ROBUST SATISFICING IN BAYESIAN OPTIMIZATION

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF ENGINEERING AND SCIENCE
OF BILKENT UNIVERSITY
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR
THE DEGREE OF
MASTER OF SCIENCE
IN
ELECTRICAL AND ELECTRONICS ENGINEERING

By
Artun Saday
August 2025

Robust Satisficing in Bayesian Optimization
By Artun Saday
August 2025

We certify that we have read this thesis and that in our opinion it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Science.

Cem Tekin(Advisor)

Çağın Ararat

Elif Vural

Approved for the Graduate School of Engineering and Science:

Orhan Arıkan
Director of the Graduate School

ABSTRACT

ROBUST SATISFICING IN BAYESIAN OPTIMIZATION

Artun Saday
M.S. in Electrical and Electronics Engineering
Advisor: Cem Tekin
August 2025

This thesis explores the use of a recent optimization framework called "robust satisficing", in the context of Bayesian optimization. Robust satisficing (RS) is an alternative to optimizing, where the goal is to find a solution to a problem that achieves a predefined threshold robustly, under environmental uncertainties. On the other hand Bayesian optimization (BO) is a well-established framework for optimizing difficult to evaluate black-box functions. Previously, the BO literature has mainly focused on optimization, robust optimization or pure satisficing. The goal of this work is to introduce RS in to the BO framework. We analyze the problem in the contextual GP setting where distribution shifts on the contextual variable introduces uncertainties, and also in an adversarial setting where actions chosen are subjected to a perturbation from an adversary. We develop novel algorithms in both settings, using both a known RS approach and a new modification of it. We prove regret bounds for our algorithms in both settings and do extensive simulations that show the advantages of our approach compared with other state-of-the-art methods such as distributionally robust optimization.

Keywords: Robust Satisficing, Gaussian process bandits, Bayesian optimization, regret minimization.

ÖZET

BAYESCI ENIYILEMEDE GÜRBÜZ YETERLILIK

Artun Saday
Elektrik-Elektonik Mühendisliği, Yüksek Lisans
Tez Danışmanı: Cem Tekin
Ağustos 2025

Bu tez, "Bayesci eniyileme" bağlamında "gürbüz yeterlilik" (robust satisficing) adı verilen yeni bir optimizasyon çerçevesinin kullanımını incelemektedir. Sağlam yeterli optimizasyon (RS), klasik optimizasyonun bir alternatifi olarak ortaya çıkmakta; burada amaç, çevresel belirsizlikler altında, önceden belirlenmiş bir eşiği sağlam bir şekilde karşılayan bir çözüm bulmaktır. Öte yandan, Bayesyen optimizasyon (BO), değerlendirilmesi zor siyah-kutu fonksiyonların optimize edilmesinde yaygın olarak kullanılan yerleşik bir yaklaşımdır. BO literatürü şimdiye dek ağırlıklı olarak optimizasyon, sağlam optimizasyon veya doğrudan tatmin edici yöntemler üzerine odaklanmıştır. Bu tezde sunulan çalışmalar, RS yaklaşımını BO çerçevesine dahil eden ilk çalışmalardandır. Problemi, bağlamsal değişkenler üzerindeki dağılım kaymalarının belirsizlik yarattığı bağlamsal Gaussian Süreçleri (GP) ortamında ve seçilen eylemlerin bir saldırgan tarafından bozuma uğratıldığı adversaryel ortamda analiz ediyoruz. Her iki durumda da, hem bilinen bir RS yaklaşımını hem de bunun yeni bir modifikasyonunu kullanarak özgün algoritmalar geliştiriyoruz. Geliştirdiğimiz algoritmalar için her iki senaryoda da pişmanlık (regret) sınırlarını ispatlıyor ve yöntemimizin, dağılımsal olarak sağlam optimizasyon gibi güncel yöntemlere kıyasla üstünlüklerini ortaya koyan kapsamlı benzetim deneyleri sunuyoruz.

Anahtar sözcükler: Sağlam yeterlilik, Gauss süreci haydutları, Bayesçi eniyileme, pişmanlık minimizasyonu.

Acknowledgement

I am deeply grateful to my advisor, Assoc. Prof. Cem Tekin, for his invaluable guidance and support throughout my studies. His insight and encouragement have been instrumental in shaping this dissertation and in my growth as a researcher.

I would also like to thank my friends Cahit, Onat, Enes, and the two Efes. Their companionship made every day at the office enjoyable and kept me motivated throughout my graduate journey. I am especially indebted to Cahit, who introduced me to the joy of true collaboration and the seamless flow of working with a capable problem-solver. I am equally grateful for our many philosophical deep dives outside office hours, which enriched my perspective far beyond research.

My deepest gratitude goes to my mom, dad, and brother, whose unwavering support gave me courage in difficult times. Their belief in me has been a constant source of strength, and I owe them more than I can ever repay.

I would also like to extend my thanks to Prof. Çağın Ararat, for rekindling my love for rigorous mathematics and for exemplifying the professionalism and dedication that define academia.

Finally, this work was supported by the Scientific and Technological Research Council of Türkiye (TÜBİTAK) under Grant 124E065.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Optimization Objectives

A large majority of problems in science and engineering can be cast as a form of optimization problem. In engineering, this can mean setting the gain of a controller to maximize performance or tuning the parameters of a circuit to minimize power consumption. In machine learning, the goal is often to find model parameters that minimize empirical loss on a given dataset. In natural sciences optimization arises in diverse forms, from inferring parameters to a physical system to finding the molecular configuration with the lowest energy in computational chemistry. In economics and operational research, it governs decision making and resource allocation under constraints. Despite the differences in the fields these problems share the common structure of minimizing or maximizing a function under constraints and uncertainty, defined over a domain of possible solutions. In this thesis we will generally denote the objective function as $f(x, c)$ where $x \in \mathcal{X}$ is the decision variable the agent aims to optimize, and $c \in \mathcal{C}$ the contextual, or the uncertain variable, that the decision maker does not have direct control over. In different problems, the contextual variable can be a random variable, observational data, or might be selected by an adversary to harm the performance of the decision-maker. Depending of the nature of the problem and the goal of

the decision-maker, different optimization objectives arise.

1.1.1 Stochastic Optimization

In many real world problems the contextual variable $c \in \mathcal{C}$ is not known directly, and is modeled as a random variable with a reference distribution $c \sim \hat{P}$. The reference distribution can be determined through empirical observation. For instance, in machine learning the contextual variable can be the data and its distribution can be the empirical distribution from the training data. When the contextual variable $c \in \mathcal{C}$ is unknown, defining the best decision variable is not straight forward, and the decision-maker should define an optimization objective. One approach known as *stochastic optimization* (SO) is to maximize the expected value of the objective function under the reference distribution.

$$\textbf{Stochastic Optimization:} \quad x_{\text{SO}} \in \arg\max_{x \in \mathcal{X}} \mathbb{E}_{c \sim \hat{P}}[f(x, c)] \tag{1.1}$$

While SO is performs well in average, it does not give strong protection against worst case scenarios. Furthermore, if the true distribution of the contextual variable is different then the reference distribution announced to the decision-maker, the SO solution can fail. In machine learning this can manifest as a *distribution shift* between the training distribution and the deployment distribution. Robustness against distributions shifts is an active area of research in modern machine learning [1].

1.1.2 Robust Optimization

One factor that plays a role in many real world applications is the decision's robustness to unforeseen variations or perturbations on the contextual variable. It is desired that even if the contextual variable is misestimated, the decision maker still achieves good performance. One approach this problem is *robust optimization* (RO) [2]. In RO, the learner assumes that the actual value of the context $c^* \in \mathcal{C}$ is given within an ambiguity set $\mathcal{U} \subseteq \mathcal{C}$. The goal is to then optimize for the worst

case outcome in the ambiguity set.

$$\textbf{Robust Optimization: } x_{\text{RO}} \in \arg\max_{x \in \mathcal{X}} \min_{c \in \mathcal{U}} f(x, c) \tag{1.2}$$

The choice of the ambiguity ball heavily influences the solution. In practice the ambiguity ball is typically defined by a reference context value and a radius parameter, $\mathcal{U}(r) := \{c \in \mathcal{C} \,|\, \|c_{\text{ref}} - c\| \leq r\}$. Selecting the radius parameter judiciously is important as picking a small r results might results in a risky solution, and picking a large r would give an overly conservative suboptimal solution [3, 4].

1.1.3 Distributionally Robust Optimization

An alternative approach which combines the SO and RO is the framework of *distributionally robust optimization* (DRO). While RO gives strong worst-case robustness guarantees, if the context is a random variable with a known estimated distribution, it makes sense to make use of this information while selecting our action. Similar to RO, DRO requires an ambiguity set, but unlike RO, the ambiguity set is a set of distributions on the context space. Let $\mathcal{P}$ denote the space of probability distributions (a simplex on the finite context case) on the context space. Then the ambiguity set for DRO is defined as $\mathcal{U}(r) := \{P \in \mathcal{P} \,|\, d(\hat{P}, P) \leq r\}$, where $d : \mathcal{P} \times \mathcal{P} \to \mathbb{R}^+$ is probability distance metric that measures how different two distributions are. Common metrics include the Wasserstein distance and KL-divergence. The DRO objective is as follows:

$$\textbf{Distributionally Robust Optimization: } x_{\text{RO}} := \arg\max_{x \in \mathcal{X}} \min_{P \in \mathcal{U}} \mathbb{E}_{c \sim P}[f(x, c)] \tag{1.3}$$

DRO maximizes the worst-case expected value of the solution, given that the true distribution of the contextual variable lies within the ambiguity set of distributions. This way DRO manages to give robustness guarantees under distribution shifts, while also not sacrificing the average performance for unlikely singular outcomes.

1.1.4 Satisficing

In some problems finding the optimum solution is either not necessary, or the problem is infeasible. In such settings *satisficing* can be a viable option. In his 1978 Nobel Prize lecture in Economics, Herbert Simon remarked that *"decision makers can satisfice either by finding optimal solutions for a simplified world or by finding satisfactory solutions for a more realistic world."* Satisficing refers to the process of achieving a satisfactory utility level, denoted by a threshold τ (often called the aspiration level), under uncertainty. This behavior is frequently observed in decision-making contexts where agents operate under risk and uncertainty while exhibiting bounded rationality due to the inherent complexity of the problem, as well as computational and time constraints [5]. Since its introduction, the concept of satisficing has been studied across a wide range of disciplines, including economics [6], management science [7, 8], psychology [9], and engineering [10, 11]. In our setting, we can formalize the satisficing objective as:

$$\textbf{Satisficing:}\ \text{Find}\ x\ \ \text{s.t.}\ \ \mathbb{E}_{c\sim\hat{P}}[f(x,c)] \geq \tau\ , \tag{1.4}$$

where τ is the threshold the decision-makaer aims to achieve. Satisficing can also be interpreted as a stopping heuristic in optimization. In many optimization procedures, the majority of sampled evaluations tend to cluster near a local optimum, yielding only marginal improvements over time. By targeting a predefined satisfactory threshold rather than the exact optimum, satisficing mitigates this inefficiency and typically requires significantly fewer samples. In general, satisficing can be seen as a stopping heuristic during the optimization process.

1.1.5 Robust Satisficing

Robust satisficing (RS) is an emerging paradigm that extends the principles of satisficing to emphasize robustness. The central idea of RS is to identify a solution that achieves a satisficing threshold τ across the widest possible range of contextual outcomes. Unlike robust optimization (RO) and distributionally robust optimization (DRO), RS does not seek to maximize the worst-case outcome.

Instead, a solution is preferred if it consistently attains the threshold τ, even if its worst-case performance may fall below that of an RO solution. Decision theorists argue that RS is often more normative for decision-making under uncertainty due to its target-oriented nature [12, 13].

This perspective aligns with many real-world decision-making patterns, where agents act based on achieving a predefined threshold, even under adverse conditions. In engineering design, this corresponds to ensuring that a product meets certain performance requirements under a variety of operating conditions [14]. In medicine, RS naturally arises when aiming to maintain physiological variables, such as blood glucose levels, within a safe range under contextual covariates such as carbohydrate intake or physical activity. Similarly, a student preparing for an exam may select topics to study with the goal of ensuring a passing grade despite uncertainty about which questions will appear, and under limited study time. These examples illustrate how RS is implicitly embedded in both professional decision-making and daily life.

One of the first mathematical formulations of RS is from [15], which formalizes the objective of RS in the same setting as DRO, where the decision-maker has access to a reference distribution $\hat{P}$ about the contextual variable, but do not know the true distribution P^*. They define RS action in this setting as:

$$\textbf{Robust Satisficing: } x_{\mathrm{RS}} \in \arg\min k \text{ s.t. } \mathbb{E}_{c\sim P}[f(x,c)] \geq \tau - kd(P,\hat{P}), \quad (1.5)$$
$$\forall P \in \mathcal{P}_0 \ , x \in \mathcal{X},\ k \geq 0 \ .$$

To find x_{RS}, we can first define the *fragility* of $x \in \mathcal{X}$ as

$$\kappa_{\tau,t}(x) = \min k \text{ s.t. } \mathbb{E}_{c\sim P}[f(x,c)] \geq \tau - kd(P,\hat{P}),\ \forall P \in \mathcal{P}_0,\ k \geq 0 \ , \qquad (1.6)$$

where $\mathcal{P}_0$ is the probability simplex over $\mathcal{C}$.When (1.5) is feasible, the RS solution is the one with the minimum fragility, i.e., $x_{\mathrm{RS}} \in \arg\min_{x\in\mathcal{X}} \kappa_{\tau,t}(x)$ and $\kappa_{\tau,t} = \inf_{x\in\mathcal{X}} \kappa_{\tau,t}(x)$. Similar to DRO, RS utilizes a reference distribution assumed to be a proxy for the true distribution. However, unlike DRO, (1.5) do not define an ambiguity set $\mathcal{U}$ that represents all plausible distributions on $\mathcal{C}$ but rather define a threshold value τ, which the decision-maker aims to satisfy. In cases where the decision-maker is not confident that the reference distribution $\hat{P}$ accurately

Table 1.1: Comparison of optimization objectives.

Method	Inputs	Objective
SO	f, $\hat{P}$	Find $x \in \mathcal{X}$ that maximize $\mathbb{E}_{c\sim\hat{P}}[f(x,c)]$
RO	f, $\mathcal{U}$	Find $x \in \mathcal{X}$ that maximize $\inf_{c\in\mathcal{U}} f(x,c)$
DRO	f, $\hat{P}$ r	Find $x \in \mathcal{X}$ that maximize $\inf_{P\in\mathcal{U}} \mathbb{E}_{c\sim P}[f(x,c)]$
S	f, τ	Find $x \in \mathcal{X}$ that achieves $\mathbb{E}_{c\sim\hat{P}}[f(x,c)] \geq \tau$
RS	f, $\hat{P}$, τ	Find $x \in \mathcal{X}$ that minimize $\kappa(x)$ where
		$\kappa(x) = \min k$ s.t. $\mathbb{E}_{c\sim P}[f(x,c)] \geq \tau - k\Delta(P,\hat{P}),\ \forall P \in \mathcal{P}_0,\ k \geq 0$

represents the true distribution P^*, finding a meaningful ambiguity set can be difficult. In contrast, τ has a meaningful interpretation and can be expressed as a percentage of the SO solution computed under the reference distribution $\hat{P}$ over $\mathcal{C}$ given by[1]

$$Z_{\text{SO}} := \max_{x\in\mathcal{X}} \mathbb{E}_{c\sim\hat{P}}[f(x,c)] \ . \tag{1.7}$$

Unlike DRO, in RS, we are certain that our formulation covers the true distribution P^*, since $P^* \in \mathcal{P}_0$. The fragility can be viewed as the minimum rate of suboptimality one can obtain with respect to the threshold per unit of distribution shift from $\hat{P}$. The success of an action is measured by whether it achieves the desired threshold value τ in expectation. Depending on the objective function and the ambiguity set, the RS solution can differ from the DRO and the SO solutions (see Figure 1.1 for an example). Table 1.1 gives a comparison of the optimization objectives presented. One of the main advantages of RS compared to RO and DRO is that RS does not require an ambiguity set, which might be difficult to estimate accurately in some applications. The RO and DRO solutions are known to be highly sensitive to the choice of the ambiguity set. If it is too narrow, it may exclude relevant uncertainties; if too broad, it can lead to conservative solutions that underperform under typical conditions [3, 4].

[1]When f is unknown, Z_{SO} cannot be computed exactly. However, one can reflect the uncertainty about f by using a surrogate model to estimate the unknown objective function.

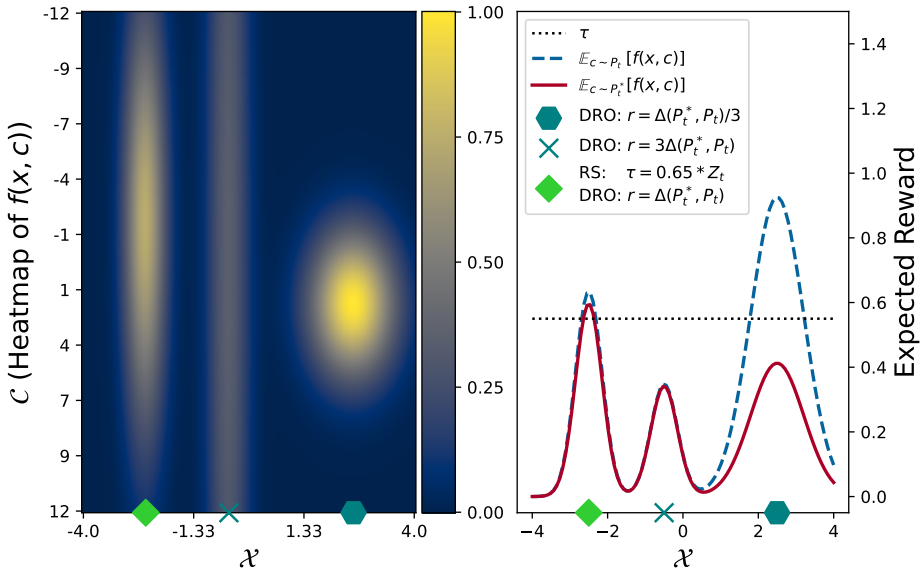


Figure 1.1: Examples of RS, DRO, and SO solutions where Z_{SO} is the solution to the SO problem given in (1.7). For DRO, $\mathcal{U}$ corresponds to a ball centered at $\hat{P}$ with radius r. Rhombus, cross, and hexagon correspond to RS and two other suboptimal solutions, respectively. Note that the SO solution corresponds to the point with a hexagon, i.e., a suboptimal solution. When the radius of the ambiguity ball of DRO captures the discrepancy between $\hat{P}$ and P^* perfectly, it selects the RS solution. When the ambiguity ball is too small or too large, it fails to find the RS solution and selects a suboptimal solution.

1.2 Bayesian Optimization

In real world problems where the objective function is unknown, costly to evaluate, and where we do not have access to its gradient, traditional optimization methods such as gradient descent or Newton's method become inapplicable. These settings are common in experimental sciences, engineering design, and hyperparameter tuning in machine learning pipelines where each function evaluation may correspond to a time consuming simulation, or training a large model. In such settings

it is crucial to choose informative points to evaluate the function. Bayesian optimization (BO) offers a principled way for evaluating such black-box functions by incorporating a probabilistic surrogate model of the unknown objective function, and choosing the point of evaluation by working with the surrogate function instead [16]. The optimal inputs are searched through an acquisition function, which aims to balance exploration and exploitation in an efficient manner. Designing acquisition functions for specific tasks is an active area of research and also will be the main focus of this thesis. BO has shown its effectiveness in applications such as hyperparameter tuning [17], engineering design [18], experimental design [19], and clinical modeling [20].

The main power of BO lies in its ability to incorporate uncertainty about the objective function, into its surrogate by using a probabilistic model. This commonly done by using a *Gaussian process* (GP).

Definition 1. A Gaussian Process is a distribution over functions, such that any finite subset of function evaluations has a joint Gaussian distribution, with consistency across different subsets. We denote a GP by $g(x) \sim \mathcal{GP}(\mu(x), k(x, x'))$.

Similar to a Gaussian random variable, a GP can be fully characterized by a mean function

$$\mu(x) = \mathbb{E}[g(x)] \ ,$$

and a covariance (or kernel) function

$$k(x, x') = \mathbb{E}[(g(x) - \mu(x))(g(x') - \mu(x'))] \ .$$

For any finite set of inputs $\{x_1, \dots x_n\}$ the corresponding function values $\{f(x_1), \dots, f(x_n)\}$ follow a multivariate normal distribution:

$$[g(x_1), \dots, g(x_n)]^\top \sim \mathcal{N}(\mu, K),$$

where $\mu_i = \mu(x_i)$ and $K_{ij} = k(x_i, x_j)$. The mean function captures the expected value of the function at each input, while the kernel function encodes assumptions about the function's smoothness, periodicity, or other structural properties. Depending on the kernel choice, functions drawn from the GP vary in behavior,

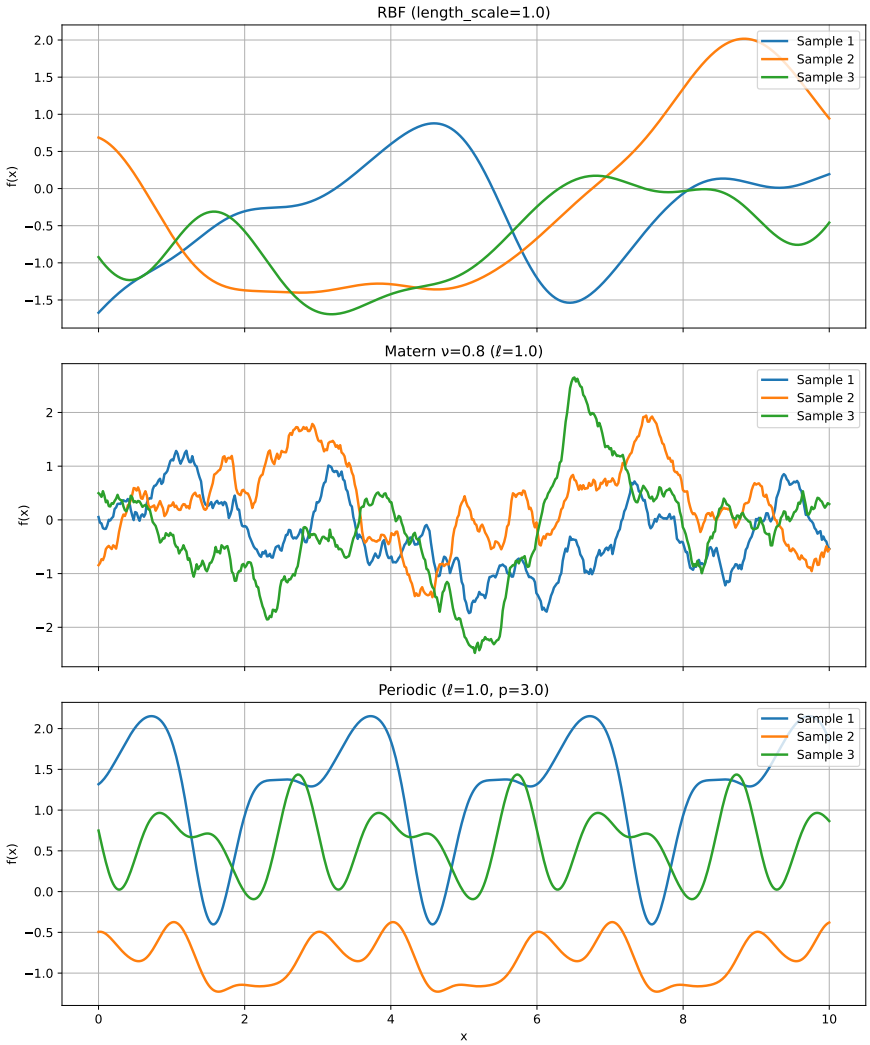


Figure 1.2: Illustration of 3 samples drawn from $f \sim \mathcal{N}(\mu, k)$ with different kernels.

hence choosing a suitable kernel function is important. Figure 1.2 shows how the kernel choice effects the functions drawn from the GP.

