Guaranteed bounds for optimal stopping problems using kernel-based non-asymptotic uniform confidence bands

Martin Glanzer[a,1], Sebastian Maier[b,1,*], Georg Ch. Pflug[c,d,1]

[a] *University of Mannheim Business School, Schloss, 68131 Mannheim, Germany*
[b] *Department of Statistical Science, University College London, Gower Street, London WC1E 6BT, United Kingdom*
[c] *Faculty of Economics and Statistics, University of Vienna, Oskar-Morgenstern Platz 1, 1090 Vienna, Austria*
[d] *Advancing Systems Analysis Program, International Institute for Applied Systems Analysis, Schlossplatz 1, 2361 Laxenburg, Austria*

In this paper, we introduce an approach for obtaining probabilistically guaranteed upper and lower bounds on the true optimal value of stopping problems. Bounds of existing simulation-and-regression approaches, such as those based on least squares Monte Carlo and information relaxation, are stochastic in nature and therefore do not come with a finite sample guarantee. Our data-driven approach is fundamentally different as it allows replacing the sampling error with a pre-specified confidence level. The key to this approach is to use high- and low-biased estimates that are guaranteed to over- and underestimate, respectively, the conditional expected continuation value that appears in the stopping problem's dynamic programming formulation with a pre-specified confidence level. By incorporating these guaranteed over- and underestimates into a backward recursive procedure, we obtain probabilistically guaranteed bounds on the problem's true optimal value. As a byproduct we present novel kernel-based non-asymptotic uniform confidence bands for regression functions from a reproducing kernel Hilbert space. We derive closed-form formulas for the cases where the data-generating distribution is either known or unknown, which makes our data-driven approach readily applicable in a range of practical situations including simulation. We illustrate the applicability of the proposed bounding procedure by valuing a Bermudan put option.

Keywords: Stochastic programming, Optimal stopping, Finite sample guarantees, Reproducing kernel Hilbert spaces, Approximate dynamic programming

1. Introduction

Optimal stopping is a fundamental class of problems in stochastic optimisation that has attracted considerable attention in the operational research (OR) literature. This is attributable to the simple elegance of optimal stopping and its wide applicability (Powell, 2019). In its standard form, an optimal stopping problem requires the decision maker to choose when to stop in order to maximise

*Corresponding author
Email addresses: martin.glanzer@uni-mannheim.de (Martin Glanzer), s.maier@ucl.ac.uk (Sebastian Maier), georg.pflug@univie.ac.at (Georg Ch. Pflug)
[1] All authors contributed equally to this work.

(minimise) an expected reward (cost)[2]. When formulated as a discrete time dynamic program using Bellman's principle of optimality, the algorithmic solution strategy simplifies to a binary decision at each point in time: Stop the process or continue until the next time period. There are numerous applications in areas such as finance (Wei and Zhu, 2022), energy (Nadarajah et al., 2017), real options (Maier et al., 2020a,b), economics (Kang, 2005), marketing (Feng and Gallego, 1995), inventory management (Van Foreest and Kilic, 2023), and healthcare (David and Yechiali, 1985). In addition, optimal stopping is an active area of methodological research, with recent contributions to the OR literature extending this problem class to, e.g., risk aversion (Pichler et al., 2022), robust optimisation (Sturt, 2023), interpretability (Ciocan and Mišić, 2022), pathwise optimisation (Desai et al., 2012), duality (Ibáñez and Velasco, 2020) and randomisation (Dong, 2024).

Despite the simple formulation, finite-horizon optimal stopping problems do not permit analytic solutions, so one has to resort to numerical approximations. Standard techniques to numerically approximate the value of optimal stopping problems apply either a finite difference scheme, a binomial/trinomial lattice (i.e. a tree) model, or Monte Carlo simulation. Simulation-based techniques first generate a set of sample paths from the underlying stochastic model in the form a scenario fan, and then apply dynamic programming to approximate the optimal value of the stopping problem using the generated scenario fan. While tree-based approaches guarantee non-anticipative policies by construction, scenario fan-based approaches have to ensure non-anticipativity in an explicit way as part of the optimisation step. So-called simulation-and-regression approaches accomplish this through approximating the continuation function (which is a conditional expectation) that appears in the dynamic programming recursion by a – parametric or nonparametric – regression estimator. Carriere (1996) was the first to present the simulation-and-regression idea using non-parametric regression estimates. The seminal work of Longstaff and Schwartz (2001) (as well as the parallel work of Tsitsiklis and Van Roy (2001)) then popularised the idea of blending tools from regression analysis with stochastic dynamic programming, which is widely studied in the OR literature (see, e.g., Nadarajah et al. (2017); Maier et al. (2020a); Wei and Zhu (2022); Ciocan and Mišić (2022)).

