the same GP posteriors and approximate the acquisition optimum of each acquisition landscape with a high-budget solver. Specifically, we evaluate 2^{16} scrambled Sobol points, refine the top 256 candidates with L-BFGS-B, and take the best value found across this procedure and all compared optimizers as the high-budget empirical reference maximum.

Table 9: **Acquisition function optimization study.** Lower acquisition gap is better, while higher Top-1% acquisition hit rate and average remaining-gap reduction are better. Bold indicates the best result for each benchmark and metric.

Optimizer	Metric	Ackley	Rastrigin	Rosenbrock	Sum of Different Powers	Michalewicz
KSOS	Acquisition Gap ↓	0.021 ± 0.011	0.018 ± 0.013	0.045 ± 0.023	0.025 ± 0.015	0.068 ± 0.043
	Top-1% Hit Rate (%) ↑	**91**	86	**95**	**93**	79
	Remaining-Gap Reduction (%) ↑	**17.52**	**3.57**	**26.85**	**35.49**	22.36
Sobol Search	Acquisition Gap ↓	0.23 ± 0.11	0.17 ± 0.11	0.44 ± 0.23	0.25 ± 0.14	0.28 ± 0.060
	Top-1% Hit Rate (%) ↑	69	81	74	76	63
	Remaining-Gap Reduction (%) ↑	0.32	0.44	3.87	2.16	3.94
L-BFGS-B	Acquisition Gap ↓	$\mathbf{0.010 \pm 0.013}$	$\mathbf{0.0092 \pm 0.0015}$	$\mathbf{0.022 \pm 0.011}$	0.019 ± 0.017	0.013 ± 0.017
	Top-1% Hit Rate (%) ↑	88	**89**	93	90	**96**
	Remaining-Gap Reduction (%) ↑	10.69	2.57	15.92	21.27	67.38
CMA-ES	Acquisition Gap ↓	0.079 ± 0.026	0.017 ± 0.014	0.087 ± 0.036	$\mathbf{0.017 \pm 0.011}$	0.037 ± 0.012
	Top-1% Hit Rate (%) ↑	84	76	90	86	85
	Remaining-Gap Reduction (%) ↑	11.03	2.43	16.35	22.28	44.27
DE	Acquisition Gap ↓	0.017 ± 0.038	0.016 ± 0.013	0.026 ± 0.014	0.035 ± 0.015	$\mathbf{0.012 \pm 0.0094}$
	Top-1% Hit Rate (%) ↑	87	83	91	89	94
	Remaining-Gap Reduction (%) ↑	11.01	2.73	17.68	20.74	**69.24**

Across 100 acquisition landscapes, we report three metrics. **Acquisition gap** measures the normalized difference between the acquisition value returned by an optimizer and the high-budget empirical reference maximum, reported as mean $\pm$ standard deviation. **Top-1% acquisition hit rate (%)** measures the fraction of landscapes for which the returned candidate lies within the top 1% of acquisition values in the independent reference set. **Average remaining-gap reduction (%)** measures the average fraction of the current objective gap eliminated by the returned candidate, with zero assigned when no improvement is obtained. Together, these metrics characterize both AFO quality and the downstream improvement produced by the selected query.

The results are reported in Table 9. On Ackley, Rastrigin, Rosenbrock, and Sum of Different Powers, KSOS achieves the highest or second-highest Top-1% hit rate and the largest average remaining-gap reduction, consistent with its strong BO performance on these benchmarks. On Michalewicz, KSOS has the second-lowest hit rate and remaining-gap reduction, consistent with its weaker performance on this task. Interestingly, the smallest acquisition gap does not always correspond to the largest downstream objective improvement. For example, although other optimizers attain smaller acquisition gaps on several benchmarks, KSOS achieves the largest remaining-gap reduction on the four benchmarks where it performs favorably in BO. These results suggest that acquisition value optimality alone may not fully explain downstream BO performance, and we leave a more detailed analysis of this relationship to future work.

D LIMITATIONS

KSOS-BO can be less effective on acquisition landscapes containing steep drops, extended plateaus, or regions of extreme curvature, where a smooth kernel representation may not adequately capture the local geometry. This behavior is reflected in its weaker performance on benchmarks such as Styblinski–Tang and Michalewicz. In addition, KSOS-BO may provide limited advantage on simple unimodal landscapes where local optimization is already effective. For example, Sphere is a separable, bowl-shaped function for which local sampling-based or gradient-based methods can achieve high final precision with little need for global exploration. Zakharov is also smooth and unimodal, but exhibits strong cross-dimensional coupling and anisotropic curvature, which can be difficult to approximate accurately from a finite set of samples, while local methods can efficiently exploit its single smooth basin. These cases suggest that the benefits of KSOS-BO are more pronounced on multimodal or structurally complex acquisition landscapes, whereas local methods may be preferable when the landscape contains a single well-behaved basin.

KSOS-BO also introduces additional computational overhead because each acquisition-optimization step requires solving a semidefinite optimization problem. Consequently, its wall-clock cost can be higher than that of Sobol Search, DE, and CMA-ES for each iteration. Finally, the quality of the recovered candidate depends on the kernel, sampling distribution and coverage, number of sampled points N, regularization parameter λ, and recovery procedure. The finite-dimensional KernelSOS formulation provides an approximation over the sampled representation rather than a full-domain

certificate of global optimality. These factors may limit performance and scalability, particularly in high-dimensional or highly irregular acquisition landscapes.