Given a GP prior over functions, we can analytically derive the posterior distribution after observing a set of noisy data using Bayes rule. Let $g \sim \mathcal{GP}(\mu(x), k(x, x'))$ be our GP prior, with mean function $m(x)$ and covariance function $k(x, x')$. Suppose we observe a dataset $\mathcal{D} = \{(x_i, y_i)\}_{i=1}^n$, where the observations are noisy evaluations of the function: $y_i = g(x_i) + \epsilon_i$, with $\epsilon_i \sim \mathcal{N}(0, \sigma^2)$ i.i.d. Gaussian noise. Denote by $\mathbf{X} = [x_1, \ldots, x_n]^\top \in \mathbb{R}^{n \times d}$, and $\mathbf{y} = [y_1, \ldots, y_n]^\top \in \mathbb{R}^n$ the

vector of observation inputs and outputs. For a test point x_*, we are interested in the posterior distribution of $g_* = g(x_*)$ given the data. Since the noise is independent, it only effects the marginal distribution of the observations, hence $y(x) \sim \mathcal{N}(\boldsymbol{\mu}, k + \sigma^2 \delta_{ii'})$, where $\delta_{ii'} = 1$ if $i = i'$ and 0 else. Under the GP prior, the joint distribution of the training inputs and the test function value is multivariate Gaussian:

$$\begin{bmatrix} \mathbf{y} \\ g_* \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \boldsymbol{\mu} \\ \mu_* \end{bmatrix}, \begin{bmatrix} K(\mathbf{X}, \mathbf{X}) + \sigma^2 I & k(\mathbf{X}, x_*) \\ k(x_*, \mathbf{X}) & k(x_*, x_*) \end{bmatrix} \right),$$

where $\boldsymbol{\mu} \in \mathbb{R}^n$ has entries $\mu(x_i)$, $\mu_* = \mu(x_*)$, $K(\mathbf{X}, \mathbf{X}) \in \mathbb{R}^{n \times n}$ is the kernel matrix with entries $K(\mathbf{X}, \mathbf{X})_{i,j} = k(x_i, x_j)$, and $k(\mathbf{X}, x_*) \in \mathbb{R}^n$ is the vector of covariances between the training inputs and the test input.

Using the standard formula for conditioning in multivariate Gaussians, the posterior distribution $g_* \mid \mathcal{D}$ is Gaussian:

$$g_* \mid \mathcal{D} \sim \mathcal{N}(\mu_*, \sigma_*^2),$$

with posterior mean and variance given by

$$\mu_* = \mu(x_*) + k(x_*, \mathbf{X})[K(\mathbf{X}, \mathbf{X}) + \sigma^2 I]^{-1}(\mathbf{y} - \boldsymbol{\mu}),$$

$$\sigma_*^2 = k(x_*, x_*) - k(x_*, \mathbf{X})[K(\mathbf{X}, \mathbf{X}) + \sigma^2 I]^{-1} k(\mathbf{X}, x_*).$$

In the special case where the prior mean function is zero (i.e., $\mu(x) = 0$), the expressions simplify to

$$\mu_* = k(x_*, \mathbf{X})[K + \sigma^2 I]^{-1}\mathbf{y}, \quad \sigma_*^2 = k(x_*, x_*) - k(x_*, \mathbf{X})[K + \sigma^2 I]^{-1} k(\mathbf{X}, x_*).$$

These expressions allow for exact Bayesian inference at test points, yielding both a prediction and a principled uncertainty estimate. This uncertainty-aware nature of GPs makes them particularly suitable for settings such as Bayesian optimization, where each evaluation of the objective is costly and uncertainty quantification is crucial for sample efficiency. For a detailed tutorial on GP regression, see [21].

1.3 Related Work and Our Contribution

1.3.1 Related Work

Bayesian optimization (BO) [22, 16] is a sample-efficient framework for optimizing complex black-box functions that are costly to evaluate. It is particularly well suited to settings where the objective is noisy, multimodal, or lacks gradient information. BO couples a probabilistic model—commonly a Gaussian process (GP)—with an acquisition strategy to select evaluation points that efficiently balance exploration and exploitation. This methodology has gained traction across a variety of sequential decision-making problems, including hyperparameter tuning in machine learning [23], vaccine and drug development [24], and dynamic treatment regimes in healthcare [25].

Contextual BO [26] extends the framework to incorporate exogenous variables, or *contexts*, which influence the outcome but are not under the learner's control. When the distribution over contexts is known, a common objective is to maximize expected utility [27, 28]. In practice, however, the assumed reference distribution may differ from the true context distribution encountered at deployment. Such distributional mismatch can lead to solutions that are suboptimal—or even unsafe—when applied in the real environment. To mitigate this, several approaches have been proposed to handle distribution shift within BO.

Robustness, and in particular adversarial robustness, has been extensively studied in the context of GP optimization. [29] address BO with outliers by combining robust GP regression with outlier diagnostics. [30], which is closely related to our work, introduce adversarial robustness in GP bandits and propose a robust optimization (RO) algorithm that assumes a known perturbation budget. In contrast, we adopt a robust satisficing (RS) approach that does not rely on this assumption. [27] consider a related setting in which the learner selects x_t but observes the output at a different location $\tilde{x}_t \sim P_{x_t}$, modeling the problem using a GP over probability distributions. In the noisy input setting, [31] propose an entropy search algorithm. Adversarial corruption of observations, rather than

inputs, is investigated by [32, 33]. Uncertainty in the context distribution is studied in contextual GP optimization. [34] address this problem using RO with a known ambiguity set. A complementary line of research focuses on the optimization of risk measures in BO [35, 36, 37], while [38] propose jointly optimizing the mean and variance of $f(x, \omega)$ using a GP.

The idea of *satisficing*, introduced by Herbert Simon [39], provides a complementary perspective to optimization. Rather than seeking a global optimum, the goal is to find a decision x whose utility meets or exceeds a predefined threshold τ (the *aspiration level*) under uncertainty. Simon famously noted in his 1978 Nobel Prize lecture that "decision makers can satisfice either by finding optimal solutions for a simplified world or by finding satisfactory solutions for a more realistic world." Satisficing behavior is frequently observed when decision makers face uncertainty, limited information, or computational and time constraints, and has been studied in domains ranging from economics [6], management science [7, 8], and psychology [9] to engineering [10, 11]. In recent years, satisficing has been formalized in the multi-armed bandit setting, both in terms of regret minimization [40, 41, 42, 43] and good-arm identification [44]. Notably, [44] show that satisficing designs can often be identified with substantially fewer samples than are required for optimal designs, while [42] demonstrate that in time-sensitive problems with discounted rewards, satisficing strategies can yield greater cumulative returns than those converging to an optimum.

Robust satisficing (RS) extends this principle by emphasizing reliability under uncertainty. In the seminal work of [12], RS is described as identifying a decision that both satisfices and remains resilient to variations in the environment. [15] formulate RS as an optimization problem and propose methods for estimating robust satisficing designs efficiently.

1.3.2 Contributions

Building on the foundations presented, we introduce *robust Bayesian satisficing* (RBS) as an alternative paradigm for robust Bayesian optimization. The aim

of RBS is to achieve performance at or above the aspiration level τ despite unrestricted shifts in the unknown covariate, ensuring satisfactory outcomes under evolving and potentially adverse conditions. We tackle the problem in two distinct settings:

- The contextual GP setting presented in [34], where the learner has access to a reference distribution about the random context, but does not have access to the true distribution.
- The adversarial GP setting introduced in [30], where an adversary applies input perturbations to the points selected by the learner.

Building on top of the work of [15], we introduce a new parametric formulation of robust satisficing, that recovers previously known formulations ([15] and [45]) as edge cases. We demonstrate how a spectrum of robust solutions can be obtained by varying the parameter. Using these formulations we present new BO algorithms tailored for robust satisficing. We provide theoretical bounds on the regrets of our algorithms under certain assumptions in both the distributional and adversarial settings. Finally, we present numerical experiments, including in-silico experiments on an FDA-approved simulator for diabetes management with realistic adversarial perturbations, to demonstrate the strengths of our algorithms addressing the shortcomings of other approaches known in the literature.

Chapter 2

Robust Bayesian Satisficing in the Distributional Setting

2.1 Problem Formulation

2.1.1 Setup

Let $f\colon \mathcal{X} \times \mathcal{C} \to \mathbb{R}$ be an *unknown* reward function defined over a parameter space with a finite action set $\mathcal{X}$ and a finite context set $\mathcal{C} := \{c_1, \ldots, c_n\}$. Let $\mathcal{P}_0$ denote the set of all probability distributions over $\mathcal{C}$. Our goal is to sequentially optimize f based on noisy observations. At each round $t \in [T]$, the environment first provides a reference distribution $P_t \in \mathcal{P}_0$ for the context variable, the learner selects an action $x_t \in \mathcal{X}$, and the environment then reveals a context $c_t \in \mathcal{C}$ together with a noisy observation

$$y_t = f(x_t, c_t) + \eta_t,$$

where η_t is conditionally σ-subgaussian given the history $(x_1, c_1, y_1), \ldots, (x_t, c_t)$. We assume that c_t is drawn independently from a time-varying, unknown distribution $P_t^* \in \mathcal{P}_0$, which may differ from P_t. Note that the learner has to decide

on their action without the knowledge of the true context c_t, but only the reference distribution. The distributions P_t and P_t^* are represented by n-dimensional nonnegative vectors w_t and w_t^*, respectively, satisfying $\|w_t\|_1 = \|w_t^*\|_1 = 1$. The distance between two distributions $P, P' \in \mathcal{P}_0$ is denoted by $\Delta(P, P')$; in this work, we adopt the *maximum mean discrepancy* (MMD) as the distance measure.

2.1.2 Regret Measures

Regret measures. We study RS through the lens of regret minimization, where the goal is to select a sequence of actions $x_1, \ldots, x_T$ that minimizes the growth rate of regret. In particular, we introduce two notions of regret. The first is the *lenient regret* [43], defined as

$$R_T^l := \sum_{t=1}^{T} \left(\tau - \mathbb{E}_{c\sim P_t^*}[f(x_t, c)]\right)^+, \tag{2.1}$$

where $(\cdot)^+ := \max\{0, \cdot\}$. Lenient regret quantifies the cumulative shortfall of the learner relative to the target threshold τ. Actions that meet or exceed τ in expectation incur no regret, whereas actions that fall short accumulate the gap. Importantly, achieving sublinear lenient regret under arbitrary distribution shifts is infeasible: even the RS-optimal action x_t^*, computed with full knowledge of f, can only guarantee an expected reward of at least $\tau - \kappa_{\tau,t}\,\Delta(P_t^*, P_t)$. Consequently, our upper bounds on lenient regret necessarily depend on the magnitude of distribution shifts.

We further introduce a new regret notion, termed *robust satisficing regret*, defined as

$$R_T^{rs} := \sum_{t=1}^{T} \left(\tau - \kappa_{\tau,t}\Delta(P_t^*, P_t) - \mathbb{E}_{c\sim P_t^*}[f(x_t, c)]\right)^+. \tag{2.2}$$

This measure quantifies the cumulative shortfall of the learner relative to the guarantee of the robust satisficing action, under the true distribution $\tau - \kappa_{\tau,t}\,\Delta(P_t^*, P_t)$. In particular, the robust satisficing action x_t^* satisfies

$$\mathbb{E}_{c\sim P_t^*}[f(x_t^*, c)] \;\;\geq\;\; \tau - \kappa_{\tau,t}\,\Delta(P_t^*, P_t)\;.$$

It follows immediately that $R_T^{rs} \leq R_T^l$. Moreover, in the absence of distribution shifts (i.e., when $P_t^* = P_t$ and (1.5) is feasible for all t), the two regret notions coincide.

To develop algorithms that minimize both (2.1) and (2.2), we adopt *Gaussian processes* (GPs) as surrogate models for the unknown function f. This requires us to impose mild regularity assumptions on f, which we detail below.

2.1.3 Regularity assumptions.

We assume that f lies in a *reproducing kernel Hilbert space* (RKHS) $\mathcal{H}_k$ associated with a kernel function k. The kernel is taken to be a product kernel of the form

$$k\big((x,c),(x',c')\big) = k_{\mathcal{X}}(x,x')\,k_{\mathcal{C}}(c,c'),$$

where $k_{\mathcal{X}}$ and $k_{\mathcal{C}}$ are kernel functions defined on the action space $\mathcal{X}$ and context space $\mathcal{C}$, respectively. We further assume that the Hilbert norm of f is bounded, i.e., $\|f\|_{\mathcal{H}_k} \leq B$. This assumption is standard in the Bayesian optimization literature and facilitates the use of Gaussian processes as surrogate models [46].

2.1.4 Maximum Mean Discrepancy

We adopt the *maximum mean discrepancy* (MMD) as our distance metric due to its ability to compare probability distributions and its relation to RKHSs. MMD operates by mapping each distribution into an RKHS through a feature map associated with a kernel, and then computing the distance between the corresponding mean embeddings. When a characteristic kernel[1] is used, this embedding uniquely represents the distribution, enabling the MMD to detect differences in the full distribution rather than only in a limited set of moments.

[1]A kernel k is called *characteristic* if the mean embedding mapping $P \mapsto \mu_P$ is injective, i.e., $\mu_P = \mu_Q$ implies $P = Q$. Common examples of characteristic kernels include the radial basis function (RBF) kernel and Laplace kernel.

Formally, let P and Q be probability distributions over a space $\mathcal{Z}$, let $k : \mathcal{Z} \times \mathcal{Z} \to \mathbb{R}$ be a positive-definite kernel with associated RKHS $\mathcal{H}_k$ and feature map $\varphi : \mathcal{Z} \to \mathcal{H}_k$. The MMD is defined as

$$\mathrm{MMD}_k(P, Q) := \sup_{\|f\|_{\mathcal{H}_k} \leq 1} \left(\mathbb{E}_{z \sim P}[f(z)] - \mathbb{E}_{z \sim Q}[f(z)] \right),$$

where z is a dummy variable representing a draw from either P or Q. Using the reproducing property $f(z) = \langle f, k(\cdot, z) \rangle_{\mathcal{H}_k} = \langle f, \varphi(z) \rangle_{\mathcal{H}_k}$, the expectations can be rewritten as

$$\mathbb{E}_{z \sim P}[f(z)] = \langle f, \mathbb{E}_{z \sim P}[\varphi(z)] \rangle_{\mathcal{H}_k}.$$

Define

$$\mu_P := \mathbb{E}_{z \sim P}[\varphi(z)]$$

as the *mean embedding* of P in $\mathcal{H}_k$. An *embedding* is simply the representation of each point $z \in \mathcal{Z}$ as a vector $\varphi(z)$ in the RKHS induced by k. The mean embedding μ_P is then the average of these feature vectors under the distribution P. Substituting and applying the Cauchy–Schwarz inequality yields

$$\begin{aligned} \mathrm{MMD}_k(P, Q) &= \sup_{\|f\|_{\mathcal{H}_k} \leq 1} \langle f, \mu_P - \mu_Q \rangle_{\mathcal{H}_k} \\ &\leq \|f\|_{\mathcal{H}_k} \|\mu_P - \mu_Q\|_{\mathcal{H}_k}. \end{aligned} \tag{2.3}$$

Thus, the supremum in equation 2.3 is achieved when f is aligned with $\mu_P - \mu_Q$ and is of unit norm, showing that the MMD is exactly the RKHS norm of the difference between the mean embeddings of P and Q.

2.1.5 Modeling the Unknown Function

Let $\mathcal{Z} := \mathcal{X} \times \mathcal{C}$. We model the unknown reward function f using a GP prior with mean function $\mu(z) = \mathbb{E}[f(z)]$ and positive definite kernel $k(z, z') = \mathbb{E}[(f(z) - \mu(z))(f(z') - \mu(z'))]$. For simplicity, we assume a zero mean prior, i.e., $\mu(z) = 0$, and that the kernel is bounded as $k(z, z) \leq 1$ for all $z \in \mathcal{Z}$. Thus, the prior over f is given by $f \sim \mathcal{GP}(0, k(z, z'))$. Assuming Gaussian likelihood with variance $\lambda > 0$, the posterior distribution of f given observations $\mathbf{y}_t = [y_1, \ldots, y_t]^\mathsf{T}$

at inputs $\mathbf{z}_t = [z_1, \ldots, z_t]^\mathsf{T}$ is again a GP. At the beginning of round $t \geq 1$, the posterior mean and covariance are given by:

$$\begin{aligned}
\mu_t(z) &= \boldsymbol{k}_{t-1}(z)^\mathsf{T}(\boldsymbol{K}_{t-1} + \lambda \boldsymbol{I}_{t-1})^{-1}\boldsymbol{y}_{t-1} \\
k_t(z, z') &= k(z, z') - \boldsymbol{k}_{t-1}(z)^\mathsf{T}(\boldsymbol{K}_{t-1} + \lambda \boldsymbol{I}_{t-1})^{-1}\boldsymbol{k}_{t-1}(z') \\
\sigma_t^2(z) &= k_t(z, z) \ ,
\end{aligned}$$

where $\boldsymbol{k}_t(z) = [k(z, z_1) \ldots k(z, z_t)]^\mathsf{T}$, $\boldsymbol{K}_t$ is the $t \times t$ kernel matrix of the observations with $(\boldsymbol{K}_t)_{ij} = k(z_i, z_j)$ and $\boldsymbol{I}_t$ is the $t \times t$ identity matrix. Let $\sigma_t^2(z) := k_t(z, z)$ represent the posterior variance of the model. Define $\mu_0 := 0$ and $\sigma_0^2(z) := k(z, z)$.

Our analysis utilizes the *maximum information gain*, which is standard in the literature.

Definition 2. The *maximum information gain* at round t is defined as

$$\gamma_t := \max_{A \subseteq \mathcal{Z},\, |A| = t} \mathrm{I}(\boldsymbol{y}_A; \boldsymbol{f}_A) \ , \tag{2.4}$$

where $\mathbf{y}_A$ denotes the set of noisy observations obtained at the design points A, $\mathbf{f}_A$ the corresponding function values, and $\mathrm{I}(\mathbf{y}_A; \mathbf{f}_A)$ the mutual information between the two.

For the set of observation points $A = \{z_1, \ldots, z_t\}$, the surrogate give the distribution $\boldsymbol{f}_A \sim \mathcal{N}(0, \boldsymbol{K}_A)$, the observations given the function are distributed as $\boldsymbol{y}|\boldsymbol{f}_A \sim \mathcal{N}(\boldsymbol{f}_A, \lambda \boldsymbol{I})$, and the observations have the marginal distribution $\mathcal{N}(0, \boldsymbol{K}_A + \lambda \boldsymbol{I}_t)$. The mutual information can be written as the difference $H(\boldsymbol{y}_A) - H(\boldsymbol{y}_A|\boldsymbol{f}_A)$. The differential entropy of a multivariate Gaussian is $H(\mathcal{N}(0, \Sigma)) = \frac{1}{2}\log((2\pi e)^t \det(\boldsymbol{\Sigma}))$. Hence we can write the mutual information as

$$\begin{aligned}
\mathrm{I}(\boldsymbol{y}_A; \boldsymbol{f}_A) &= H(\boldsymbol{y}_A) - H(\boldsymbol{y}_A|\boldsymbol{f}_A) \\
&= \frac{1}{2}\log\big((2\pi e)^t \det(\boldsymbol{K}_A + \lambda \boldsymbol{I}_t)\big) - \frac{1}{2}\log\big((2\pi e)^t \det(\lambda \boldsymbol{I}_t)\big) \\
&= \frac{1}{2}\log\left(\frac{\det(\boldsymbol{K}_A + \lambda \boldsymbol{I}_t)}{\det(\boldsymbol{I}_t)}\right) = \frac{1}{2}\log\big(\lambda^{-1}\boldsymbol{K}_A + \boldsymbol{I}_t\big) \ ,
\end{aligned}$$

hence the maximum information gain can be written as

$$\gamma_t = \max_{A \subseteq \mathcal{Z}\, |A| = t} , \frac{1}{2}\log\big(\lambda^{-1}\boldsymbol{K}_A + \boldsymbol{I}_t\big) \ .$$

The next result, due to [47] and building on [48, Theorem 2], provides sharp confidence bounds for functions f whose Hilbert norm in the RKHS is bounded.

Lemma 1. *[47, Theorem 1] Let $\delta \in (0,1)$, $\bar{\lambda} := \max\{1, \lambda\}$, and*

$$\beta_t(\delta) := \frac{\sigma}{\sqrt{\lambda}} \sqrt{\log\big(\det\big(\bar{\lambda}/\lambda \boldsymbol{K}_{t-1} + \bar{\lambda} \boldsymbol{I}_{t-1}\big)\big) + 2\log\left(\frac{1}{\delta}\right)} + B \ .$$

Then, the following holds with probability at least $1-\delta$:

$$|\mu_t(x,c) - f(x,c)| \leq \beta_t(\delta)\sigma_t(x,c), \ \forall x \in \mathcal{X}, \ \forall c \in \mathcal{C}, \ \forall t \geq 1 \ .$$

For simplicity, we set $\lambda = 1$ throughout the remainder of this chapter, and note that $\log\det(\mathbf{K}_{t-1} + \mathbf{I}_{t-1}) \leq 2\gamma_{t-1}$. When the confidence parameter δ is clear from context, we write β_t in place of $\beta_t(\delta)$ for brevity. We define the *upper confidence bound* (UCB) and *lower confidence bound* (LCB) for each $(x,c) \in \mathcal{X} \times \mathcal{C}$ as

$$\mathrm{ucb}_t(x,c) := \mu_t(x,c) + \beta_t \sigma_t(x,c), \qquad \mathrm{lcb}_t(x,c) := \mu_t(x,c) - \beta_t \sigma_t(x,c).$$

For a fixed action $x \in \mathcal{X}$, we denote by $\mathrm{ucb}_x^t = [\mathrm{ucb}_t(x,c_1), \ldots, \mathrm{ucb}_t(x,c_n)]^\mathsf{T} \in \mathbb{R}^n$ and $\mathrm{lcb}_x^t = [\mathrm{lcb}_t(x,c_1), \ldots, \mathrm{lcb}_t(x,c_n)]^\mathsf{T} \in \mathbb{R}^n$ the corresponding UCB and LCB vectors. We also write $f_x = [f(x,c_1), \ldots, f(x,c_n)]^\mathsf{T}$ for the vector of true function values.

2.2 RoBOS for Robust Bayesian Satisficing

To implement *robust Bayesian satisficing* (RBS), we introduce a learning algorithm called *Robust Bayesian Optimistic Satisficing* (RoBOS), whose pseudocode is provided in Algorithm 1. At the beginning of each round t, RoBOS observes the reference distribution w_t. It then computes the UCB index $\mathrm{ucb}_t(x,c)$ for each action–context pair $(x,c) \in \mathcal{X} \times \mathcal{C}$, using the GP posterior mean and variance at round t. The UCB values, together with the threshold τ and reference distribution w_t, are used to calculate the *estimated fragility* of each action x at round t, defined

as

$$\hat{\kappa}_{\tau,t}(x) = \begin{cases} \max_{w \in \Delta(\mathcal{C}) \setminus \{w_t\}} \dfrac{\tau - \langle w, \mathrm{ucb}_x^t \rangle}{\|w - w_t\|_M} & \text{if } \langle w_t, \mathrm{ucb}_x^t \rangle \geq \tau, \\ +\infty & \text{if } \langle w_t, \mathrm{ucb}_x^t \rangle < \tau, \end{cases} \tag{2.5}$$

where $\Delta(\mathcal{C})$ denotes the probability simplex over $\mathcal{C}$, M is the $n \times n$ kernel matrix, and $\|w\|_M := \sqrt{w^\top M w}$ corresponds to the MMD metric. Specifically, given the context kernel $k_\mathcal{C} : \mathcal{C} \times \mathcal{C} \to \mathbb{R}_+$, the matrix M is defined by $M_{ij} = k_\mathcal{C}(c_i, c_j)$. The estimated fragility $\hat{\kappa}_{\tau,t}(x)$ serves as an optimistic surrogate for the true fragility $\kappa_{\tau,t}(x)$.

Observe that if $\hat{\kappa}_{\tau,t}(x) \leq 0$, then the threshold τ is satisfied under *any* context distribution, given the UCB estimates of the rewards for action x. Conversely, if $\tau > \mathbb{E}_{c \sim P_t}[\mathrm{ucb}_t(x, c)]$, the threshold cannot be met under the reference distribution based on the UCB indices, which implies $\hat{\kappa}_{\tau,t}(x) = \infty$. The following lemma establishes the relationship between $\hat{\kappa}_{\tau,t}(x)$ and the true fragility $\kappa_{\tau,t}(x)$.

Lemma 2. *Fix $\tau \in \mathbb{R}$. With probability at least $1 - \delta$, $\hat{\kappa}_{\tau,t}(x) \leq \kappa_{\tau,t}(x)$ for all $x \in \mathcal{X}$ and $t \geq 1$.*

Proof. Assume that the confidence intervals in Lemma 1 hold. When $\hat{\kappa}_{\tau,t}(x) = \infty$, we also have $\kappa_{\tau,t}(x) = \infty$. When $\hat{\kappa}_{\tau,t}(x) < \infty$ and $\kappa_{\tau,t}(x) = \infty$, the inequality holds. Recall that when $\langle w_t, \mathrm{ucb}_x^t \rangle \geq \tau$, $\bar{w}_x^t := \arg\max_{w \in \Delta(\mathcal{C}) \setminus \{w_t\}} \frac{\tau - \langle w, \mathrm{ucb}_x^t \rangle}{\|w - w_t\|_M}$ represents the context distribution that achieves $\hat{\kappa}_{\tau,t}(x)$. When $\langle w_t, f_x \rangle \geq \tau$, let $\bar{\bar{w}}_x^t := \arg\max_{w \in \Delta(\mathcal{C}) \setminus \{w_t\}} \frac{\tau - \langle w, f_x \rangle}{\|w - w_t\|_M}$ be the context distribution that achieves $\kappa_{\tau,t}(x)$. When $\hat{\kappa}_{\tau,t}(x) < \infty$ and $\kappa_{\tau,t}(x) < \infty$, we have

$$\begin{aligned} \hat{\kappa}_{\tau,t}(x) - \kappa_{\tau,t}(x) &\leq \frac{\tau - \langle \bar{w}_x^t, \mathrm{ucb}_x^t \rangle}{\|\bar{w}_x^t - w_t\|_M} - \frac{\tau - \langle \bar{\bar{w}}_x^t, f_x \rangle}{\|\bar{\bar{w}}_x^t - w_t\|_M} \\ &\leq \frac{\tau - \langle \bar{w}_x^t, \mathrm{ucb}_x^t \rangle}{\|\bar{w}_x^t - w_t\|_M} - \frac{\tau - \langle \bar{w}_x^t, f_x \rangle}{\|\bar{w}_x^t - w_t\|_M} \\ &= \frac{\langle \bar{w}_x^t, f_x - \mathrm{ucb}_x^t \rangle}{\|\bar{w}_x^t - w_t\|_M} \\ &\leq 0 \ . \end{aligned}$$

□

To carry out robust satisficing, RoBOS selects the action with the smallest estimated fragility, i.e., $x_t = \arg\min_{x \in \mathcal{X}} \hat{\kappa}_{\tau,t}(x)$. After the action x_t is chosen, the context c_t and outcome y_t are observed from the environment, and these are subsequently used to update the GP posterior.

Algorithm 1 RoBOS

Inputs $\mathcal{X}$, $\mathcal{C}$, τ, GP kernel k, $\mu_0 = 0$, σ, confidence parameter δ, B

$t = 1, 2, \ldots$ 1. Observe the reference distribution w_t

2. Compute $\mathrm{ucb}_t(x, c) = \mu_t(x, c) + \beta_t \sigma_t(x, c)$ for all $x \in \mathcal{X}$ and $c \in \mathcal{C}$

3. Compute $\hat{\kappa}_{\tau,t}(x)$ as in (2.5)

4. Choose action $x_t = \arg\min_{x \in \mathcal{X}} \hat{\kappa}_{\tau,t}(x)$

5. Observe $c_t \sim P_t^*$ and $y_t = f(x_t, c_t) + \eta_t$

6. Use $\{x_t, c_t, y_t\}$ to compute μ_{t+1} and σ_{t+1}

2.3 Regret Analysis

We analyze the regret under the assumption that (1.5) is feasible.

Assumption 1. Let $\hat{x}_t := \arg\max_{x \in \mathcal{X}} \langle w_t, f_x \rangle$ denote the optimal action under the reference distribution. We assume that $\tau \leq \langle w_t, f_{\hat{x}_t} \rangle$, $\forall t \in [T]$, i.e., the threshold does not exceed the solution to the SO problem.

If Assumption 1 fails to hold in round t, then (1.5) becomes infeasible. In this case, we have $\kappa_{\tau,t} = \infty$, implying that no robust satisficing solution exists. Consequently, measuring regret in such rounds is not meaningful. In practice, if the learner is willing to adjust its aspiration level, Assumption 1 can be relaxed by dynamically selecting the threshold τ at each round so that $\tau < \langle w_t, \mathrm{lcb}^t_{\hat{x}'_t} \rangle$, where $\hat{x}'_t := \arg\max_{x \in \mathcal{X}} \langle w_t, \mathrm{lcb}^t_x \rangle$.[2]

[2]Our algorithm can be readily adapted to handle dynamic thresholds $\{\tau_t\}_{t \geq 1}$. Specifically, if we also replace the thresholds in the regret definitions (2.1) and (2.2) with τ_t, and update Assumption 1 to require $\tau_t \leq \langle w_t, f_{\hat{x}_t} \rangle$ for all $t \in [T]$, then all derived regret bounds remain valid.

The next theorem establishes an upper bound on the robust satisficing regret of RoBOS in terms of the maximum information gain defined in (2.4).

Theorem 2. *Let $\delta \in (0,1)$. Suppose RoBOS is executed under Assumption 1 with confidence parameter $\delta/2$, and let $\beta_t := \beta_t(\delta/2)$ be defined as in Lemma 1. Then, with probability at least $1-\delta$, the robust satisficing regret R_T^{rs} of RoBOS satisfies*

$$R_T^{rs} \;\leq\; 4\beta_T \sqrt{T\Big(2\gamma_T + 2\log\Big(\tfrac{12}{\delta}\Big)\Big)}\;.$$

Proof Sketch: Assume that confidence bounds in Lemma 1 hold (happens with probability at least $1-\delta/2$). Let $r_t^{rs} := \tau - \kappa_{\tau,t}\Delta(P_t^*, P_t) - \mathbb{E}_{c\sim P_t^*}[f(x_t,c)]$ be the *instantaneous regret*. Robust satisficing regret can be written as $R_T^{rs} = \sum_{t=1}^T r_t^{rs}\mathbb{I}(r_t^{rs}\geq 0)$. We have

$$\begin{aligned}
r_t^{rs} &\leq \tau - \kappa_{\tau,t}\|w_t^* - w_t\|_M - \langle w_t^*, \mathrm{ucb}_{x_t}^t\rangle + 2\beta_t\langle w_t^*, \sigma_t(x_t,\cdot)\rangle && (2.6)\\
&\leq \tau - \kappa_{\tau,t}\|w_t^* - w_t\|_M - (\tau - \hat{\kappa}_{\tau,t}\|w_t^* - w_t\|_M) + 2\beta_t\langle w_t^*, \sigma_t(x_t,\cdot)\rangle && (2.7)\\
&= \|w_t^* - w_t\|_M(\hat{\kappa}_{\tau,t} - \kappa_{\tau,t}) + 2\beta_t\langle w_t^*, \sigma_t(x_t,\cdot)\rangle\\
&\leq 2\beta_t\langle w_t^*, \sigma_t(x_t,\cdot)\rangle\;,
\end{aligned}$$

where (2.6) comes from the confidence bounds in Lemma 1, (2.7) comes from the fact that our algorithm at each round guarantees $\langle w_t^*, \mathrm{ucb}_{x_t}^t\rangle \geq \tau - \hat{\kappa}_{\tau,t}\|w_t^* - w_t\|_M$, and the last inequality is due to Lemma 2. The rest of the proof follows the standard methods used in the Gaussian process literature together with an auxiliary concentration lemma. □

Our analysis relies on the fact that, under Assumption 1, both the true fragility $\kappa_{\tau,t}$ and its estimate $\hat{\kappa}_{\tau,t}$ remain finite. Moreover, observe that if $\tau > \langle w_t, f_{\hat{x}_t}\rangle$ while $P_t \neq P_t^*$, then $\kappa_{\tau,t} = \infty$, in which case the instantaneous regret evaluates to

$$\big(\tau - \kappa_{\tau,t}\Delta(P_t^*, P_t) - \mathbb{E}_{c\sim P_t^*}[f(x_t,c)]\big)^+ = 0\;.$$

With β_T chosen as in Lemma 1, Theorem 2 then yields a regret bound of $\tilde{O}(\gamma_T\sqrt{T})$. The following corollary applies bounds on γ_T from [49] to obtain regret guarantees for specific kernel families.

Corollary 3. *Suppose the kernel $k(z, z')$ is either a Matérn-ν kernel or a squared exponential kernel. Then, on an input domain of dimension d, the robust satisficing regret of RoBOS is bounded by $\tilde{O}\left(T^{\frac{2\nu+3d}{4\nu+2d}}\right)$ and $\tilde{O}(\sqrt{T})$, respectively.*

Next, we analyze the lenient regret of RoBOS. As discussed in the problem formulation, lenient regret is a stricter performance measure, under which even the benchmark action x_t^* can accumulate linear regret. Consequently, our lenient regret bound depends on the magnitude of distribution shift, quantified as $\epsilon_t := \|w_t^* - w_t\|_M$, for $t \in [T]$. We emphasize that the regret analysis of RoBOS differs from that of DRBO in [50], where the assumption is $\|w_t^* - w_t\|_M \leq \epsilon_t'$ for some $\epsilon_t' \geq 0$ and all $t \in [T]$. The key distinction is that DRBO requires the sequence $(\epsilon_t')_{t\in[T]}$ as input, whereas RoBOS operates without access to $(\epsilon_t)_{t\in[T]}$. Note also that $\epsilon_t \leq \epsilon_t'$ always holds.

Before presenting the lenient regret bound, we define $B' := \max_x \|f_x\|_{M^{-1}}$. It is known that $B' \leq B\sqrt{\lambda_{\max}(M^{-1})\, n}$, where $\lambda_{\max}(M^{-1})$ denotes the largest eigenvalue of M^{-1} and B is an upper bound on the RKHS norm of f [50].

Theorem 4. *Fix $\delta \in (0, 1)$. Under Assumption 1, if RoBOS is executed with confidence parameter $\delta/2$ and $\beta_t := \beta_t(\delta/2)$, where β_t is defined in Lemma 1, then with probability at least $1-\delta$, the lenient regret R_T^l of RoBOS admits the following upper bound:*

$$R_T^l \leq 4\beta_T \sqrt{T\left(2\gamma_T + 2\log\left(\frac{12}{\delta}\right)\right)} + B' \sum_{t=1}^{T} \epsilon_t \ .$$

Proof Sketch: When $\langle w_t, \mathrm{ucb}_x^t \rangle \geq \tau$, let $\bar{w}_x^t := \arg\max_{w\in\Delta(\mathcal{C})\setminus\{w_t\}} \frac{\tau - \langle w, \mathrm{ucb}_x^t\rangle}{\|w - w_t\|_M}$ be the context distribution that achieves $\hat{\kappa}_{\tau,t}(x)$. We have

$$\begin{aligned} r_t^l &:= \tau - \langle w_t^*, f_{x_t}\rangle \\ &= \tau - \langle w_t^*, \mathrm{ucb}_{x_t}^t\rangle + \langle w_t^*, \mathrm{ucb}_{x_t}^t - f_{x_t}\rangle \\ &\leq \tau - \langle w_t^*, \mathrm{ucb}_{x_t}^t\rangle + 2\beta_t \langle w_t^*, \sigma_t(x_t, \cdot)\rangle \ , \end{aligned} \tag{2.8}$$

where (2.8) follows from the confidence bounds in Lemma 1.

Consider $w_t^* = w_t$. By Assumption 1 and the selection rule of RoBOS (i.e., $\langle w_t, \mathrm{ucb}_{x_t}^t \rangle \geq \tau$), we obtain

$$r_t^l \leq \tau - \langle w_t, \mathrm{ucb}_{x_t}^t \rangle + 2\beta_t \langle w_t^*, \sigma_t(x_t, \cdot) \rangle \leq 2\beta_t \langle w_t^*, \sigma_t(x_t, \cdot) \rangle \ .$$

Consider $w_t^* \neq w_t$. Continuing from (2.8)

$$\begin{aligned}
r_t^l &\leq \|w_t^* - w_t\|_M \frac{\tau - \langle w_t^*, \mathrm{ucb}_{x_t}^t \rangle}{\|w_t^* - w_t\|_M} + 2\beta_t \langle w_t^*, \sigma_t(x_t, \cdot) \rangle \\
&\leq \|w_t^* - w_t\|_M \hat{\kappa}_{\tau,t}(x_t) + 2\beta_t \langle w_t^*, \sigma_t(x_t, \cdot) \rangle && (2.9) \\
&\leq \|w_t^* - w_t\|_M \hat{\kappa}_{\tau,t}(\hat{x}_t) + 2\beta_t \langle w_t^*, \sigma_t(x_t, \cdot) \rangle && (2.10) \\
&\leq \|w_t^* - w_t\|_M \frac{\langle w_t, f_{\hat{x}_t} \rangle - \langle \bar{w}_{\hat{x}_t}^t, \mathrm{ucb}_{\hat{x}_t}^t \rangle}{\left\|\bar{w}_{\hat{x}_t}^t - w_t\right\|_M} + 2\beta_t \langle w_t^*, \sigma_t(x_t, \cdot) \rangle && (2.11) \\
&= \|w_t^* - w_t\|_M \|f_{\hat{x}_t}\|_{M^{-1}} + 2\beta_t \langle w_t^*, \sigma_t(x_t, \cdot) \rangle && (2.12) \\
&\leq \epsilon_t B' + 2\beta_t \langle w_t^*, \sigma_t(x_t, \cdot) \rangle \ , && (2.13)
\end{aligned}$$

where (2.9) comes from the definition of $\hat{\kappa}_{\tau,t}(x_t)$; (2.10) follows from $x_t = \arg\min_{x \in \mathcal{X}} \hat{\kappa}_{\tau,t}(x)$; (2.11) results from the assumption on τ in Assumption 1; (2.12) utilizes Lemma 1 and the Cauchy-Schwarz inequality; and finally (2.13) follows from the definition of ϵ_t. The remainder of the proof follows similar arguments as in the proof of Theorem 2. □

Theorem 4 establishes that the lenient regret scales as $\tilde{O}(\gamma_T \sqrt{T} + E_T)$, where $E_T := \sum_{t=1}^T \epsilon_t$. A particularly relevant case where E_T is sublinear arises in *data-driven optimization*. In this setting, the reference distribution is fixed, i.e., $P_t^* = P^*$ for all $t \in [T]$, while P_1 is the uniform distribution over the context space. For $t > 1$, P_t corresponds to the empirical distribution of the observed contexts, i.e., $P_t = \sum_{s=1}^{t-1} \delta_{c_s}$, where δ_c denotes the Dirac measure on $c \in \mathcal{C}$ such that $\delta_c(A) = 1$ if $c \in A$ and 0 otherwise. The following corollary is derived through steps analogous to those in the proof of [50, Corollary 4].

Lemma 3. *Consider data-driven optimization model with $\delta \in (0, 1)$. Under Assumption 1, when RoBOS is run with confidence parameter $\delta/3$ with $\beta_t := \beta_t(\delta/3)$, where β_t is defined in Lemma 1, then with a probability of at least $1 - \delta$*

$$R_T^l \leq 4\beta_T \sqrt{T\left(2\gamma_T + 2\log\left(\frac{12}{\delta}\right)\right)} + B'\epsilon_1 + B'2\sqrt{T}\left(2 + \sqrt{2\log\left(\frac{\pi^2 T^2}{2\delta}\right)}\right) \ .$$

2.4 Experiments

We assess the performance of RoBOS in both a synthetic setting and a real-world environment, comparing its lenient and robust satisficing regret against those of the following benchmark algorithms. Our implementation is available at `http://github.com/Bilkent-CYBORG/RoBOS`.

SO: As the representative of standard optimization, we employ a stochastic variant of the GP-UCB algorithm, which at each round selects $x_t = \arg\max_{x \in \mathcal{X}} \mathbb{E}_{c \sim P_t}[\mathrm{ucb}_t(x, c)]$.

DRBO: For the distributionally robust optimization approach, we use the DRBO algorithm from [50]. At each round, DRBO samples $x_t = \arg\max_{x \in \mathcal{X}} \inf_{P \in \mathcal{U}_t} \mathbb{E}_{c \sim P}[\mathrm{ucb}_t(x, c)]$.

WRBO: For the worst-case robust optimization approach, we consider a max–min algorithm, which we denote WRBO. This method maximizes the worst-case reward over the context set. Notably, when the ambiguity set $\mathcal{U}_t$ in DRBO is taken as $\mathcal{P}_0$, i.e., the set of all probability distributions on $\mathcal{C}$, DRBO reduces to WRBO, which at each round selects $x_t = \arg\max_{x \in \mathcal{X}} \min_{c \in \mathcal{C}} \mathrm{ucb}_t(x, c)$.

2.4.1 Synthetic Environment

The objective function is illustrated in Figure 1.1(Left). Both the reference distribution and the true distribution are stationary, with $P_t = \mathcal{N}(2, 1)$ and $P_t^* = \mathcal{N}(0, 25)$. We denote the true MMD distance $\Delta(P_t, P_t^*)$ by ϵ, and run DRBO with ambiguity balls of radii $r = 3\epsilon$, $r = \epsilon$, and $r = \epsilon/3$. Notably, when $r = \epsilon$, DRBO has exact knowledge of the distributional shift. The threshold τ is set to 60% of the maximum value of the objective function, i.e., $\tau = 0.6 \max_{(x,c) \in \mathcal{X} \times \mathcal{C}} f(x, c) \approx 0.65 Z_t$. For the surrogate model, we use an RBF kernel with Automatic Relevance Determination (ARD), assigning lengthscales of 0.2 and 5 to the $\mathcal{X}$ and $\mathcal{C}$ dimensions, respectively. Simulations are performed with observation noise $\eta_t \sim \mathcal{N}(0, 0.02^2)$.

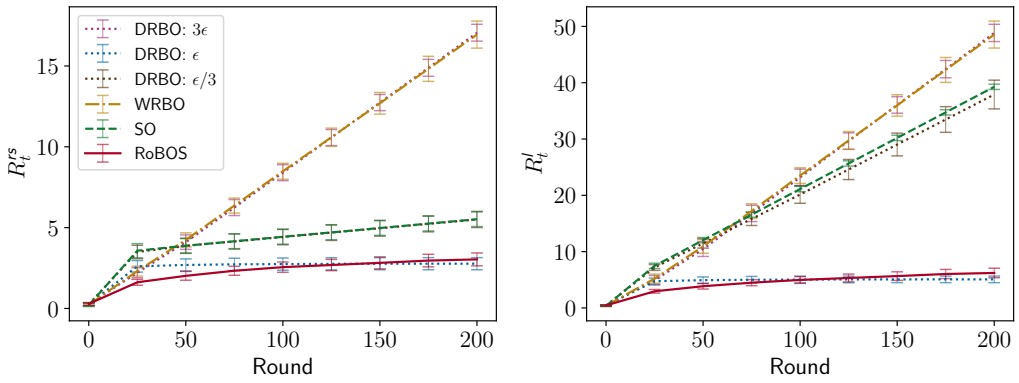


Figure 2.1: Synthetic environment results. Left: robust satisficing regret. Right: lenient regret. Results are averaged over 50 independent runs, with error bars denoting half the standard deviation.

Results for the robust satisficing and lenient regrets are presented in Figure 2.1. In this setup, the RS solution (Figure 1.1) is the only solution that satisfies the threshold τ. Consequently, any algorithm converging to a different solution incurs linear lenient regret, as shown in Figure 2.1(Right). When DRBO is run with an overconfident ambiguity set ($r = \epsilon/3$), it converges to a suboptimal solution (hexagon); when the ambiguity set is underconfident ($r = 3\epsilon$), it converges to another suboptimal solution (cross). Only when the ambiguity radius matches the true shift ($r = \epsilon$) does DRBO converge to the RS solution. The specific solutions reached are illustrated in Figure 1.1.

We also examine the sensitivity of RoBOS to the choice of τ, compared with DRBO's dependence on the ambiguity radius r. Figure 2.2 plots the average rewards of both methods across a range of τ and r values. While RoBOS is designed to satisfice rather than maximize rewards, its performance remains competitive with DRBO over diverse hyperparameter settings. Interestingly, for $\tau \in [0.1, 0.3]$, RoBOS selects the solution marked by a cross in Figure 1.1, reflecting a trade-off: with smaller τ, RoBOS prioritizes robustness guarantees at the expense of maximum reward.

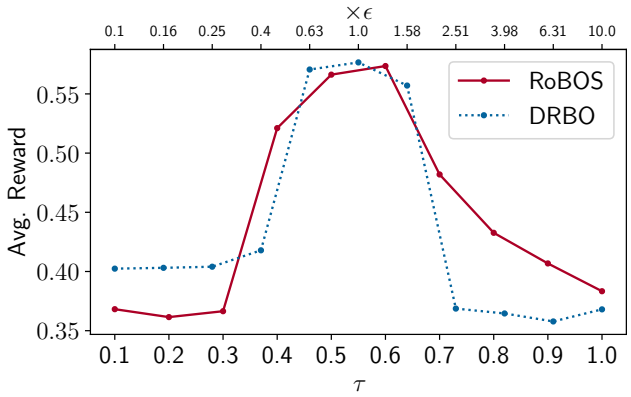


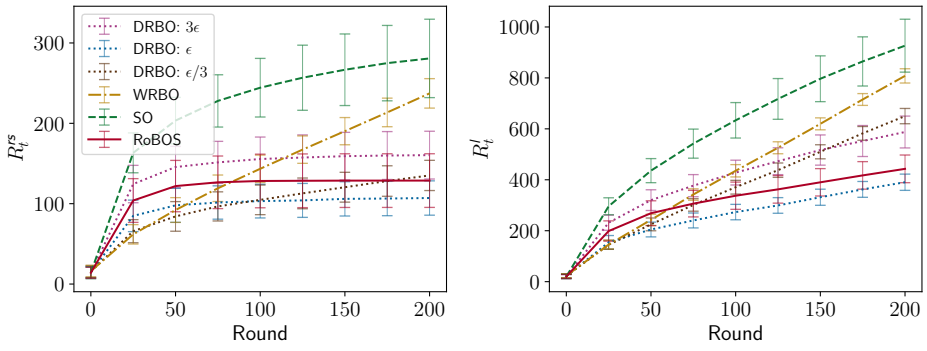
Figure 2.2: Average reward for different τ values in RoBOS and ambiguity radii r in DRBO. τ is chosen linearly between the minimum and maximum function values. For DRBO, r ranges from 0.1ϵ to 10ϵ. Each curve averages 10 independent runs of 200 rounds.

2.4.2 Insulin Dosage for T1DM

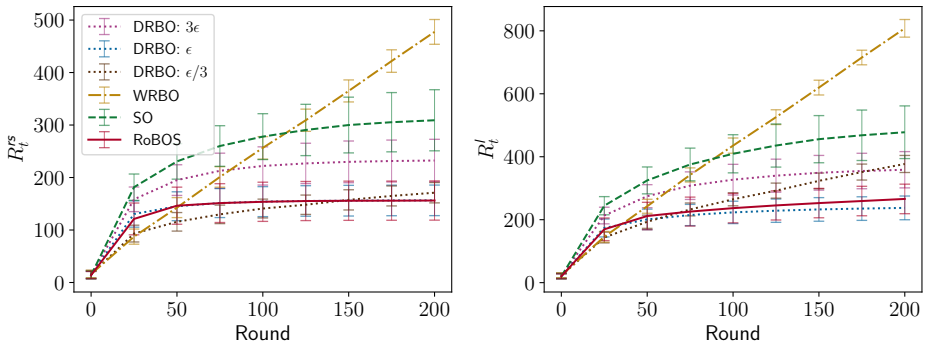
We evaluate our algorithm on the problem of personalized insulin dose allocation for Type 1 Diabetes Mellitus (T1DM) patients using the open-source implementation of the FDA-approved UVA/PADOVA T1DM simulator [51, 25]. The simulator takes as input the patient's fasting blood glucose level, carbohydrate intake during a meal, and the insulin dose administered immediately after the meal, and outputs the blood glucose level 150 minutes later. Following [25], we set the target blood glucose level to $K = 112.5$ mg/dl and define the pseudo-reward as $r(t) = -|o(t) - K|$, where $o(t)$ is the observed blood glucose at round t. A safe blood glucose range is defined as 102.5–122.5 mg/dl, which corresponds to a threshold of $\tau = -10$.

At each round t, the environment samples the true and reference distributions as $P_t \sim \mathcal{N}(\zeta_t, 2.25)$ and $P_t^* \sim \mathcal{N}(\zeta_t + N, 9)$, where $\zeta_t \sim \mathcal{U}(20, 80)$ and N represents the distributional shift. The action set $\mathcal{X}$ is the insulin dose, and the context set $\mathcal{C}$ is the carbohydrate intake. In this setting, the distributional shift corresponds to errors in estimating carbohydrate intake. We consider two cases: $N \sim \mathcal{U}(-6, 6)$ and $N_t \sim \mathcal{U}(-6/\log(t+2),\ 6/\log(t+2))$. Simulations include observation noise $\eta_t \sim \mathcal{N}(0, 1)$. For the GP surrogate, we employ a Matérn-ν kernel with length-scale

10.



(a)



(b)

Figure 2.3: Results for the insulin dose allocation simulation, where ϵ_t denotes the true MMD distance between the true and reference distributions at round t, and the threshold $\tau = -10$ corresponds to a satisficing requirement of staying within 10 mg/dl of the target blood glucose level. In panel (b), the true MMD distance decays as $\epsilon_t \approx \epsilon_0/\log(t)$. The plots display robust satisficing regret (Left) and lenient regret (Right), averaged over 50 independent runs with error bars representing half the standard deviation.

As shown in Figures 2.3a and 2.3b, DRBO outperforms RoBOS when the distributional shift is known exactly. However, when the shift is either underestimated or overestimated, RoBOS achieves superior results. Additional plots of average cumulative rewards are provided in the appendix.

2.5 Conclusion and Future Work

We introduced *robust Bayesian satisficing* (RBS) as a new sequential decision-making framework that provides reliable protection against incalculable distribution shifts. To this end, we proposed the first RBS algorithm, RoBOS, and established information-gain–based bounds on both lenient regret and robust satisficing regret.

Our experiments revealed that when the range of distributional shifts can be tightly estimated with an uncertainty set $\mathcal{U}_t$ centered around the reference distribution, DRBO [50] may outperform RoBOS, particularly in maximizing cumulative reward. This outcome is unsurprising, as more accurate knowledge of the shift naturally enables better performance. However, when the estimated shift fails to capture the true shift, or when the objective is to guarantee a satisfactory performance level rather than to maximize reward, RoBOS emerges as a more reliable alternative.

This foundational study on RBS opens several promising avenues for future research. One direction is to extend RoBOS to continuous action and context spaces, which would necessitate refined regret analyses and efficient methods for computing fragility estimates at each round. Another compelling direction is to develop alternative acquisition strategies tailored to RBS; for example, investigating a Thompson sampling–based strategy as an alternative to the UCB approach explored in this work.

Chapter 3

Robust Bayesian Satisficing in the Adversarial Setting

3.1 Problem Formulation

Let f be an unknown function defined on a domain $\mathcal{X} \subset \mathbb{R}^m$, equipped with a pseudometric $d(\cdot, \cdot) : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$. Suppose that f belongs to a Reproducing Kernel Hilbert Space (RKHS) $\mathcal{H}$ with reproducing kernel $k(\cdot, \cdot) : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$, and that its Hilbert norm is bounded as $\|f\|_{\mathcal{H}} \leq B$ for some $B > 0$. This assumption implies smoothness properties with respect to the kernel metric[1] [52], which in turn enables the construction of GP-based confidence bounds [46].

At each round $t \in [T]$, the learner selects a candidate point $\tilde{x}_t \in \mathcal{X}$. The adversary then applies a perturbation δ_t, and the learner observes the perturbed point $x_t = \tilde{x}_t + \delta_t$ together with a noisy evaluation $y_t = f(x_t) + \eta_t$, where η_t are independent R-subGaussian random variables. If the adversary were unconstrained, it could always choose δ_t such that $\tilde{x}_t + \delta_t = \arg\min_{x \in \mathcal{X}} f(x)$,

[1]Specifically, for the kernel metric we have $|f(x) - f(x')| = \langle f, k(\cdot, x) - k(\cdot, x') \rangle \leq \|f\|_{\mathcal{H}} \, d_k(x, x')$. Note that f is not necessarily Lipschitz continuous under an arbitrary metric on $\mathcal{X}$. Hence, throughout our analysis, we take $d(\cdot, \cdot)$ to be the kernel metric. Moreover, when the feature map $k(\cdot, x)$ is injective, this pseudometric is a true metric.

rendering the problem intractable. To avoid this, we assume that at each round t, the adversary is restricted by an *unknown perturbation budget* ϵ_t, which bounds the admissible perturbations.

Formally, for any $x \in \mathcal{X}$ and perturbation budget ϵ, we define the *ambiguity ball*

$$\Delta_\epsilon(x) := \{\, x' - x : x' \in \mathcal{X},\ d(x, x') \leq \epsilon \,\}. \tag{3.1}$$

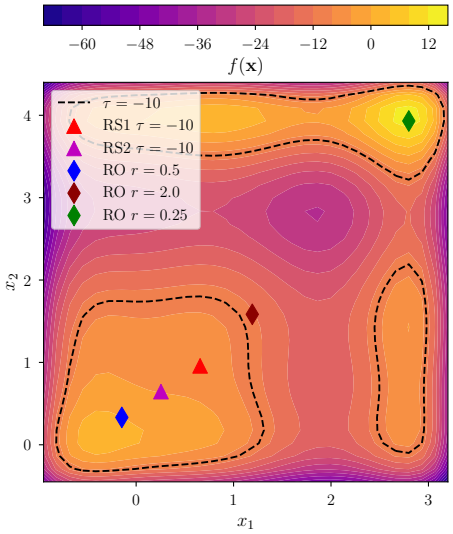
Thus, in response to the learner's choice $\tilde{x}_t$, the adversary selects $\delta_t \in \Delta_{\epsilon_t}(\tilde{x}_t)$.

Remark 1. *In this setting, it is generally impossible to learn the true optimum of the function. To illustrate, consider a simple two-armed bandit problem with* $\mathcal{X} = \{x_1, x_2\}$ *where* $f(x_1) = 0$ *and* $f(x_2) = 1$*. At each round* t*, an adversary with sufficient power can always apply a perturbation* δ_t *that forces the learner to play* $x_t = x_1$*. Consequently, the learner never observes the optimal action and can not identify the optimum.*

3.1.1 Robust Satisficing Formulations

The fundamental distinction between RS and RO lies in their objectives. RO aims to optimize performance under worst-case scenarios, whereas RS focuses on ensuring satisfactory outcomes across a range of possible situations. RO critically depends on the specification of an ambiguity set, often modeled as a ball of radius r centered at a reference value of the uncertain variable. The choice of r strongly influences the resulting solution: overestimating r produces overly conservative solutions that overweight rare worst-case events, while underestimating r yields overly optimistic solutions. Figure 3.1 (left) illustrates the impact of different choices of r, as well as the solutions for the RS formulations which are presented in the next section.

In the present setting, the contingent variables correspond to adversarial perturbations rather than contextual variables, as considered in Chapter 2. Unlike RO, which requires the specification of an ambiguity set, RS depends solely on the satisficing threshold τ, representing the desired outcome. This threshold is



Objective	Formulation
Standard	$\arg\max_x f(x)$
RO	$\arg\max_x \min_{\delta \in \Delta_\epsilon(x)} f(x+\delta)$
RS-1	$\arg\min_x k$ s.t. $f(x+\delta) \geq \tau - kd(x, x+\delta), \forall \delta \in \Delta_\infty(x), \quad k \geq 0$
RS-2	$\arg\max_x \epsilon$ s.t. $f(x+\delta) \geq \tau, \quad \forall \delta \in \Delta_\epsilon(x), \epsilon \geq 0$

Figure 3.1: **(Left)** Illustration of RO and RS solutions. Highlighted contours represent $f(x) = \tau$, i.e., the satisficing threshold. **(Right)** Formulations of different optimization objectives.

both interpretable and practical, as it can be directly chosen by domain experts to reflect problem-specific requirements. We now present two formulations of the RS approach under the assumption that the function f is known.

RS-1: We adapt the formulation of [15], given in equation (1.5), to our adversarial setting. The objective here is to identify $x^{\text{RS-1}} \in \mathcal{X}$ that solves $\kappa_\tau := \min_{x \in \mathcal{X}} \kappa_\tau(x)$, where $\kappa_\tau(x)$ denotes the *fragility* of action $x \in \mathcal{X}$, defined as

$$\kappa_\tau(x) := \min k \quad \text{s.t.} \quad f(x+\delta) \geq \tau - kd(x, x+\delta), \quad \forall \delta \in \Delta_\infty(x) \ . \tag{3.2}$$

This is equivalent to solving

$$\kappa_\tau(x) := \begin{cases} \max_{\delta \in \Delta_\infty(x) \setminus 0} \frac{\tau - f(x+\delta)}{d(x, x+\delta)} & \text{if } f(x) \geq \tau \\ +\infty & \text{if } f(x) < \tau. \end{cases} \tag{3.3}$$

When (3.2) is feasible, the RS-1 action is defined as $x^{\text{RS-1}} = \arg\min_{x \in \mathcal{X}} \kappa_\tau(x)$. Intuitively, the fragility of an action represents the minimum rate at which its performance deteriorates relative to the threshold τ, per unit perturbation, across all possible perturbations. Importantly, unlike the RO formulation, which restricts perturbations to those within a prescribed budget ϵ, the RS-1 formulation (3.2)

accounts for *all* perturbations. In this way, RS provides safety guarantees even when the adversary's perturbation budget is unknown. Figure 3.1 (left) illustrates the RS-1 action $x^{\text{RS-1}}$ in comparison to other optimization objectives.

RS-2: Building on the literature on robust satisficing, we introduce a second formulation, referred to as RS-2, which is closely related to the approach of [45]. The goal of RS-2 is to find $x^{\text{RS-2}} \in \mathcal{X}$ that solves the optimization problem $\epsilon_\tau := \max_{x \in \mathcal{X}} \epsilon_\tau(x)$, where $\epsilon_\tau(x)$ denotes the *critical radius* of action $x \in \mathcal{X}$, defined as

$$\epsilon_\tau(x) := \max_{x' \in \mathcal{X}} d(x, x') \quad \text{s.t.} \quad f(x+\delta) \geq \tau, \quad \forall \delta \in \Delta_{d(x,x')}(x) \ . \tag{3.4}$$

The robust satisficing action is defined as the one with the largest critical radius, $x^{\text{RS-2}} = \arg\max_{x \in \mathcal{X}} \epsilon_\tau(x)$. If $f(x) = \tau$ and x is a local maximum, then $\epsilon_\tau(x) = 0$ since the only perturbation satisfying the condition is $\delta = 0$. If $f(x) < \tau$, we set $\epsilon_\tau(x) = -\infty$, as no ϵ can satisfy the requirement, and by convention the maximum of the empty set is $-\infty$. In contrast to RO, RS-2 does not assume prior knowledge of the adversary's budget ϵ_t; instead, it identifies the action that achieves the threshold τ under the widest range of perturbations. This target-driven perspective embodies the philosophy of robust satisficing. Figure 3.1 (left) illustrates $x^{\text{RS-2}}$ alongside other optimization objectives. If $\tau > f(\hat{x})$, where $\hat{x} := \arg\max_{x \in \mathcal{X}} f(x)$ denotes the optimal action, then no action can satisfy the threshold, even in the absence of perturbations. Therefore, we make the following assumption going forward.

Assumption 5. $\tau \leq \max_{x \in \mathcal{X}} f(x)$.

Assumption 5 guarantees that the optimization problems (3.2) for RS-1 and (3.4) for RS-2 are feasible, thereby ensuring that a robust satisficing action is well-defined in both cases. Since f is unknown, however, the learner may not be able to immediately verify whether a given threshold τ satisfies this assumption. If the learner is flexible in choosing the satisficing threshold, Assumption 5 can be relaxed by dynamically selecting τ at each round such that $\tau < \text{lcb}_t(\hat{x}'_t)$ where $\hat{x}'_t := \arg\max_{x \in \mathcal{X}} \text{lcb}_t(x)$. Here, lcb_t denotes the lower confidence bound of the function. In many real-world applications, domain knowledge allows τ to be chosen

in a way that satisfies Assumption 5. For instance, in hyperparameter tuning where the objective $f(x)$ corresponds to model accuracy, the optimal value $f(\hat{x})$ is naturally bounded above by 1 (perfect accuracy) and below by the accuracy of the best-known baseline model (e.g., 0.97).

3.1.2 GP Regression

We rely on the standard GP regression framework similar to the approach in Section 2.1.5. Given observations $\mathcal{D}_t = \{(x_i, y_i)\}_{i=1}^t$ with Gaussian noise parameter $\lambda > 0$ and kernel k, the posterior mean and variance are

$$\mu_t(x) = \boldsymbol{k}_t(x)^\intercal (\boldsymbol{K}_t + \lambda I)^{-1} \boldsymbol{y}_t, \quad \sigma_t^2(x) = k(x,x) - \boldsymbol{k}_t(x)^\intercal (\boldsymbol{K}_t + \lambda I)^{-1} \boldsymbol{k}_t(x), \tag{3.5}$$

where $\boldsymbol{k}_t(x) = [k(x_i, x)]_{i=1}^t$ and $(\boldsymbol{K}_t)_{ij} = k(x_i, x_j)$. These posterior statistics define confidence intervals for $f(x)$, which in turn allow us to bound the fragility $\kappa_\tau(x)$ and the critical radius $\epsilon_\tau(x)$ in our adversarial setting.

3.1.3 Regret Measures

We adapt the lenient regret and robust satisficing regret introduced in Section 2.1.2 to the adversarial setting as:

$$R_T^l := \sum_{t=1}^T (\tau - f(x_t))^+, \qquad R_T^{rs} := \sum_{t=1}^T (\tau - \kappa_\tau \epsilon_t - f(x_t))^+ \ . \tag{3.6}$$

Lenient regret, introduced by [43], is a natural metric for satisficing, measuring cumulative loss relative to the threshold τ. However, sublinear lenient regret is impossible under unconstrained adversarial attacks, since a sufficiently large perturbation budget can always force rewards below τ. Consequently, any upper bound on lenient regret must either depend explicitly on the perturbation magnitude or restrict the budget ϵ_t and threshold τ. Robust satisficing regret addresses this by comparing the realized reward of x_t with the guaranteed reward of the RS-1 action $x^{\text{RS-1}}$ under perturbations of size at most ϵ_t. By construction, $x^{\text{RS-1}}$ ensures $f(x^{\text{RS-1}}) \geq \tau - \kappa_\tau \epsilon_t$, providing a benchmark for performance guarantees.

3.2 Algorithms and Theoretical Results

Our algorithms employ GP-based confidence bounds, defined as

$$\mathrm{ucb}_t(x) = \mu_t(x) + \beta_t \sigma_t(x), \quad \mathrm{lcb}_t(x) = \mu_t(x) - \beta_t \sigma_t(x),$$

where β_t is a time-dependent exploration parameter. For theoretical guarantees, β_t is chosen according to Lemma 1, based on [48] but argued to be less conservative and more practical in [47]. For reference, we present below the lemma again once more.

Lemma 4. *[47, Theorem 1] Let $\zeta \in (0,1)$, $\overline{\lambda} := \max\{1, \lambda\}$, and*

$$\beta_t(\zeta) := B + R\sqrt{\log\det\left(\boldsymbol{K}_t + \overline{\lambda}\boldsymbol{I}_t\right) + 2\log\left(\tfrac{1}{\zeta}\right)}\ .$$

Then, with probability at least $1-\zeta$, it holds that

$$\mathrm{lcb}_t(x) \le f(x) \le \mathrm{ucb}_t(x), \quad \forall x \in \mathcal{X},\ \forall t \ge 1.$$

We denote the event in Lemma 4 by $\mathcal{E}$ and refer to it as the *good event*. Our analysis also relies on the notion of *maximum information gain* γ_t defined in Definition 2. Following [46], γ_t has become a standard complexity measure in GP-based analysis. Since f is an unknown black-box function, the fragility κ_τ (RS-1) and critical radius ϵ_τ (RS-2) cannot be computed directly. Instead, our algorithms approximate these quantities using upper and lower confidence bounds while balancing exploration and exploitation.

3.2.1 Algorithm for RS-1

To address the RS-1 objective in BO, we introduce the *Adversarially Robust Satisficing-1* (AdveRS-1) algorithm, whose pseudo-code is provided in Algorithm 2. At each round t, AdveRS-1 first computes the upper confidence bound $\mathrm{ucb}_t(x)$ for every action $x \in \mathcal{X}$. Using these estimates within (3.2), it then defines the *optimistic fragility* as:

$$\overline{\kappa}_{\tau,t}(x) := \max_{\delta \in \Delta_\infty(x)\setminus 0} \frac{\tau - \mathrm{ucb}_t(x+\delta)}{d(x, x+\delta)}\ , \tag{3.7}$$

If $\text{ucb}_t(x) \geq \tau$, we set $\overline{\kappa}_{\tau,t}(x)$ as in (3.7); otherwise, we assign $\overline{\kappa}_{\tau,t}(x) := \infty$. We then define $\overline{\kappa}_{\tau,t} := \min_{x \in \mathcal{X}} \overline{\kappa}_{\tau,t}(x)$. The optimistic fragility $\overline{\kappa}_{\tau,t}(x)$ serves as an optimistic proxy for the true fragility $\kappa_\tau(x)$. At each round, AdveRS-1 selects $\tilde{x}_t = \arg\min_{x \in \mathcal{X}} \overline{\kappa}_{\tau,t}(x)$, thereby favoring the action with the lowest optimistic fragility and promoting exploration. We further define the *pessimistic fragility* $\underline{\kappa}_{\tau,t}(x)$, obtained by replacing the ucb_t with lcb_t in (3.7), and set $\underline{\kappa}_{\tau,t} := \min_{x \in \mathcal{X}} \underline{\kappa}_{\tau,t}(x)$. The next lemma establishes the relationship between these estimated fragilities and the true fragility of an action.

Algorithm 2 AdveRS-1

Input: Kernel function k, $\mathcal{X}$, τ, R, B, confidence parameter ζ, time horizon T
Initialize: Set $\mathcal{D}_0 = \emptyset$ (empty dataset), $\mu_0(x) = 0$, $\sigma_0(x) = 1$ for all x
for $t = 0$ to $T - 1$ **do**
  Compute $\text{ucb}_t(x) = \mu_t(x) + \beta_t \sigma_t(x)$, $\forall x \in \mathcal{X}$
  Compute $\overline{\kappa}_{\tau,t}(x)$ as in (3.7), $\forall x \in \mathcal{X}$
  Select point $\tilde{x}_t = \arg\min_{x \in \mathcal{X}} \overline{\kappa}_{\tau,t}(x)$
  Adversary selects perturbation $\delta_t \in \Delta_{\epsilon_t}(\tilde{x}_t)$
  Sample $x_t = \tilde{x}_t + \delta_t$, observe $y_t = f(x_t) + \eta_t$
  Update dataset $\mathcal{D}_t = \mathcal{D}_{t-1} \cup \{(x_t, y_t)\}$
  Update GP posterior as in (3.5)
end for

Lemma 5. *On the event $\mathcal{E}$, the true fragility is bounded by $\overline{\kappa}_{\tau,t}(x) \leq \kappa_\tau(x) \leq \underline{\kappa}_{\tau,t}(x)$, for all $x \in \mathcal{X}$ and all $t \geq 1$.*

Lemma 5 follows from the monotonicity of fragility established in [15], with the proof given in the Appendix. Building on this result and the RS-1 objective (3.2), the following corollary shows how $\underline{\kappa}_{\tau,t}(x)$ and $\overline{\kappa}_{\tau,t}(x)$ characterize the robustness of an action x at round t.

Corollary 6. *Under Assumption 5, with probability at least $1 - \zeta$, the following holds for all $t \in [T]$:*

$$f(\tilde{x}_t + \delta) \geq \tau - \underline{\kappa}_{\tau,t} \cdot d(\tilde{x}_t, \tilde{x}_t + \delta), \quad \forall \delta \in \Delta_\infty(\tilde{x}_t).$$

Conversely, there exists some $\delta' \in \Delta_\infty(\tilde{x}_t)$ *such that*

$$f(\tilde{x}_t + \delta') \;\leq\; \tau - \overline{\kappa}_{\tau,t} \cdot d(\tilde{x}_t, \tilde{x}_t + \delta').$$

3.2.2 Algorithm for RS-2

Algorithm 3 AdveRS-2

5: Compute $\overline{\epsilon}_{\tau,t}(x)$ as in (3.8), $\forall x \in \mathcal{X}$

6: Select point $\tilde{x}_t = \arg\max_{x \in \mathcal{X}} \overline{\epsilon}_{\tau,t}(x)$

For the RS-2 objective, we proceed analogously and introduce the *Adversarially Robust Satisficing-2* (AdveRS-2) algorithm, presented in Algorithm 3. The structure of AdveRS-2 largely mirrors that of Algorithm 2, with only the modified lines shown alongside their original line numbers. Similar to AdveRS-1, the algorithm leverages confidence bounds to form optimistic estimates of the critical radius $\epsilon_\tau(x)$ for each action. In particular, we define the *optimistic critical radius* as

$$\overline{\epsilon}_{\tau,t}(x) := \max_{x' \in \mathcal{X}} d(x, x') \quad \text{s.t.} \quad \mathrm{ucb}_t(x + \delta) \geq \tau, \quad \forall \delta \in \Delta_{d(x,x')}(x), \tag{3.8}$$

with $\overline{\epsilon}_{\tau,t} := \max_{x \in \mathcal{X}} \overline{\epsilon}_{\tau,t}(x)$. AdveRS-2 employs the optimistic critical radius to guide exploration, selecting at each round t the action $\tilde{x}_t = \arg\max_{x \in \mathcal{X}} \overline{\epsilon}_{\tau,t}(x)$, i.e., the one with the largest optimistic estimate. In addition, we define the *pessimistic critical radius* $\underline{\epsilon}_{\tau,t}(x)$, obtained by replacing the ucb with the lcb in (3.8). The following lemma establishes bounds on the true critical radius $\epsilon_\tau(x)$ for any action x.

Lemma 6. *Conditioned on the event* $\mathcal{E}$*, the following inequality holds for all* $x \in \mathcal{X}$ *and* $t \geq 1$*:*

$$\underline{\epsilon}_{\tau,t}(x) \leq \epsilon_\tau(x) \leq \overline{\epsilon}_{\tau,t}(x).$$

Under Assumption 5, with probability at least $1 - \zeta$, $\forall t \in [T]$, it holds that $f(\tilde{x}_t + \delta) \geq \tau$, $\forall \delta \in \Delta_{\underline{\epsilon}_{\tau,t}}(\tilde{x}_t)$ and that $\exists \delta \in \Delta_{\overline{\epsilon}_{\tau,t}}(\tilde{x}_t)$ s.t. $f(\tilde{x}_t + \delta) \leq \tau$. Corollary 3.2.2 provides insight into the attainability of the satisficing threshold τ. In practice, this information can guide the learner in adjusting τ toward either a more ambitious or a more conservative target.

3.2.3 Regret Analysis of Advers-1

For the regret analysis of Algorithm 2, we employ the kernel-induced distance function $d(x, x') = |k(x, x)^2 - 2k(x, x') + k(x', x')^2|$, and make use of the Lipschitz continuity property of f.

Theorem 7. *Let $\zeta \in (0, 1)$, and suppose $f \in \mathcal{H}$ with $\|f\|_{\mathcal{H}} \leq B$. Let $k(\cdot, \cdot)$ denote the reproducing kernel of $\mathcal{H}$, and assume the noise terms η_t are conditionally R-subgaussian. Under Assumption 5, when $d(\cdot, \cdot)$ is the kernel metric and β_t is defined as in Lemma 4, the lenient regret and robust satisficing regret of AdveRS-1 satisfy*

$$R_T^l \leq 4\beta_T \sqrt{T\gamma_T} + B \sum_{t=1}^{T} \epsilon_t \ , \qquad R_T^{rs} \leq 4\beta_T \sqrt{T\gamma_T} \ . \tag{3.9}$$

Proposition 8. *Within the setting of Theorem 7, the linear term $B \sum_{t=1}^{T} \epsilon_t$ in the lenient regret bound is unavoidable. In particular, there exists a problem instance for which this term constitutes a lower bound on the lenient regret of any algorithm.*

3.2.4 Regret Analysis of Advers-2

The RS-1 formulation guarantees performance under all perturbations $\delta \in \Delta_\infty(x^{\text{RS-1}})$, ensuring that

$$f(x^{\text{RS-1}} + \delta) \geq \tau - \kappa_\tau \cdot d(x^{\text{RS-1}}, x^{\text{RS-1}} + \delta).$$

In contrast, the RS-2 formulation does not provide guarantees against every perturbation, but instead offers a stronger condition,

$$f(x^{\text{RS-2}} + \delta) \geq \tau,$$

for perturbations $\delta \in \Delta_{\epsilon_\tau}(x^{\text{RS-2}})$, where ϵ_τ is defined in (3.4). This fundamental difference is reflected in the regret analysis of AdveRS-2. To bound its regret, we impose the following assumption:

Assumption 9. Assume that the perturbation budget of the adversary $\epsilon_t \leq \epsilon_\tau$ for all $t \geq 1$.

Remark 2. *Assumption 9, which is stronger than Assumption 5, ensures that at each round $t \leq T$, there exists an action $x \in \mathcal{X}$ meeting the threshold τ against any adversary-selected perturbation $\delta_t \in \Delta_{\epsilon_t}(x)$. Without Assumption 9, achieving sublinear lenient regret is impossible, as there will always be a perturbation $\delta \in \Delta_{\epsilon_t}(x)$ causing $f(x+\delta) \leq \tau$ for any x.*

Theorem 10. *Under the assumptions of Theorem 7 together with Assumption 9, the lenient regret and the robust satisficing regret of AdveRS-2 satisfy the bound*

$$R_T^l,\ R_T^{rs} \leq 4\beta_T \sqrt{T\gamma_T}\ .$$

Although the lenient regret bound of AdveRS-2 requires the additional assumption on the adversary's perturbation budget, it is also stronger than the bound obtained for AdveRS-1, and it matches the standard regret rate $\tilde{\mathcal{O}}(\gamma_T\sqrt{T})$ achieved by GP-UCB [48]. In particular, when the kernel $k(\cdot,\cdot)$ is chosen as the RBF kernel or the Mátern-ν kernel, the bound specializes to $\tilde{\mathcal{O}}(\sqrt{T})$ and $\tilde{\mathcal{O}}(T^{\frac{2\nu+3m}{4\nu+2m}})$, respectively, where m denotes the input dimension [49].

3.3 Unifying the RS Formulations

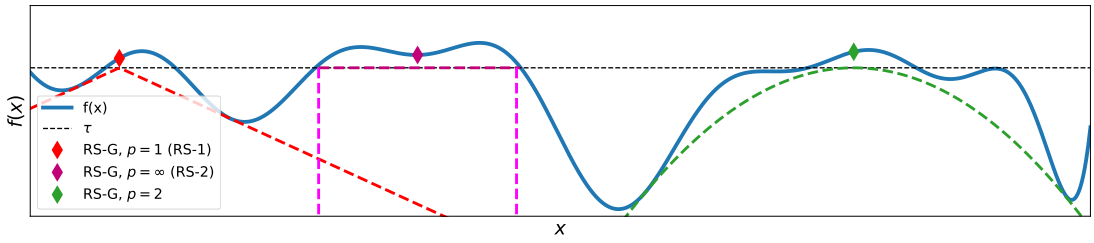


Figure 3.2: Illustration of how different RS formulations balance reward guarantees against perturbation magnitudes during action selection. Each formulation can be viewed as selecting an action (diamonds) by enforcing the widest fragility-cone (dashed lines), which limits performance degradation below the threshold τ under perturbations. The fragility-cone corresponds to the reward guarantee of the solution $x^{\text{RS-G}}$, given by $\tau - [\kappa_{\tau,p} d(x^{\text{RS-G}}, x^{\text{RS-G}} + \delta)]^p$ in (3.10). It characterizes the shape of the constraint that governs how performance deteriorates with increasing perturbation magnitude.

The reward guarantee provided by the RS-1 solution in (3.2) decreases linearly with the magnitude of the perturbation. In contrast, the guarantee of the RS-2 solution remains constant up to a perturbation of size ϵ_τ, after which it vanishes entirely. In practice, however, one may prefer a more flexible form of guarantee. For instance, it may be desirable for the solution to remain close to the threshold τ under small perturbations, while placing less emphasis on larger perturbations that are considered unlikely given the problem structure. Motivated by this, we propose a generalized robust satisficing formulation, RS-General (RS-G). For $p \geq 1$, we define the p-fragility of an action as

$$\kappa_{\tau,p}(x) := \min k \quad \text{s.t.} \quad f(x+\delta) \geq \tau - [k\, d(x, x+\delta)]^p\,, \quad \forall \delta \in \Delta_\infty(x)\ . \tag{3.10}$$

The RS-G action is $x^{\text{RS-G}} := \arg\min_{x\in\mathcal{X}} \kappa_{\tau,p}(x)$.

Proposition 11. *Let $x^{RS\text{-}G}$ denote the solution obtained from the RS-G formulation with power parameter p. Then, in the limit as $p \to \infty$, the RS-G solution converges to the RS-2 solution, i.e.,*

$$\lim_{p\to\infty} x^{RS\text{-}G} = x^{RS\text{-}2}\ .$$

The parameter p regulates the trade-off in the RS-G formulation. At $p = 1$, the reward guarantee decreases linearly with the perturbation magnitude. As p increases, the guarantee becomes stronger for small perturbations while weakening for larger ones. In the limit as $p \to \infty$, the formulation converges to RS-2, yielding robust guarantees within $\Delta_{\epsilon_\tau}(x^{\text{RS-2}})$ but none beyond this set. Figure 3.2 shows how the parameter p affects the solution and the reward guarantee of the RS-G formulation. Algorithm 2 can be directly extended to the RS-G framework by substituting optimistic fragilities with optimistic p-fragilities $\overline{\kappa}_{\tau,p,t}(x)$, computed via (3.10) on $\text{ucb}_t(x)$ instead of $f(x)$. We refer to this generalized algorithm as *AdveRS-G*.Similarly, the RS regret can be extended to the RS-G formulation as

$$R_T^{rs\text{-}g} := \sum_{t=1}^{T} \left(\tau - [\kappa_\tau \epsilon_t]^p - f(x_t)\right)^+ .$$

This quantity measures, at each round, the gap between the obtained reward and the guarantee provided by the $x^{\text{RS-G}}$ solution for a chosen value of p. The following corollary establishes an upper bound on the RS-G regret.

Corollary 12. *Under the assumptions of Theorem 7, AdveRS-G satisfies* $R_T^{rs\text{-}g} \leq 4\beta_T\sqrt{T\gamma_T}$.

3.4 Experiments

In each experiment, we evaluate the performance of AdveRS-1 and AdveRS-2 against the RO baseline [30]. The RO method is tested with an ambiguity ball of radius r that is set equal to, smaller than, and larger than the true perturbation budget ϵ_t at each round, using the Euclidean distance as the perturbation metric. We consider the following attack schemes:

$$\textbf{Random:} \quad \delta_t \sim \mathrm{Uni}(\Delta_{\epsilon_t}(\tilde{x}_t)),$$

$$\textbf{LCB:} \quad \delta_t = \arg\min_{\delta \in \Delta_{\epsilon_t}(\tilde{x}_t)} \mathrm{lcb}_t(\tilde{x}_t),$$

$$\textbf{Worst Case:} \quad \delta_t = \arg\min_{\delta \in \Delta_{\epsilon_t}(\tilde{x}_t)} f(\tilde{x}_t).$$

The results of all experiments are averaged over 100 runs, with error bars representing std/2. Additional experiments and robust satisficing regret plots are provided in the Appendix.

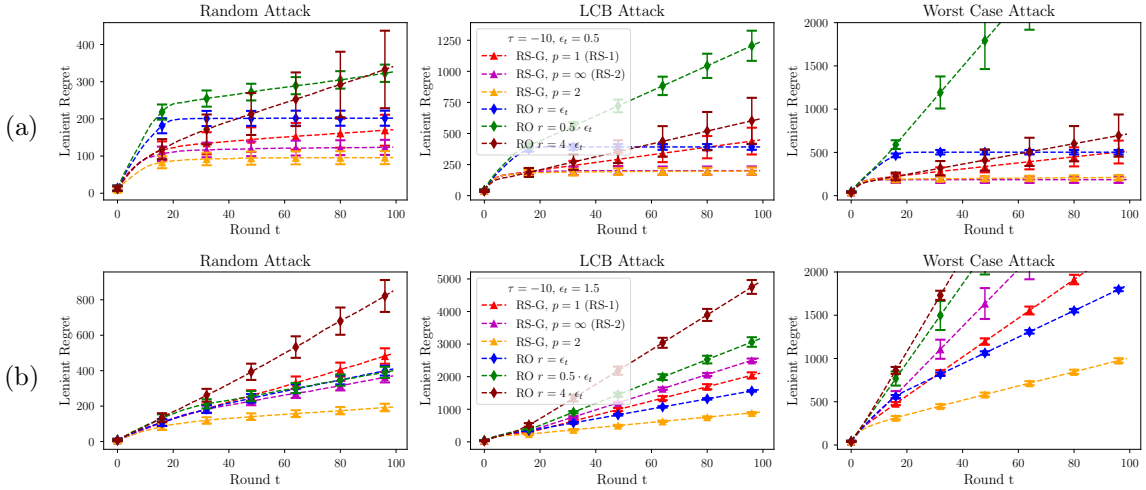


Figure 3.3: Lenient regret results for synthetic experiment shown in two scenarios: (a) satisfying and (b) failing to satisfy Assumption 9.

Synthetic Experiment: In the first experiment, we consider the proof-of-concept function shown in Figure 3.1, which is a modified version of the synthetic function from [30]. We set the threshold $\tau = -10$ and examine two scenarios: (a) $\epsilon_t = 0.5$, where Assumption 9 holds, and (b) $\epsilon_t = 1.5$, where Assumption 9 does *not* hold. As a representative of the RO approach, we evaluate STABLEOPT [30] with radius parameters $r = \epsilon_t$, $r = 4\epsilon_t$, and $r = 0.5\epsilon_t$. The observation noise is drawn from $\eta_t \sim \mathcal{N}(0, 1)$, and the GP kernel is a polynomial kernel trained on 500 samples of the function.

Figure 3.3a demonstrates that STABLEOPT performs poorly when the adversary's perturbation budget is misestimated, resulting in linear regret, whereas AdveRS-2 consistently achieves sublinear lenient regret by satisfying the threshold τ. Although AdveRS-1 does not always reach τ, it still outperforms STABLEOPT when r is misspecified. As shown in Figure 3.3b, when Assumption 9 fails, all algorithms incur linear regret, consistent with Remark 2. Importantly, this figure also highlights that the AdveRS algorithms remain robust in striving toward the target τ even when it is unattainable. Furthermore, RS-G with $p = 2$ achieves the lowest regret, suggesting that tuning the parameter p can yield performance benefits in certain settings.

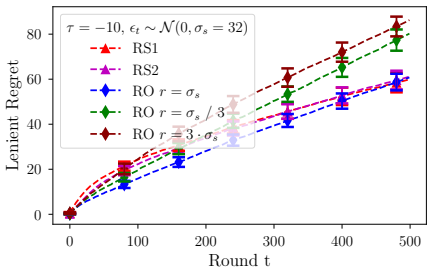


Figure 3.4: Lenient regret of insulin dosage experiment.

Insulin Dosage Experiment: In adversarial contextual GP optimization, the learner first selects an action $x_t \in \mathcal{X}$ after observing a reference context $c_t^{\text{ref}} \in \mathcal{C}$, which is then perturbed by an adversary. The learner subsequently observes the perturbed context $c_t = c_t^{\text{ref}} + \delta_t$ together with the noisy outcome $f(x_t, c_t) + \eta_t$. We apply this framework to the problem of insulin dosage selection for patients with

Type 1 Diabetes Mellitus (T1DM), using the UVA/PADOVA T1DM simulator [51]. T1DM patients require bolus insulin, typically administered before meals [53]. While carbohydrate intake information can significantly improve postprandial blood glucose (PBG) prediction [54], meal announcements often deviate from the true carbohydrate intake, making robustness to contextual shifts essential for reliable diabetes management. Maintaining PBG within a τ-neighbourhood of a target level K serves as the satisficing objective. The action space consists of insulin doses in the range $[0, 15]$, and the context is the carbohydrate intake perturbed by $\delta_t \sim \mathcal{N}(0, \sigma_s)$, with $\sigma_s = 32$, corresponding to approximately two slices of bread. As illustrated in Figure 3.4, although none of the algorithms achieve sublinear regret, both AdveRS-1 and AdveRS-2 consistently outperform STABLEOPT when the radius r is not optimally chosen. STABLEOPT performs best when $r = \sigma_s$, but tuning this parameter is challenging in practice. In contrast, the satisficing threshold τ is more intuitive for clinicians to set, as it aligns with standard clinical practice [53].

Robustness to Worst Case Attack: We evaluate the robustness of the solutions from Figure 3.1 (left) under worst-case adversarial perturbations of varying magnitude. Specifically, we consider $\min_{\delta \in \Delta_\epsilon(x^*)} f(x^* + \delta)$, where x^* denotes the action selected by each objective. To quantify robustness with respect to achieving the threshold τ, we define

$$\text{Area} = \int_\epsilon \max\left(0,\, \tau - \min_{\delta \in \Delta_\epsilon(x^*)} f(x^* + \delta)\right) d\epsilon.$$

Smaller values of this measure indicate stronger robustness across a range of perturbations. Our results reveal a clear performance gap between the proposed methods and the RO baseline: both RS-1 and RS-2 yield solutions that are significantly more robust. Moreover, while RS-2 excels in achieving τ, RS-1 offers greater resilience when τ is unattainable.

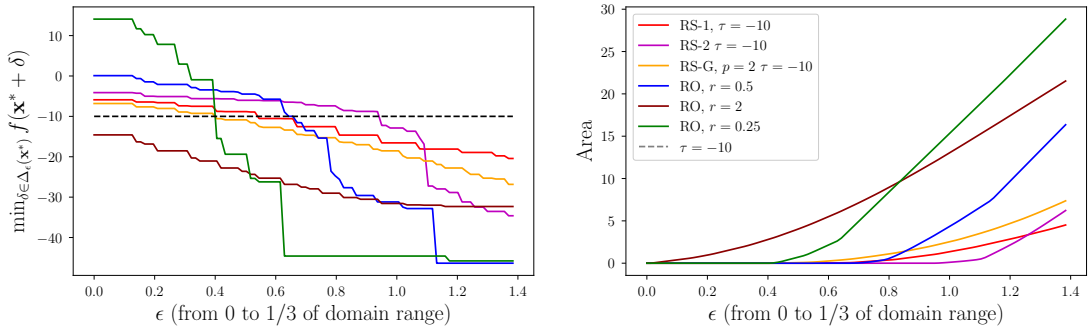


Figure 3.5: **(Left)** Rewards for the points selected by each algorithm, under worst case attack with perturbation budget ϵ. **(Right)** Area under the curves as a function of perturbation magnitudes.

3.5 Conclusion

In this work, we studied GP optimization under adversarial perturbations, where both the perturbation budget and the adversary's strategy are unknown and variable. Standard robust optimization (RO) methods rely on predefined ambiguity sets, which are difficult to specify under such uncertainty. To address this, we proposed a robust satisficing (RS) framework with two formulations, RS-1 and RS-2, designed to guarantee performance relative to a threshold τ. We further unified these formulations under a general RS-G framework. The RS-1 algorithm achieves a lenient regret bound that scales with perturbation magnitude, as well as a sublinear RS regret bound. The RS-2 algorithm provides sublinear bounds for both lenient and RS regret, provided that τ is attainable. Empirical results demonstrate that both algorithms consistently outperform RO methods, especially when the ambiguity set is mis-specified. While promising, the framework has limitations. Choosing an appropriate τ may require domain expertise, and the theoretical guarantees of AdveRS-2 rely on the attainability of τ. Future directions include adaptive approaches to selecting τ, extending the RS formulations to settings such as distribution shifts and supervised learning, and dynamically learning the parameter p in the RS-G formulation during runtime.

Bibliography

[1] P. W. Koh, S. Sagawa, H. Marklund, S. M. Xie, M. Zhang, A. Balsubramani, W. Hu, M. Yasunaga, R. L. Phillips, I. Gao, *et al.*, "Wilds: A benchmark of in-the-wild distribution shifts," in *International conference on machine learning*, pp. 5637–5664, PMLR, 2021.

[2] A. Ben-Tal and A. Nemirovski, "Robust optimization–methodology and applications," *Mathematical programming*, vol. 92, pp. 453–480, 2002.

[3] H. Rahimian and S. Mehrotra, "Distributionally robust optimization: A review," *arXiv preprint arXiv:1908.05659*, 2019.

[4] V. Gupta, "Near-optimal bayesian ambiguity sets for distributionally robust optimization," *Management Science*, vol. 65, no. 9, pp. 4242–4260, 2019.

[5] J. C. Mao, "Survey of capital budgeting: Theory and practice," *Journal of Finance*, pp. 349–360, 1970.

[6] R. Bordley and M. LiCalzi, "Decision analysis using targets instead of utility functions," *Decisions in Economics and Finance*, vol. 23, no. 1, pp. 53–74, 2000.

[7] T. M. Moe, "The new economics of organization," *American Journal of Political Science*, pp. 739–777, 1984.

[8] S. G. Winter, "The satisficing principle in capability learning," *Strategic Management Journal*, vol. 21, no. 10-11, pp. 981–996, 2000.

[9] B. Schwartz, A. Ward, J. Monterosso, S. Lyubomirsky, K. White, and D. R. Lehman, "Maximizing versus satisficing: happiness is a matter of choice," *Journal of Personality and Social Psychology*, vol. 83, no. 5, p. 1178, 2002.

[10] H. Nakayama and Y. Sawaragi, "Satisficing trade-off method for multiobjective programming," in *Interactive Decision Analysis*, pp. 113–122, 1984.

[11] M. A. Goodrich, W. C. Stirling, and R. L. Frost, "A theory of satisficing decisions and control," *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, vol. 28, no. 6, pp. 763–779, 1998.

[12] B. Schwartz, Y. Ben-Haim, and C. Dacso, "What makes a good decision? robust satisficing as a normative standard of rational decision making," *Journal for the Theory of Social Behaviour*, vol. 41, no. 2, pp. 209–227, 2011.

[13] A. L. Mogensen and D. Thorstad, "Tough enough? robust satisficing as a decision norm for long-term policy analysis," *Synthese*, vol. 200, no. 1, p. 36, 2022.

[14] L. J. Ball, L. Maskill, and T. C. Ormerod, "Satisficing in engineering design: Causes, consequences and implications for design support," *Automation in construction*, vol. 7, no. 2-3, pp. 213–227, 1998.

[15] D. Z. Long, M. Sim, and M. Zhou, "Robust satisficing," *Operations Research*, vol. 71, no. 1, pp. 61–82, 2023.

[16] R. Garnett, *Bayesian optimization*. Cambridge University Press, 2023.

[17] J. Wu, X.-Y. Chen, H. Zhang, L.-D. Xiong, H. Lei, and S.-H. Deng, "Hyperparameter optimization for machine learning models based on bayesian optimization," *Journal of Electronic Science and Technology*, vol. 17, no. 1, pp. 26–40, 2019.

[18] R. Lam, M. Poloczek, P. Frazier, and K. E. Willcox, "Advances in bayesian optimization with applications in aerospace engineering," in *2018 AIAA Non-Deterministic Approaches Conference*, p. 1656, 2018.

[19] M. Imani and S. F. Ghoreishi, "Bayesian optimization objective-based experimental design," in *2020 American control conference (ACC)*, pp. 3405–3411, IEEE, 2020.

[20] A. Alaa and M. van der Schaar, "Autoprognosis: Automated clinical prognostic modeling via bayesian optimization with structured kernel learning," in *International conference on machine learning*, pp. 139–148, PMLR, 2018.

[21] C. E. Rasmussen and H. Nickisch, "Gaussian processes for machine learning (gpml) toolbox," *The Journal of Machine Learning Research*, vol. 11, pp. 3011–3015, 2010.

[22] N. Srinivas, A. Krause, S. M. Kakade, and M. W. Seeger, "Information-theoretic regret bounds for Gaussian process optimization in the bandit setting," *IEEE Transactions on Information Theory*, vol. 58, no. 5, pp. 3250–3265, 2012.

[23] J. Snoek, H. Larochelle, and R. P. Adams, "Practical Bayesian optimization of machine learning algorithms," *Advances in Neural Information Processing Systems*, vol. 25, 2012.

[24] J. Bergstra, R. Bardenet, Y. Bengio, and B. Kégl, "Algorithms for hyper-parameter optimization," in *Advances in Neural Information Processing Systems*, vol. 24, 2011.

[25] I. Demirel, A. A. Celik, and C. Tekin, "ESCADA: Efficient safety and context aware dose allocation for precision medicine," in *Advances in Neural Information Processing Systems*, vol. 35, pp. 27441–27454, 2022.

[26] A. Krause and C. Ong, "Contextual Gaussian process bandit optimization," *Advances in Neural Information Processing Systems*, vol. 24, 2011.

[27] R. Oliveira, L. Ott, and F. Ramos, "Bayesian optimisation under uncertain inputs," in *The 22nd international conference on artificial intelligence and statistics*, pp. 1177–1184, PMLR, 2019.

[28] J. Kirschner and A. Krause, "Stochastic bandits with context distributions," in *Advances in Neural Information Processing Systems*, vol. 32, 2019.

[29] R. Martinez-Cantin, K. Tee, and M. McCourt, "Practical bayesian optimization in the presence of outliers," in *International conference on artificial intelligence and statistics*, pp. 1722–1731, PMLR, 2018.

[30] I. Bogunovic, J. Scarlett, S. Jegelka, and V. Cevher, "Adversarially robust optimization with gaussian processes," *Advances in neural information processing systems*, vol. 31, 2018.

[31] L. Fröhlich, E. Klenske, J. Vinogradska, C. Daniel, and M. Zeilinger, "Noisy-input entropy search for efficient robust bayesian optimization," in *International Conference on Artificial Intelligence and Statistics*, pp. 2262–2272, PMLR, 2020.

[32] I. Bogunovic, A. Krause, and J. Scarlett, "Corruption-tolerant gaussian process bandit optimization," in *International Conference on Artificial Intelligence and Statistics*, pp. 1071–1081, PMLR, 2020.

[33] I. Bogunovic, Z. Li, A. Krause, and J. Scarlett, "A robust phased elimination algorithm for corruption-tolerant gaussian process bandits," *Advances in Neural Information Processing Systems*, vol. 35, pp. 23951–23964, 2022.

[34] J. Kirschner, I. Bogunovic, S. Jegelka, and A. Krause, "Distributionally robust bayesian optimization," in *International Conference on Artificial Intelligence and Statistics*, pp. 2174–2184, PMLR, 2020.

[35] S. Cakmak, R. Astudillo Marban, P. Frazier, and E. Zhou, "Bayesian optimization of risk measures," *Advances in Neural Information Processing Systems*, vol. 33, pp. 20130–20141, 2020.

[36] Q. P. Nguyen, Z. Dai, B. K. H. Low, and P. Jaillet, "Value-at-risk optimization with gaussian processes," in *International Conference on Machine Learning*, pp. 8063–8072, PMLR, 2021.

[37] Q. P. Nguyen, Z. Dai, B. K. H. Low, and P. Jaillet, "Optimizing conditional value-at-risk of black-box functions," *Advances in Neural Information Processing Systems*, vol. 34, pp. 4170–4180, 2021.

[38] S. Iwazaki, Y. Inatsu, and I. Takeuchi, "Mean-variance analysis in bayesian optimization under uncertainty," in *International Conference on Artificial Intelligence and Statistics*, pp. 973–981, PMLR, 2021.

[39] H. A. Simon, "A behavioral model of rational choice," *The Quarterly Journal of Economics*, vol. 69, no. 1, pp. 99–118, 1955.

[40] P. Reverdy, V. Srivastava, and N. E. Leonard, "Satisficing in multi-armed bandit problems," *IEEE Transactions on Automatic Control*, vol. 62, no. 8, pp. 3788–3803, 2016.

[41] A. Hüyük and C. Tekin, "Multi-objective multi-armed bandit with lexicographically ordered and satisficing objectives," *Machine Learning*, vol. 110, no. 6, pp. 1233–1266, 2021.

[42] D. Russo and B. Van Roy, "Satisficing in time-sensitive bandit learning," *Mathematics of Operations Research*, vol. 47, no. 4, pp. 2815–2839, 2022.

[43] N. Merlis and S. Mannor, "Lenient regret for multi-armed bandits," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 35, pp. 8950–8957, 2021.

[44] X. Cai, S. Gomes, and J. Scarlett, "Lenient regret and good-action identification in gaussian process bandits," in *International Conference on Machine Learning*, pp. 1183–1192, PMLR, 2021.

[45] M. Zacksenhouse, S. Nemets, M. A. Lebedev, and M. A. Nicolelis, "Robust satisficing linear regression: Performance/robustness trade-off and consistency criterion," *Mechanical systems and signal processing*, vol. 23, no. 6, pp. 1954–1964, 2009.

[46] N. Srinivas, A. Krause, S. M. Kakade, and M. Seeger, "Gaussian process optimization in the bandit setting: No regret and experimental design," *arXiv preprint arXiv:0912.3995*, 2009.

[47] C. Fiedler, C. W. Scherer, and S. Trimpe, "Practical and rigorous uncertainty bounds for gaussian process regression," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 35, pp. 7439–7447, 2021.

[48] S. R. Chowdhury and A. Gopalan, "On kernelized multi-armed bandits," in *International Conference on Machine Learning*, pp. 844–853, PMLR, 2017.

[49] S. Vakili, K. Khezeli, and V. Picheny, "On information gain and regret bounds in gaussian process bandits," in *International Conference on Artificial Intelligence and Statistics*, pp. 82–90, PMLR, 2021.

[50] J. Kirschner, I. Bogunovic, S. Jegelka, and A. Krause, "Distributionally robust Bayesian optimization," vol. 108, pp. 2174–2184, 2020.

[51] B. P. Kovatchev, M. Breton, C. Dalla Man, and C. Cobelli, "In silico preclinical trials: a proof of concept in closed-loop control of type 1 diabetes," 2009.

[52] I. Steinwart and A. Christmann, *Support Vector Machines*. Information Science and Statistics, Springer New York, 2008.

[53] T. Danne, M. Phillip, B. A. Buckingham, P. Jarosz-Chobot, B. Saboo, T. Urakami, T. Battelino, R. Hanas, and E. Codner, "Ispad clinical practice consensus guidelines 2018: Insulin treatment in children and adolescents with diabetes," *Pediatric diabetes*, vol. 19, pp. 115–135, 2018.

[54] C. Zecchin, A. Facchinetti, G. Sparacino, and C. Cobelli, "How much is short-term glucose prediction in type 1 diabetes improved by adding insulin delivery and meal content information to cgm data? a proof-of-concept study," *Journal of diabetes science and technology*, vol. 10, no. 5, pp. 1149–1160, 2016.

[55] K. Muandet, K. Fukumizu, B. Sriperumbudur, B. Schölkopf, *et al.*, "Kernel mean embedding of distributions: A review and beyond," *Foundations and Trends® in Machine Learning*, vol. 10, no. 1-2, pp. 1–141, 2017.

[56] A. Saday, Y. C. Yıldırım, and C. Tekin, "Robust bayesian satisficing," *Advances in Neural Information Processing Systems*, vol. 36, 2024.

[57] A. Marco, P. Hennig, J. Bohg, S. Schaal, and S. Trimpe, "Automatic lqr tuning based on gaussian process global optimization," in *2016 IEEE international conference on robotics and automation (ICRA)*, pp. 270–277, IEEE, 2016.

[58] A. Marco, F. Berkenkamp, P. Hennig, A. P. Schoellig, A. Krause, S. Schaal, and S. Trimpe, "Virtual vs. real: Trading off simulations and physical experiments in reinforcement learning with bayesian optimization," in *2017 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 1557–1563, IEEE, 2017.

Appendix A

Appendix for Chapter 2

A.1 Table of notations

A.2 Additional experimental results

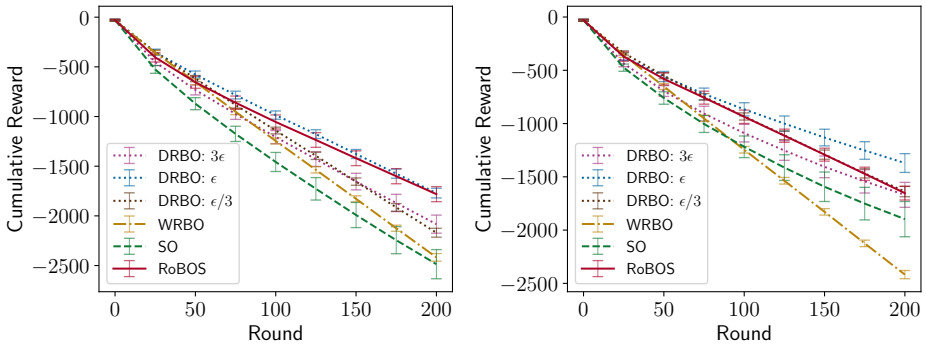


Figure A.1: Cumulative rewards of the insulin dosage for T1DM experiments given in the main paper. Left and right plots are continuations of Figures 2.3a and 2.3b, respectively. Plots show average of 50 independent runs with error bars corresponding to standard deviation divided by 2. The pseudo-reward function is defined as $r(t) = -|o(t) - K|$.

Table A.1: Table of notations

Notation	Description
τ	aspiration level (threshold)
$\|w\|_M$	$\sqrt{w^\mathsf{T} M w}$ MMD measure with kernel matrix M
$\mathcal{X}$	Action set
$\mathcal{C}$	Context set
w_t^*	True distribution at round t
w_t	Reference distribution at round t
f_x	$[f(x, c_1), \ldots, f(x, c_n)]^\mathsf{T}$
ucb_x^t	upper confidence bound of f_x given by $[\mathrm{ucb}_t(x, c_1), \ldots, \mathrm{ucb}_t(x, c_n)]^\mathsf{T}$
$\kappa_{\tau,t}(x)$	$\begin{cases} \max\left\{\max_{w \in \Delta(\mathcal{C}) \setminus \{w_t\}} \frac{\tau - \langle w, f_x \rangle}{\|w - w_t\|_M}, 0\right\} & \text{if } \langle w_t, f_x \rangle \geq \tau \\ +\infty & \text{if } \langle w_t, f_x \rangle < \tau \end{cases}$
$\hat{\kappa}_{\tau,t}(x)$	$\begin{cases} \max_{w \in \Delta(\mathcal{C}) \setminus \{w_t\}} \frac{\tau - \langle w, \mathrm{ucb}_x^t \rangle}{\|w - w_t\|_M} & \text{if } \langle w_t, \mathrm{ucb}_x^t \rangle \geq \tau \\ +\infty & \text{if } \langle w_t, \mathrm{ucb}_x^t \rangle < \tau \ . \end{cases}$
$\hat{x}_t$	$\arg\max_{x \in \mathcal{X}} \langle w_t, f_x \rangle$
x_t^*	$\arg\min_{x \in \mathcal{X}} \kappa_{\tau,t}(x)$ when $\langle w_t, f_{\hat{x}_t} \rangle \geq \tau$
x_t	$\arg\min_{x \in \mathcal{X}} \hat{\kappa}_{\tau,t}(x)$ when $\langle w_t, f_{\hat{x}_t} \rangle \geq \tau$
$\kappa_{\tau,t}$	$\kappa_{\tau,t}(x_t^*)$ when $\langle w_t, f_{\hat{x}_t} \rangle \geq \tau$
$\hat{\kappa}_{\tau,t}$	$\hat{\kappa}_{\tau,t}(x_t)$ when $\langle w_t, f_{\hat{x}_t} \rangle \geq \tau$
$\bar{\bar{w}}_x^t$	$\arg\max_{w \in \Delta(\mathcal{C}) \setminus \{w_t\}} \frac{\tau - \langle w, f_x \rangle}{\|w - w_t\|_M}$ when $\langle w_t, f_x \rangle \geq \tau$
$\bar{w}_x^t$	$\arg\max_{w \in \Delta(\mathcal{C}) \setminus \{w_t\}} \frac{\tau - \langle w, \mathrm{ucb}_x^t \rangle}{\|w - w_t\|_M}$ when $\langle w_t, \mathrm{ucb}_x^t \rangle \geq \tau$

A.3 Proofs

A.3.1 Proof of Lemma 2

Assume that the confidence intervals in Lemma 1 hold. When $\hat{\kappa}_{\tau,t}(x) = \infty$, we also have $\kappa_{\tau,t}(x) = \infty$. When $\hat{\kappa}_{\tau,t}(x) < \infty$ and $\kappa_{\tau,t}(x) = \infty$, the inequality holds.

Recall that when $\langle w_t, \mathrm{ucb}_x^t \rangle \geq \tau$, $\bar{w}_x^t := \arg\max_{w \in \Delta(\mathcal{C}) \setminus \{w_t\}} \frac{\tau - \langle w, \mathrm{ucb}_x^t \rangle}{\|w - w_t\|_M}$ represents the context distribution that achieves $\hat{\kappa}_{\tau,t}(x)$. When $\langle w_t, f_x \rangle \geq \tau$, let $\bar{\bar{w}}_x^t := \arg\max_{w \in \Delta(\mathcal{C}) \setminus \{w_t\}} \frac{\tau - \langle w, f_x \rangle}{\|w - w_t\|_M}$ be the context distribution that achieves $\kappa_{\tau,t}(x)$. When

$\hat{\kappa}_{\tau,t}(x) < \infty$ and $\kappa_{\tau,t}(x) < \infty$, we have

$$\begin{aligned}
\hat{\kappa}_{\tau,t}(x) - \kappa_{\tau,t}(x) &\leq \frac{\tau - \langle \bar{w}_x^t, \mathrm{ucb}_x^t \rangle}{\|\bar{w}_x^t - w_t\|_M} - \frac{\tau - \langle \bar{\bar{w}}_x^t, f_x \rangle}{\|\bar{\bar{w}}_x^t - w_t\|_M} \\
&\leq \frac{\tau - \langle \bar{w}_x^t, \mathrm{ucb}_x^t \rangle}{\|\bar{w}_x^t - w_t\|_M} - \frac{\tau - \langle \bar{w}_x^t, f_x \rangle}{\|\bar{w}_x^t - w_t\|_M} \\
&= \frac{\langle \bar{w}_x^t, f_x - \mathrm{ucb}_x^t \rangle}{\|\bar{w}_x^t - w_t\|_M} \\
&\leq 0 \ .
\end{aligned}$$

A.3.2 An Auxiliary Concentration Lemma

Lemma 7. *([50, Lemma 7]) Let $S_t \geq 0$ be a non-negative stochastic process with filtration $\mathcal{F}_t$, and define $m_t = \mathbb{E}[S_t | \mathcal{F}_{t-1}]$. Further assume that $S_t \leq B$ for $B \geq 1$. Then for any $T \geq 1$, with probability at least $1 - \delta$ it holds that*

$$\begin{aligned}
\sum_{t=1}^{T} m_t &\leq 2 \sum_{t=1}^{T} S_t + 4B \log \frac{1}{\delta} + 8B \log(4B) + 1 \\
&\leq 2 \sum_{t=1}^{T} S_t + 8B \log \frac{6B}{\delta} \ .
\end{aligned}$$

A.3.3 Proof of Theorem 2

The robust satisficing regret is upper bounded by the sum of instantaneous regrets $r_t^{rs} := \tau - \kappa_{\tau,t} \|w_t^* - w_t\|_M - \langle w_t^*, f_{x_t} \rangle$ as follows:

$$\begin{aligned}
R_T^{rs} &= \sum_{t=1}^{T} \left(\tau - \kappa_{\tau,t} \Delta(P_t^*, P_t) - \mathbb{E}_{c \sim P_t^*}[f(x_t, c)] \right)^+ \\
&= \sum_{t=1}^{T} (\tau - \kappa_{\tau,t} \|w_t^* - w_t\|_M - \langle w_t^*, f_{x_t} \rangle)^+ \\
&= \sum_{t=1}^{T} r_t^{rs} \mathbb{I}(r_t^{rs} \geq 0) \ . \qquad \text{(A.1)}
\end{aligned}$$

Assume that the confidence intervals in Lemma 1 hold (an event that happens with probability at least $1-\delta/2$). For the instantaneous regret we have

$$\begin{aligned}
r_t^{rs} &= \tau - \kappa_{\tau,t}\|w_t^* - w_t\|_M - \langle w_t^*, \mathrm{ucb}_{x_t}^t\rangle + \langle w_t^*, \mathrm{ucb}_{x_t}^t - f_{x_t}\rangle \\
&\leq \tau - \kappa_{\tau,t}\|w_t^* - w_t\|_M - \langle w_t^*, \mathrm{ucb}_{x_t}^t\rangle + 2\beta_t\langle w_t^*, \sigma_t(x_t,\cdot)\rangle \qquad \text{(A.2)} \\
&\leq \tau - \kappa_{\tau,t}\|w_t^* - w_t\|_M - (\tau - \hat{\kappa}_{\tau,t}\|w_t^* - w_t\|_M) + 2\beta_t\langle w_t^*, \sigma_t(x_t,\cdot)\rangle \qquad \text{(A.3)} \\
&= \|w_t^* - w_t\|_M(\hat{\kappa}_{\tau,t} - \kappa_{\tau,t}) + 2\beta_t\langle w_t^*, \sigma_t(x_t,\cdot)\rangle \\
&\leq 2\beta_t\langle w_t^*, \sigma_t(x_t,\cdot)\rangle \ . \qquad \text{(A.4)}
\end{aligned}$$

In the derivation above, (A.2) follows from the confidence bounds given in Lemma 1; (A.3) follows from $\langle w_t^*, \mathrm{ucb}_{x_t}^t\rangle \geq \tau - \hat{\kappa}_{\tau,t}(x_t)||w_t^* - w_t||_M$ and $\hat{\kappa}_{\tau,t} = \hat{\kappa}_{\tau,t}(x_t)$; (A.4) is due to Lemma 2.

From this point on, we bound the robust satisficing regret by following standard steps for bounding regret of GP bandits (see e.g., [50]). First, by plugging the upper bound on r_t^{rs} obtained in (A.4) to (A.1), and using monotonicity of β_t, we obtain

$$\begin{aligned}
R_T^{rs} &\leq 2\beta_T \sum_{t=1}^{T} \langle w_t^*, \sigma_t(x_t,\cdot)\rangle \qquad \text{(A.5)} \\
&\leq 2\beta_T \sqrt{T\sum_{t=1}^{T} \langle w_t^*, \sigma_t(x_t,\cdot)\rangle^2} \qquad \text{(A.6)} \\
&\leq 2\beta_T \sqrt{T\sum_{t=1}^{T} \langle w_t^*, \sigma_t(x_t,\cdot)^2\rangle} \ , \qquad \text{(A.7)}
\end{aligned}$$

where (A.6) uses the Cauchy-Schwarz inequality and (A.7) uses the Jensen inequality.

To complete the proof we use Lemma 7 to relate the expectation of the posterior variance in (A.7) to the posterior variance of the observations. Namely, we have with probability at least $1-\delta/2$

$$\sum_{t=1}^{T} \langle w_t^*, \sigma_t(x_t,\cdot)^2\rangle \leq 2\sum_{t=1}^{T} \sigma_t(x_t,c_t)^2 + 8\log\frac{12}{\delta} \ . \qquad \text{(A.8)}$$

Next, to relate the sum of variances that appears in (A.8) to the maximum information gain. We use $x \leq 2\log(1+x)$ for all $x \in [0,1]$ to obtain

$$\begin{aligned}\sum_{t=1}^{T} \sigma_t(x_t, c_t)^2 &\leq \sum_{t=1}^{T} 2\log\big(1+\sigma_t(x_t,c_t)^2\big) \\ &\leq 4\gamma_T \ , \end{aligned} \tag{A.9}$$

where (B.12) follows from [48, Lemma 3].

Finally, to bound the robust satisficing regret, we plug the upper bound in (B.12) to the r.h.s. of (A.8), and use it to upper bound (A.7) as follows

$$R_T^{rs} \leq 4\beta_T \sqrt{T\left(2\gamma_T + 2\log\left(\frac{12}{\delta}\right)\right)} \ .$$

Let A and B be the events that Lemma 1 and Lemma 7 hold. Setting the confidence of each event to $1-\delta/2$, we can bound from below the probability that both events hold

$$P(A \cap B) = 1 - P(\bar{A} \cup \bar{B}) \geq 1 - P(\bar{A}) - P(\bar{B}) = 1 - \delta \ ,$$

completing the proof of Theorem 2.

A.3.4 Proof of Theorem 4

Define the instantaneous regret as $r_t^l := \tau - \langle w_t^*, f_{x_t} \rangle$. We have

$$R_T^l := \sum_{t=1}^{T} \big(\tau - \mathbb{E}_{c\sim P_t^*}[f(x_t,c)]\big)^+ = \sum_{t=1}^{T} r_t^l \mathbb{I}(r_t^l \geq 0) \ . \tag{A.10}$$

Assume that the confidence intervals in Lemma 1 hold (an event that happens with probability at least $1-\delta/2$). Then, for the instantaneous regret we have

$$\begin{aligned} r_t^l &= \tau - \langle w_t^*, \mathrm{ucb}_{x_t}^t \rangle + \langle w_t^*, \mathrm{ucb}_{x_t}^t - f_{x_t} \rangle \\ &\leq \tau - \langle w_t^*, \mathrm{ucb}_{x_t}^t \rangle + 2\beta_t \langle w_t^*, \sigma_t(x_t, \cdot) \rangle \ . \end{aligned} \tag{A.11}$$

When, $w_t^* = w_t$, by Assumption 1 and the selection rule of RoBOS (i.e., $\langle w_t, \text{ucb}_{x_t}^t \rangle \geq \tau$), we obtain

$$r_t^l \leq \tau - \langle w_t, \text{ucb}_{x_t}^t \rangle + 2\beta_t \langle w_t^*, \sigma_t(x_t, \cdot) \rangle \leq 2\beta_t \langle w_t^*, \sigma_t(x_t, \cdot) \rangle \ .$$

When $w_t^* \neq w_t$, continuing from (A.11), we obtain

$$\begin{aligned}
r_t^l &\leq \|w_t^* - w_t\|_M \frac{\tau - \langle w_t^*, \text{ucb}_{x_t}^t \rangle}{\|w_t^* - w_t\|_M} + 2\beta_t \langle w_t^*, \sigma_t(x_t, \cdot) \rangle \\
&\leq \|w_t^* - w_t\|_M \frac{\tau - \langle \bar{w}_{x_t}^t, \text{ucb}_{x_t}^t \rangle}{\left\|\bar{w}_{x_t}^t - w_t\right\|_M} + 2\beta_t \langle w_t^*, \sigma_t(x_t, \cdot) \rangle && \text{(A.12)} \\
&\leq \|w_t^* - w_t\|_M \frac{\tau - \langle \bar{w}_{\hat{x}_t}^t, \text{ucb}_{\hat{x}_t}^t \rangle}{\left\|\bar{w}_{\hat{x}_t}^t - w_t\right\|_M} + 2\beta_t \langle w_t^*, \sigma_t(x_t, \cdot) \rangle && \text{(A.13)} \\
&\leq \|w_t^* - w_t\|_M \frac{\langle w_t, f_{\hat{x}_t} \rangle - \langle \bar{w}_{\hat{x}_t}^t, \text{ucb}_{\hat{x}_t}^t \rangle}{\left\|\bar{w}_{\hat{x}_t}^t - w_t\right\|_M} + 2\beta_t \langle w_t^*, \sigma_t(x_t, \cdot) \rangle && \text{(A.14)} \\
&\leq \|w_t^* - w_t\|_M \frac{\langle w_t - \bar{w}_{\hat{x}_t}^t, f_{\hat{x}_t} \rangle}{\left\|\bar{w}_{\hat{x}_t}^t - w_t\right\|_M} + 2\beta_t \langle w_t^*, \sigma_t(x_t, \cdot) \rangle && \text{(A.15)} \\
&\leq \|w_t^* - w_t\|_M \frac{\left\|\bar{w}_{\hat{x}_t}^t - w_t\right\|_M \|f_{\hat{x}_t}\|_{M^{-1}}}{\left\|\bar{w}_{\hat{x}_t}^t - w_t\right\|_M} + 2\beta_t \langle w_t^*, \sigma_t(x_t, \cdot) \rangle && \text{(A.16)} \\
&= \|w_t^* - w_t\|_M \|f_{\hat{x}_t}\|_{M^{-1}} + 2\beta_t \langle w_t^*, \sigma_t(x_t, \cdot) \rangle \\
&\leq 2\beta_t \langle w_t^*, \sigma_t(x_t, \cdot) \rangle + \epsilon_t B' \ . && \text{(A.17)}
\end{aligned}$$

Since the final bound is non-negative, it can be used to bound all terms $r_t^l \mathbb{I}(r_t^l \geq 0)$ in R_T^l. In the derivation above, (A.11) follows from Lemma 1; (A.12) uses the fact that $\bar{w}_{x_t}^t$ is the distribution that achieves $\hat{\kappa}_{\tau,t}(x_t)$; (A.13) uses the fact that x_t minimizes $\hat{\kappa}_{\tau,t}(x)$, (A.14) uses Assumption 1, (A.15) again uses Lemma 1, (A.16) follows from the Cauchy-Schwarz inequality; and (A.17) follows from the facts $||w_t^* - w_t||_M = \epsilon_t$ and $B' = \max_x \|f_x\|_{M^{-1}}$. From this point on, the regret bound follows the same arguments as in the proof of Theorem 2. In particular, using (A.17), we write

$$R_T^l \leq 2\beta_T \sum_{t=1}^{T} \langle w_t^*, \sigma_t(x_t, \cdot) \rangle + B' \sum_{t=1}^{T} \epsilon_t \ . \qquad \text{(A.18)}$$

We complete the proof by continuing from (A.5), by using the same steps as in the proof of Theorem 2 starting from (A.5), which leads to the regret bound in the statement of Theorem 4.

A.3.5 Proof of Lemma 3

The first term of Lemma 3 directly follows from Theorem 4. For the second term, we use the result of [55, Theorem 3.4], which is restated below.

[55, Theorem 3.4] Assume $k(c_i, c_j) \leq 1$ for all $c_i, c_j \in \mathcal{C}$. Then, with probability at least $1 - \delta$

$$d(P^*, \hat{P}_t) \leq \frac{1}{\sqrt{t-1}}(2 + \sqrt{2\log(1/\delta)}) \ .$$

To use the lemma above, let $\delta_t = \frac{2\delta}{\pi^2 t^2}$. Then,

$$\mathbb{P}\left(d(P^*, \hat{P}_t) > \frac{1}{\sqrt{t-1}}\left(2 + \sqrt{2\log\left(\frac{\pi^2 t^2}{2\delta}\right)}\right)\right) \leq \frac{2\delta}{\pi^2 t^2} \ .$$

Using a union bound, we obtain

$$\mathbb{P}\left(\exists t \in [T] : d(P^*, \hat{P}_t) > \frac{1}{\sqrt{t-1}}\left(2 + \sqrt{2\log\left(\frac{\pi^2 t^2}{2\delta}\right)}\right)\right) \leq \sum_{t=1}^{T} \frac{2\delta}{\pi^2 t^2} = \frac{2\delta}{\pi^2}\frac{\pi^2}{6} = \frac{\delta}{3} \ .$$

The inequality above implies that

$$\mathbb{P}\left(\forall t \in [T] : d(P^*, \hat{P}_t) > \frac{1}{\sqrt{t-1}}\left(2 + \sqrt{2\log\left(\frac{\pi^2 t^2}{2\delta}\right)}\right)\right) \leq 1 - \frac{\delta}{3} \ .$$

Since $\epsilon_t \leq \frac{1}{\sqrt{t-1}}\left(2 + \sqrt{2\log\left(\frac{\pi^2 t^2}{2\delta}\right)}\right)$, we have with probability at least $1 - \delta$

$$\begin{aligned}
R_T^l &\leq 4\beta_T\sqrt{T\left(2\gamma_T + 2\log\left(\frac{12}{\delta}\right)\right)} + B'\sum_{t=1}^{T}\epsilon_t \ . \\
&\leq 4\beta_T\sqrt{T\left(2\gamma_T + 2\log\left(\frac{12}{\delta}\right)\right)} + B'\epsilon_1 + B'\sum_{t=2}^{T}\frac{1}{\sqrt{t-1}}\left(2 + \sqrt{2\log\left(\frac{\pi^2 t^2}{2\delta}\right)}\right) \\
&\leq 4\beta_T\sqrt{T\left(2\gamma_T + 2\log\left(\frac{12}{\delta}\right)\right)} + B'\epsilon_1 + B'\sum_{t=2}^{T}\frac{1}{\sqrt{t-1}}\left(2 + \sqrt{2\log\left(\frac{\pi^2 T^2}{2\delta}\right)}\right) \\
&= 4\beta_T\sqrt{T\left(2\gamma_T + 2\log\left(\frac{12}{\delta}\right)\right)} + B'\epsilon_1 + B'\sum_{t=1}^{T-1}\frac{1}{\sqrt{t}}\left(2 + \sqrt{2\log\left(\frac{\pi^2 T^2}{2\delta}\right)}\right) \\
&\leq 4\beta_T\sqrt{T\left(2\gamma_T + 2\log\left(\frac{12}{\delta}\right)\right)} + B'\epsilon_1 + B'2\sqrt{T}\left(2 + \sqrt{2\log\left(\frac{\pi^2 T^2}{2\delta}\right)}\right) \ .
\end{aligned}$$

Also, note that

$$\epsilon_1 = \|w^* - w_1\|_M = \left\| w^* - \frac{\mathbf{1}}{n} \right\|_M \leq \sup_{\{w:\|w\|_1=1\}} \left\| w - \frac{\mathbf{1}}{n} \right\|_M = O(1) \ .$$

Appendix B

Appendix for Chapter 3

B.1 Table of Notations

B.2 Unified Formulation RS-G

The p-fragility $\kappa_{\tau,p}(x)$ of action $x \in \mathcal{X}$ is

$$\kappa_{\tau,p}(x) := \min k \text{ s.t. } f(x+\delta) \geq \tau - [kd(x, x+\delta)]^p \,,\ \forall \delta \in \Delta_\infty(x) \,.$$

This means the RS-G action $x_p^{\text{RS-G}}$ gives the reward guarantee of $f(x+\delta) \geq \tau - [\kappa_{\tau,p} d(x_p^{\text{RS-G}}, x_p^{\text{RS-G}} + \delta)]^p$. When $p = 1$, this guaratee is a linear cone facing downward from τ, at the choosen point. Since the p-fragility is is the minimum k that satisfies this inequality, this means the guarantee of the $x_p^{\text{RS-G}}$ action can be represented by the widest cone drawn from τ. When $p = 2$, instead of a linear cone, the guarantee takes quadratic form and so on. We generalize the notion of a cone for all p and define the *fragility-cone* of an action $x' \in \mathcal{X}$ as the following function: $\text{Cone}_{x'}^p(x) := \tau - [\kappa_{\tau,p} d(x', x)]^p$. Then we can say that the solution $x^{\text{RS-G}}$ which minimizes $\kappa_{\tau,p}$, is the solution that maximizes the *wideness* of the fragility-cone, as $\kappa_{\tau,p}$ controls how wide the cone is. Figure 3.2 gives a visual representation of RS-1, RS-2 and RS-G with $p = 2$ solutions, and their respective

fragility-cones. Which formulation should be used depends on the structure of the problem and the goal of the optimizer. If, for example large perturbations are considered unlikely, RS-G with a larger p might be better suited.

B.3 Proofs

B.3.1 Proof of Lemma 5

Assume the good event $\mathcal{E}$ holds. Consider the first inequality $\overline{\kappa}_{\tau,t}(x) \leq \kappa_\tau(x)$. When $\overline{\kappa}_{\tau,t}(x) = \infty$, then $\kappa_\tau(x) = \infty$ as well. When $\overline{\kappa}_{\tau,t}(x) < \infty$ and $\kappa_\tau(x) = \infty$, the inequality holds. Let $\bar{\delta}_{x,t} := \arg\max_{\delta \in \Delta_\infty(x)\setminus 0} \frac{\tau - \mathrm{ucb}_t(x+\delta)}{d(x,x+\delta)}$ and $\delta_{x,t} := \arg\max_{\delta \in \Delta_\infty(x)\setminus 0} \frac{\tau - f(x+\delta)}{d(x,x+\delta)}$. When $\overline{\kappa}_{\tau,t}(x), \kappa_\tau(x) < \infty$ we have

$$\begin{aligned}
\overline{\kappa}_{\tau,t}(x) - \kappa_\tau(x) &= \frac{\tau - \mathrm{ucb}_t(x+\bar{\delta}_{x,t})}{d(x, x+\bar{\delta}_{x,t})} - \frac{\tau - f(x+\delta_{x,t})}{d(x, x+\delta_{x,t})} \\
&\leq \frac{\tau - \mathrm{ucb}_t(x+\bar{\delta}_{x,t})}{d(x, x+\bar{\delta}_{x,t})} - \frac{\tau - f(x+\bar{\delta}_{x,t})}{d(x, x+\bar{\delta}_{x,t})} \\
&= \frac{f(x+\bar{\delta}_{x,t}) - \mathrm{ucb}_t(x+\bar{\delta}_{x,t})}{d(x, x+\bar{\delta}_{x,t})} \\
&\leq 0 \ .
\end{aligned}$$

Similarly $\kappa_\tau(x) \leq \underline{\kappa}_{\tau,t}(x)$ follows.

B.3.2 Proof of Lemma 6

Assume the good event $\mathcal{E}$ holds. Consider the inequality $\epsilon_\tau(x) \leq \overline{\epsilon}_{\tau,t}(x)$. Assume Lemma 6 is false, then $\exists \delta$ s.t. $\delta \in \Delta_{\epsilon_\tau(x)}(x)$ and $\delta \notin \Delta_{\overline{\epsilon}_{\tau,t}(x)}(x)$ since $\Delta_{\overline{\epsilon}_{\tau,t}(x)}(x) \subset \Delta_{\epsilon_\tau(x)}(x)$, and $f(x+\delta) \geq \tau > \mathrm{ucb}_t(x+\delta)$, which is a contradiction. Therefore Lemma 6 is true.

B.3.3 Proof of Theorem 7

Lenient Regret Bound of AdveRS-1 The regret bound of AdveRS-1 follows a similar structure to the bound in [56], but instead of the MMD metric, we utilize the Lipschitz continuity of f. Define the instantaneous regret at round t as $r_t^l := \tau - f(\tilde{x}_t + \delta_t)$. The cumulative regret is $R_T^l = \sum_{t=1}^{T} (r_t^l)^+$. Assume that the good event $\mathcal{E}$ holds. Then the instantaneous regret can be bounded as

$$r_t^l = \tau - f(\tilde{x}_t + \delta_t) \leq \tau - \mathrm{ucb}_t(\tilde{x}_t + \delta_t) + 2\beta_t \sigma_t(\tilde{x}_t + \delta_t) \ . \tag{B.1}$$

If the adversary does not perturb the action, i.e., $\delta_t = 0$, then we have

$$r_t^l \leq \tau - \mathrm{ucb}_t(\tilde{x}_t) + 2\beta_t \sigma_t(\tilde{x}_t)$$

by Assumption 5 and the selection rule of the algorithm we have $\mathrm{ucb}_t(\tilde{x}_t) \geq \tau$, hence

$$r_t^l \leq 2\beta_t \sigma_t(\tilde{x}_t) = 2\beta_t \sigma_t(x_t) \ .$$

If $\delta_t \neq 0$, continuing from (B.1),

$$\begin{aligned}
r_t^l &\leq \tau - \mathrm{ucb}_t(\tilde{x}_t + \delta_t) + 2\beta_t \sigma_t(\tilde{x}_t + \delta_t) && \text{(B.2)} \\
&\leq d(\tilde{x}_t, \tilde{x}_t + \delta_t) \frac{\tau - \mathrm{ucb}_t(\tilde{x}_t + \delta_t)}{d(\tilde{x}_t, \tilde{x}_t + \delta_t)} + 2\beta_t \sigma_t(\tilde{x}_t + \delta_t) && \text{(B.3)} \\
&\leq d(\tilde{x}_t, \tilde{x}_t + \delta_t) \overline{\kappa}_{\tau,t}(\tilde{x}_t) + 2\beta_t \sigma_t(\tilde{x}_t + \delta_t) && \text{(B.4)} \\
&\leq d(\tilde{x}_t, \tilde{x}_t + \delta_t) \overline{\kappa}_{\tau,t}(\hat{x}) + 2\beta_t \sigma_t(\tilde{x}_t + \delta_t) && \text{(B.5)}
\end{aligned}$$

Define $\bar{\delta}_{x,t} := \arg\max_{\delta \in \Delta_\infty(x) \setminus 0} \frac{\tau - \mathrm{ucb}_t(x+\delta)}{d(x, x+\delta)}$. Then we can write above as,

$$\begin{aligned}
&= d(\tilde{x}_t, \tilde{x}_t + \delta_t) \frac{\tau - \mathrm{ucb}_t(\hat{x} + \bar{\delta}_{\hat{x},t})}{d(\hat{x}, \hat{x} + \bar{\delta}_{\hat{x},t})} + 2\beta_t \sigma_t(\tilde{x}_t + \delta_t) && \text{(B.6)} \\
&\leq d(\tilde{x}_t, \tilde{x}_t + \delta_t) \frac{f(\hat{x}) - f(\hat{x} + \bar{\delta}_{\hat{x},t})}{d(\hat{x}, \hat{x} + \bar{\delta}_{\hat{x},t})} + 2\beta_t \sigma_t(\tilde{x}_t + \delta_t) && \text{(B.7)} \\
&\leq d(\tilde{x}_t, \tilde{x}_t + \delta_t) \frac{B d(\hat{x}, \hat{x} + \bar{\delta}_{\hat{x},t})}{d(\hat{x}, \hat{x} + \bar{\delta}_{\hat{x},t})} + 2\beta_t \sigma_t(\tilde{x}_t + \delta_t) && \text{(B.8)} \\
&\leq 2\beta_t \sigma_t(\tilde{x}_t + \delta_t) + B\epsilon_t = 2\beta_t \sigma_t(x_t) + B\epsilon_t \ . && \text{(B.9)}
\end{aligned}$$

(B.4) comes from the definition (3.7) and (B.5) follows from $\tilde{x}_t$ being the minimizer of $\overline{\kappa}_{\tau,t}(x)$. (B.7) follows from Assumption 5 and the confidence bounds. Finally (B.8) follows from the Lipschitz continuity of f with respect to the kernel metric d, and (B.9) from $d(\tilde{x}_t, \tilde{x}_t + \delta_t) \leq \epsilon_t$. From this point on, we bound the lenient regret by following standard steps for bounding regret of GP bandits. First note that since (B.9) is nonnegative, it is also a bound for $(r_t^l)^+$. By plugging this upper bound to (3.6), and using monotonicity of β_t, we obtain

$$R_T^l \leq 2\beta_T \sum_{t=1}^{T} \sigma_t(x_t) + B \sum_{t=1}^{T} \epsilon_t \tag{B.10}$$

$$\leq 2\beta_T \sqrt{T \sum_{t=1}^{T} \sigma_t(x_t)^2} + B \sum_{t=1}^{T} \epsilon_t \tag{B.11}$$

where (B.11) uses the Cauchy-Schwartz inequality. Next, to relate the sum of variances that appears in (B.11) to the maximum information gain, we use $\alpha \leq 2\log(1+\alpha)$ for all $\alpha \in [0,1]$ to obtain

$$\sum_{t=1}^{T} \sigma_t(x_t)^2 \leq \sum_{t=1}^{T} 2\log\big(1 + \sigma_t(x_t)^2\big)$$

$$\leq 4\gamma_T \ , \tag{B.12}$$

where (B.12) follows from [48, Lemma 4]. Putting (B.12) in the regret definition we get the final bound

$$R_T^l \leq 4\beta_T \sqrt{T\gamma_T} + B \sum_{t=1}^{T} \epsilon_t \ .$$

□

Robust Satisficing Regret Bound of AdveRS-1 Define the instantaneous regret at round t as $r_t^{rs} := \tau - \kappa_\tau \epsilon_t - f(\tilde{x}_t + \delta_t)$. The cumulative regret is $R_T^{rs} = \sum_{t=1}^{T} (r_t^{rs})^+$. Assume that the good event $\mathcal{E}$ holds. Then the instantaneous

regret can be bounded as

$$r_\tau^{\text{rs}} \leq \tau - \kappa_\tau \epsilon_t - \text{ucb}_t(\tilde{x}_t + \delta_t) + 2\beta_t \sigma_t(\tilde{x}_t + \delta_t) \tag{B.13}$$

$$\leq \tau - \kappa_\tau \epsilon_t - (\tau - \overline{\kappa}_{\tau,t}(\tilde{x}_t) d(\tilde{x}_t, \tilde{x}_t + \delta_t)) + 2\beta_t \sigma_t(\tilde{x}_t + \delta_t) \tag{B.14}$$

$$\leq \tau - \kappa_\tau \epsilon_t - (\tau - \overline{\kappa}_{\tau,t}(\tilde{x}_t) \epsilon_t) + 2\beta_t \sigma_t(\tilde{x}_t + \delta_t) \tag{B.15}$$

$$= \epsilon_t(\overline{\kappa}_{\tau,t} - \kappa_\tau) + 2\beta_t \sigma_t(\tilde{x}_t + \delta_t) \tag{B.16}$$

$$\leq 2\beta_t \sigma_t(\tilde{x}_t + \delta_t) = 2\beta_t \sigma_t(x_t) \ . \tag{B.17}$$

(B.14) follows from the guarantee of the RS formulation $\text{ucb}_t(\tilde{x}_t + \delta) \geq \tau - \overline{\kappa}_{\tau,t} d(\tilde{x}_t, \tilde{x}_t + \delta)$ for any $\delta \in \Delta_\infty(\tilde{x}_t)$. (B.17) follows from $\overline{\kappa}_{\tau,t} \leq \kappa_{\tau,t}$ since $\text{ucb}_t(x) \geq f(x)$ for all x. Then we bound the cumulative robust satisficing regret, following the same standard steps as in the proof of the cumulative lenient regret bound, to obtain an upper bound of:

$$R_T^{rs} \leq 4\beta_T \sqrt{T\gamma_T} \ .$$

□

B.3.4 Proof of Theorem 10

Under the good event $\mathcal{E}$, Assumption 5 and Assumption 9, we can bound the lenient regret of AdveRS-2 as:

$$r_\tau^l := \tau - f(\tilde{x}_t + \delta_t) \tag{B.18}$$

$$\leq \min_{\delta \in \Delta_{\overline{\epsilon}_{\tau,t}}(\tilde{x}_t)} \text{ucb}_t(\tilde{x}_t + \delta) - f(\tilde{x}_t + \delta_t) \tag{B.19}$$

$$\leq \min_{\delta \in \Delta_{\epsilon_\tau}(\tilde{x}_t)} \text{ucb}_t(\tilde{x}_t + \delta) - f(\tilde{x}_t + \delta_t) \tag{B.20}$$

$$\leq \min_{\delta \in \Delta_{\epsilon_t}(\tilde{x}_t)} \text{ucb}_t(\tilde{x}_t + \delta) - f(\tilde{x}_t + \delta_t) \tag{B.21}$$

$$\leq \text{ucb}_t(\tilde{x}_t + \delta_t) - f(\tilde{x}_t + \delta_t) \tag{B.22}$$

$$\leq 2\beta_t \sigma_t(x_t) \ . \tag{B.23}$$

Note that from Assumption 5, under $\mathcal{E}$ we have $\text{ucb}_t(\hat{x}) \geq \tau$ which means $\overline{\epsilon}_{\tau,t}(\hat{x}) \geq 0$. Then by definition, for any $\delta \in \Delta_{\overline{\epsilon}_t}(\tilde{x}_t)$, $\text{ucb}_t(\tilde{x}_t + \delta_t) \geq \tau$, which gives

(B.19). (B.20) is true because $\epsilon_\tau \leq \overline{\epsilon}_{\tau,t}(x^{\text{RS-2}})$ by Lemma 6 and $\overline{\epsilon}_{\tau,t}(x^{\text{RS-2}}) \leq \overline{\epsilon}_{\tau,t}$ by the definition (3.8). Hence $\Delta_{\epsilon_\tau}(\tilde{x}_t) \subseteq \Delta_{\overline{\epsilon}_{\tau,t}}(\tilde{x}_t)$. Similarly, (B.21) follows from Assumption 9. For (B.22), observe that $\delta_t \in \Delta_{\epsilon_t}(\tilde{x}_t)$, no matter the attacking strategy of the adversary. Finally (B.23) holds under the good event $\mathcal{E}$. The rest of the proof is the same as the previous ones. Noting that $R^{rs} \leq R^l$ by definition, and hence shares the same upper bound completes the proof of Theorem 10. $\square$

B.3.5 Proof of Corollary 6

By the definition of $\underline{\kappa}_{\tau,t}(x)$, we have $\text{lcb}_t(\tilde{x}_t+\delta) \geq \tau - \underline{\kappa}_{\tau,t} d(\tilde{x}_t, \tilde{x}_t+\delta) \quad \forall \delta \in \Delta_\infty(\tilde{x}_t)$. Noting that under $\mathcal{E}$ $f(x) \geq \text{lcb}_t(x)$, we obtain the first inequality. For the second inequality observe that from the definition of $\overline{\kappa}_{\tau,t}(x)$ we know that $\exists \delta' \in \Delta_\infty(\tilde{x}_t)$ such that $\text{ucb}_t(\tilde{x}_t + \delta') = \tau - \overline{\kappa}_{\tau,t} d(\tilde{x}_t, \tilde{x}_t + \delta')$. Noting that $f(x) \leq \text{ucb}(x)$ for all x under $\mathcal{E}$ concludes the proof.

B.3.6 Proof of Corollary 3.2.2

By Lemma 6 we have $\Delta_{\underline{\epsilon}_{\tau,t}}(\tilde{x}_t) \subseteq \Delta_{\epsilon_\tau}(\tilde{x}_t)$, then for any $\delta \in \Delta_{\underline{\epsilon}_{\tau,t}}(\tilde{x}_t)$, we have $f(\tilde{x}_t + \delta) \geq \text{lcb}_t(\tilde{x}_t + \delta) \geq \tau$ under the selection rule and the good event $\mathcal{E}$. Conversely again by Lemma 6 we have $\Delta_{\epsilon_\tau}(\tilde{x}_t) \subseteq \Delta_{\overline{\epsilon}_{\tau,t}}(\tilde{x}_t)$. Assuming $\mathcal{X}$ is continuous, $\exists \delta' \in \Delta_{\epsilon_\tau}$ such that $f(\tilde{x}_t + \delta') = \tau$. Noting that $\delta' \in \Delta_{\overline{\epsilon}_{\tau,t}}$ completes the proof.

B.3.7 Proof of Proposition 8

Consider the following instance of our problem. Let $\mathcal{X} = [0, 1]$ and $f(x) = x$, noting that this is an element of RKHS with a linear kernel. The Lipschitz constant of this function is 1. Let the satisficing goal $\tau = 1$ which is the function maximum. If $\epsilon_t < 1$, no matter what action is chosen by the learner, the adversary can choose a perturbation $\delta_t = -\epsilon_t$ to obtain a reward $f(x_t + \delta_r) \leq \tau - \epsilon_t$, making

the instantaneous lenient regret $r^{\mathrm{l}} = (\tau - f(x_t + \delta_t))^+ \geq \epsilon_t$. This constitute a worst case lower bound, hence the bound is tight and the linear penalty term is not avoidable in general.

B.3.8 Proof of Proposition 11

WLOG assume that $\exists \delta \in \Delta_\infty(x)$ such that $f(x+\delta) \leq \tau$. Let $\Delta'(x) = \{\delta \in \Delta_\infty(x) \,|\, f(x+\delta) \leq \tau\}$. Notice that (3.10) is then equivalent to

$$\kappa_{\tau,p}(x) := \min k \text{ s.t. } f(x+\delta) \geq \tau - [kd(x, x+\delta)]^p \,, \quad \forall \delta \in \Delta'(x) \,,$$

since for any $\delta \notin \Delta'(x)$, the inequality constraint already holds for any $k \geq 0$.

Then we can write the equivalent formulation for an action that satisfies $f(x) \geq \tau$

$$\kappa_{\tau,p}(x) := \max_{\delta \in \Delta'(x)} \frac{[\tau - f(x+\delta)]^{1/p}}{d(x, x+\delta)} \,.$$

as $p \to \infty$ we have

$$\kappa_{\tau,p}(x) = \max_{\delta \in \Delta'(x)} \frac{1}{d(x, x+\delta)}$$
$$\frac{1}{\kappa_{\tau,p}(x)} = \min_{\delta \in \Delta'(x)} d(x, x+\delta) = \epsilon_\tau(x) \,.$$

Then $\lim_{p\to\infty} x^{\text{RS-G}} = \min_{x\in\mathcal{X}} \lim_{p\to\infty} \kappa_{\tau,p} = \min_{x\in\mathcal{X}} \frac{1}{\epsilon_\tau(x)} = \max_{x\in\mathcal{X}} \epsilon_\tau(x) = x^{\text{RS-2}}$.

B.3.9 Proof of Corollary 12

The proof is almost identical to the RS regret proof for AdveRS-1, we include it for completeness. Define the instantaneous regret at round t as $r_t^{rs\text{-}g} := \tau - [\kappa_\tau \epsilon_t]^p - f(\tilde{x}_t + \delta_t)$. The cumulative regret is $R_T^{rs\text{-}g} = \sum_{t=1}^T (r_t^{rs\text{-}g})^+$. Assume that

the good event $\mathcal{E}$ holds. Then the instantaneous regret can be bounded as

$$
\begin{aligned}
r_\tau^{\text{rs-g}} &\leq \tau - [\kappa_\tau \epsilon_t]^p - \text{ucb}_t(\tilde{x}_t + \delta_t) + 2\beta_t \sigma_t(\tilde{x}_t + \delta_t) && \text{(B.24)} \\
&\leq \tau - [\kappa_\tau \epsilon_t]^p - (\tau - [\overline{\kappa}_{\tau,p,t}(\tilde{x}_t) d(\tilde{x}_t, \tilde{x}_t + \delta_t)]^p) + 2\beta_t \sigma_t(\tilde{x}_t + \delta_t) && \text{(B.25)} \\
&\leq \tau - [\kappa_\tau \epsilon_t]^p - (\tau - [\overline{\kappa}_{\tau,p,t}(\tilde{x}_t) \epsilon_t)]^p + 2\beta_t \sigma_t(\tilde{x}_t + \delta_t) && \text{(B.26)} \\
&= \epsilon_t^p (\overline{\kappa}_{\tau,p,t}^p - \kappa_\tau^p) + 2\beta_t \sigma_t(\tilde{x}_t + \delta_t) && \text{(B.27)} \\
&\leq 2\beta_t \sigma_t(\tilde{x}_t + \delta_t) = 2\beta_t \sigma_t(x_t) \ . && \text{(B.28)}
\end{aligned}
$$

(B.25) follows from the guarantee of the RS-G formulation $\text{ucb}_t(\tilde{x}_t + \delta) \geq \tau - [\overline{\kappa}_{\tau,p,t} d(\tilde{x}_t, \tilde{x}_t + \delta)]^p$ for any $\delta \in \Delta_\infty(\tilde{x}_t)$. (B.27) follows from $\overline{\kappa}_{\tau,p,t} \leq \kappa_{\tau,p,t}$ since $\text{ucb}_t(x) \geq f(x)$ for all x. Then we bound the cumulative robust satisficing regret, following the same standard steps to obtain an upper bound of:

$$R_T^{rs\text{-}g} \leq 4\beta_T \sqrt{T\gamma_T} \ .$$

□

B.4 Implementation details of the algorithms

We note that maximizing UCB and LCB over a compact domain is non-trivial due to the non-convex nature of the posterior mean and variance. Like foundational works (*e.g.*, [46]), we focus on theoretical guarantees and assume an oracle optimizer, abstracting computational details. This challenge is common across all GP-based methods, including the RO algorithms we compare against, and in practical BO, various optimization (*e.g.*, gradient descent or quasi-Newton) techniques are used to address it.

In our experiments, we work with discretized domains and the implementation of the acquisition functions of our algorithms have complexity $\mathcal{O}(N^2)$ where $N = |\mathcal{X}|$. Notably, this complexity is the same as that of RO-based algorithms.

In all algorithms that we implement (RS or RO based), action-specific calculations are performed in an inner loop, while an outer loop is used the selection of best action.

B.5 Additional Experimental Results

B.5.1 Inverted Pendulumn Experiment

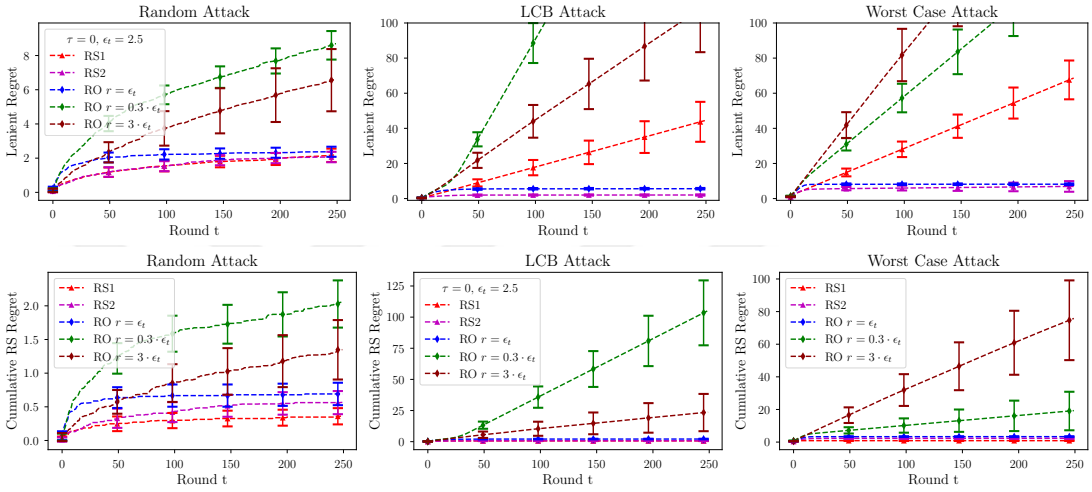


Figure B.1: Lenient and RS regret results of the inverted pendulum experiment. Perturbation budget is $\epsilon_t = 2.5$ and $\tau = 0$ for all t. STABLEOPT run with: $r = \epsilon_t$, $r = 3\epsilon_t$, and $r = 0.3\epsilon_t$. Attack schemes are indicated in the supertitles.

In this experiment, we focus on optimizing a parameterized controller to achieve robust performance in controlling an inverted pendulum. The state vector $\boldsymbol{s}_k$ consists of four variables: cart position, pendulum angle, and their derivatives. The simulation is conducted using OpenAI's Gym environment. The system dynamics are described by $\boldsymbol{s}_{k+1} = h(\boldsymbol{s}_k, u_k)$, where $\boldsymbol{s}_k$ is the 4-dimensional state vector, and $u_k \in [-10, 10]$ is the control action representing the voltage applied to the controller at step k. We implement a static feedback controller similar to the one in [57], defined as $u_k = \boldsymbol{F}\boldsymbol{s}_k$, where $\boldsymbol{F} \in \mathbb{R}^{1\times 4}$ is the gain matrix. We structure $\boldsymbol{F}$ as a Linear Quadratic Regulator (LQR) using the linearized system dynamics $(\boldsymbol{A}, \boldsymbol{B})$ around the equilibrium point. The parameterization, which has been shown to be learnable by GP's [57] is given by $\boldsymbol{F} = \text{dlqr}(\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{W}_s(\boldsymbol{\theta}), \boldsymbol{W}_u(\boldsymbol{\theta}))$ with the following parameterization:

$$\begin{aligned} \boldsymbol{W}_s(\boldsymbol{\theta}) &= \text{diag}(10^{\theta_1}, 10^{\theta_2}, 10^{\theta_3}, 0.1), & \theta_{1,2,3} &\in [-3, 2], \\ \boldsymbol{W}_u(\boldsymbol{\theta}) &= 10^{-\theta_4}, & \theta_4 &\in [1, 5]\ . \end{aligned}$$

We discretize the parameter space into 4096 points and compute the cost for each parameter by simulating it over 2000 seconds, using the same cost function as in [58]. After the simulation, the cost function is standardized and clipped to the range $[-2, 2]$. For the GP kernel, we use a Radial Basis Function (RBF) kernel with Automatic Relevance Determination (ARD), with hyperparameters selected using 400 samples prior to the experiment.

For adversarial perturbations, we consider an adversary that damages the learning process by perturbing the selected parameter. The unknown perturbation budget is constant with $\epsilon_t = 2$. The satisficing goal is set to $\tau = 0$ which corresponds to ~ 25 percentile of the objective function. Figure B.1 shows the regrets of the algorithms over a time horizon of $T = 250$. AdveRS-2 again outperforms STABLEOPT even when it is run with the perfect knowledge of the adversarial budget. While AdveRS-1 achieves linear regret, it can still perform better than STABLEOPT when the perturbation budget is misspecified.

B.5.2 Robust Satisficing Regrets of Main Experiments

Figures B.2, B.3 present the RS regret plots for the synthetic and insulin dosage experiments from the main paper.

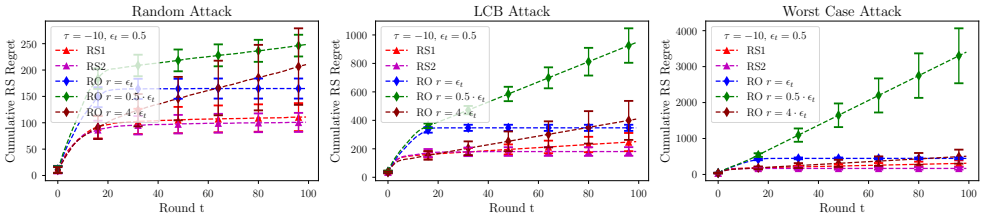


Figure B.2: Robust satisficing regret results for synthetic experiment with a perturbation budget of $\epsilon_t = 0.5$ and $\tau = -10$ for all t. STABLEOPT run with: $r = \epsilon_t$, $r = 4\epsilon_t$, and $r = 0.5\epsilon_t$. Attack schemes are indicated in the supertitles.

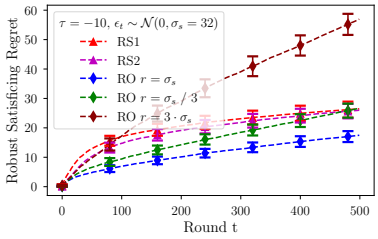


Figure B.3: Lenient regret results of insulin dosage experiment with $\tau = -10$ for all t. STABLEOPT run with: $r = \sigma_s$, $r = 3\sigma_s$, and $r = \sigma_s/3$.

B.5.3 Misspecified Kernel

In this experiment, we evaluate the performance of our algorithms when the GP is run with a kernel that is not aligned with the true function. Specifically, we use the function shown in Figure 3.1, but unlike Experiment 1, we replace the polynomial kernel with an RBF kernel, using a lengthscale and variance of both 1. All algorithms are run with the same parameters as in Experiment 1 with $\tau = 10$ and $\epsilon_t = 0.5$. Figure B.4 shows that AdveRS-2 maintains strong performance even when the kernel is misspecified, while AdveRS-1 outperforms STABLEOPT when the ambiguity ball is not correctly estimated. Figure B.5 shows the RS regrets and we see that our algorithms achieve sublinear results, while STABLEOPT with misspecified ambiguity radius can achieve linear RS regret. All plots show the mean results over 100 independent runs with error bars showing std/2.

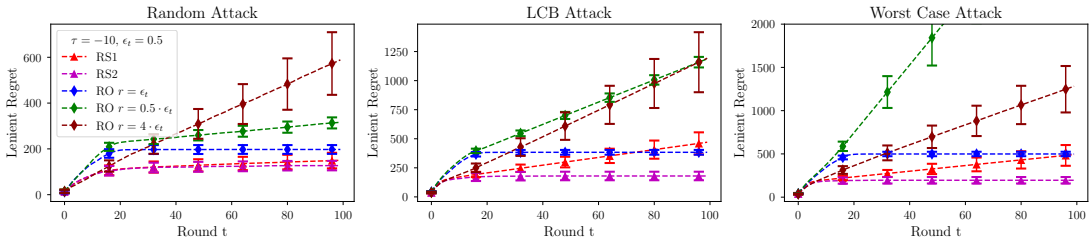


Figure B.4: Lenient regret results for synthetic experiment using a misspecified kernel, with a perturbation budget of $\epsilon_t = 0.5$ and $\tau = -10$ for all t. STABLEOPT run with: $r = \epsilon_t$, $r = 4\epsilon_t$, and $r = 0.5\epsilon_t$.

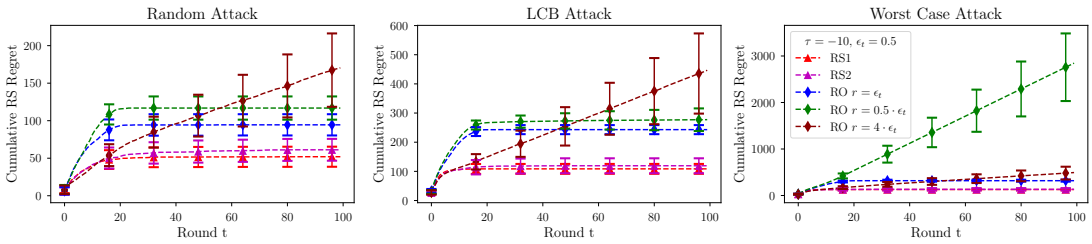


Figure B.5: Robust satisficing regret results for synthetic experiment using a misspecified kernel, with a perturbation budget of $\epsilon_t = 0.5$ and $\tau = -10$ for all t. STABLEOPT run with: $r = \epsilon_t$, $r = 4\epsilon_t$, and $r = 0.5\epsilon_t$.

B.5.4 Effect of τ and r

Table B.2 shows the varying effects of the parameters. The lenient regret of all algorithms are calculated on the synthetic function from Experiment 1. The perturbations are sampled iid from a normal distribution $\mathcal{N}(0, 0.3)$. Results show that the target oriented nature of RS approaches makes them better suited for the satisficing objective under unknown perturbations.

Alg. \ τ	-30.00	-20.00	-10.00	0.00	10.00
RS1	0.32	0.65	1.36	4.36	12.81
RS2	**0.18**	**0.32**	**0.72**	**1.99**	**3.14**
RO, r=0.10	0.50	0.95	1.53	2.22	3.25
RO, r=0.57	0.42	0.83	1.36	2.67	12.46
RO, r=1.05	0.32	0.61	1.00	5.69	15.68
RO, r=1.52	0.19	0.38	1.59	8.24	18.24
RO, r=2.00	0.20	2.01	6.75	15.75	25.72

Table B.2: Average lenient regrets for synthetic experiment for RS-1, RS-2 and RO algorithms with varying τ and r values.

Notation	Description
$\mathcal{X}$	Action set
τ	satisficing level (threshold)
$d(\cdot,\cdot) : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$	Distance metric on $\mathcal{X}$
$\Delta_\epsilon(x)$	$\{\mathbf{x}' - \mathbf{x} : \mathbf{x}' \in \mathcal{X} \text{ and } d(\mathbf{x}, \mathbf{x}') \leq \epsilon\}$, set of perturbations in ϵ-ball
$\hat{x}$	$\arg\max_{x \in \mathcal{X}} f(x)$
$\tilde{x}_t$	Action selected by each algorithm
x_t	Point sampled after perturbation, $\tilde{x}_t + \delta_t$
δ_t	Perturbation of the adversary at round t
ϵ_t	True perturbation budget of the adversary at time t
$\kappa_\tau(x)$	$\begin{cases} \max_{\delta \in \Delta_\infty(x) \setminus 0} \ \frac{\tau - f(x+\delta)}{d(x, x+\delta)} & \text{if } f(x) \geq \tau \\ +\infty & \text{if } f < \tau \end{cases}$
κ_τ	$\min_{x \in \mathcal{X}} \kappa_\tau(x)$
$x^{\text{RS-1}}$	$\arg\min_{x \in \mathcal{X}} \kappa_{\tau,t}(x)$
$\overline{\kappa}_{\tau,t}(x)$	$\begin{cases} \max_{\delta \in \Delta_\infty(x) \setminus 0} \ \frac{\tau - \text{ucb}_t(x+\delta)}{d(x, x+\delta)} & \text{if } \text{ucb}_t(x) \geq \tau \\ +\infty & \text{if } \text{ucb}_t(x) < \tau \end{cases}$
$\overline{\kappa}_{\tau,t}$	$\min_{x \in \mathcal{X}} \overline{\kappa}_{\tau,t}(x)$
$\underline{\kappa}_{\tau,t}(x)$	$\begin{cases} \max_{\delta \in \Delta_\infty(x) \setminus 0} \ \frac{\tau - \text{lcb}_t(x+\delta)}{d(x, x+\delta)} & \text{if } \text{lcb}_t(x) \geq \tau \\ +\infty & \text{if } \text{lcb}_t(x) < \tau \end{cases}$
$\underline{\kappa}_{\tau,t}$	$\min_{x \in \mathcal{X}} \underline{\kappa}_{\tau,t}(x)$
$\epsilon_\tau(x)$	$\max_{x' \in \mathcal{X}} d(x, x') \ s.t. \ f(x+\delta) \geq \tau, \ \forall \delta \in \Delta_\epsilon(x)$
ϵ_τ	$\max_{x \in \mathcal{X}} \epsilon_\tau(x)$
$x^{\text{RS-2}}$	$\arg\max_{x \in \mathcal{X}} \epsilon_\tau(x)$
$\overline{\epsilon}_{\tau,t}(x)$	$\max_{x' \in \mathcal{X}} d(x, x') \ s.t. \ \text{ucb}_t(x+\delta) \geq \tau, \ \forall \delta \in \Delta_\epsilon(x)$
$\overline{\epsilon}_{\tau,t}$	$\max_{x \in \mathcal{X}} \overline{\epsilon}_{\tau,t}(x)$
$\underline{\epsilon}_{\tau,t}(x)$	$\max_{x' \in \mathcal{X}} d(x, x') \ s.t. \ \text{lcb}_t(x+\delta) \geq \tau, \ \forall \delta \in \Delta_\epsilon(x)$
$\underline{\epsilon}_{\tau,t}$	$\max_{x \in \mathcal{X}} \underline{\epsilon}_{\tau,t}(x)$

Table B.1: Table of notations