Existing simulation-and-regression approaches can be divided into those that provide lower bounds and those that provide upper bounds on the optimal value of a stopping problem. Sub-optimal policies that give a lower bound are often comparatively simple to determine since any feasible policy will result in a lower bound. For example, the use of regression-based approximations of the continuation functions in the least-squares Monte Carlo method of Longstaff and Schwartz (2001) leads to sub-optimal decisions, and consequently gives a lower bound. Finding (good) upper bounds, by contrast, is much more involved. The existing stream of upper-bound literature is mainly related to a set of techniques collectively referred to as information relaxation (see the recent monograph of Brown and Smith (2022)). The key idea behind these techniques is to first provide the decision maker with more information by relaxing the nonanticipativity constraint, and then to penalise violations of the temporal feasibility constraints. The resulting lower and upper

[2]We focus on the problem of reward maximisation throughout the paper, but all our results directly apply to the symmetric problem of cost minimisation upon reversing the signs.

bound estimates then provide a range for the optimal value and the gap between these two bounds serves as an indicator for the quality of determined (sub-optimal) stopping policies.

However, lower and upper bounds of existing simulation-and-regression approaches depend on the sampled paths of the underlying stochastic process and therefore do not come with probabilistic guarantees. In other words, the resulting performance bounds are stochastic in nature, so are not guaranteed to be valid with respect to the true underlying distribution. This is because even if the generated scenario fan contains a large number of sample paths and is sufficiently rich, the numerical estimate that results from the subsequent optimisation step still depends on the specific set of sample paths generated. To the best of our knowledge, no guaranteed bounds for stopping problems are available in the existing literature. Here we close this gap by introducing a data-driven approach for obtaining probabilistically guaranteed bounds on the true optimal value of stopping problems. In doing so, our work contributes to the extensive literature on optimal stopping and approximate dynamic programming (Powell, 2022), but especially to the large stream of literature on bounding methods, with its growing focus on data-driven approaches that provide finite sample guarantees for performance bounds of stochastic optimisation problems (see, e.g., Guigues et al. (2017); Bertsimas and Koduri (2022); Gao (2023); Baardman et al. (2023); Sadana et al. (2024)).

The main contribution of this paper is the development of the first approach for constructing probabilistically guaranteed bounds on the true optimal value of stopping problems. Unlike existing upper- and lower-bound approaches, whose estimates are only guaranteed to be valid with respect to the optimal stopping problem's sample-based approximation, our approach results in performance bounds that are guaranteed to be valid with respect to the original optimal stopping problem. While the numerical values of our bounds also depend on the given set of sample paths, the associated probabilistic guarantees that we derive are valid with respect to the true underlying distribution, meaning our guaranteed bounds allow replacing the sampling error with a pre-specified confidence level. The key to achieving this is to bound the true (but difficult or impossible to determine) continuation function that appears in the problem's dynamic programming formulation by high- and low-biased regression estimates that are (probabilistically) guaranteed to over- and underestimate, respectively, the true conditional expectation. So, rather than directly approximating the exact continuation function, we use guaranteed over- and underestimates of the true continuation function at every stage of the backward recursive procedure to ensure that we obtain guaranteed upper and lower bounds, respectively, on the stopping problem's true optimal value. We illustrate the applicability of the proposed data-driven approach by valuing a Bermudan-style put option.

The secondary contribution of this paper is a new construction of non-asymptotic uniform confidence bands for regression functions based on the theory of reproducing kernel Hilbert spaces (RKHSs), and demonstrating how these can be leveraged to approximate continuation functions in a stochastic dynamic programming procedure. To the best of our knowledge, this newly established link between non-parametric statistics and OR is the first application of non-asymptotic uniform confidence bands that goes beyond a mere theoretical interest. There is a large body of statistical literature on constructing confidence bands for regression functions (see, e.g., Bjerve et al. (1985);

Knafl et al. (1985); Eubank and Speckman (1993); Sun and Loader (1994); Hall and Horowitz (2013); Cai et al. (2014)). To establish non-asymptotic uniform bands, some assumptions are required about the class of functions that contains the true regression function, f_0. Existing results are based on relatively strong assumptions, such as the monotonicity of f_0 (see Gauffriau et al. (2021)), the Lipschitz continuity of f_0 with known Lipschitz constant (see Lederer et al. (2019)), or the exact knowledge of the modulus of continuity of f_0 (see Jiang (2019)). Such assumptions allow to bound the form of the regression function between two observation points. More involved is the assumption used in Csáji and Horváth (2022) that f_0 is band-limited, i.e., its Fourier transform vanishes outside a compact interval. They also have to assume that the unknown f_0 is defined on the entire real line, even though their regression estimate $\hat{f}(x)$ is only defined for $x \in [0, 1]$. These assumptions are too restrictive for our purposes, so we use RKHS theory to develop novel finite sample guarantees in the form of kernel-based non-asymptotic uniform confidence bands for general regression problems.

As a final contribution, with the generalisable guaranteed bounding approach developed here we hope to have laid the groundwork for methodological extensions and applications to other data-driven decision problems. Importantly, our main regularity assumption that the true but unknown regression function is an element of an RKHS is relatively mild since the RKHS is dense in the space of continuous functions. No restrictive assumptions are made on structural properties – such as convexity or monotonicity – of the true regression function. Dommel and Pichler (2023) have recently studied a generalised stochastic optimisation problem whose objective function needs to be estimated and have shown the weak consistency of their uniform regression (function) estimator using the RKHS framework. Their asymptotic result will also follow as a direct corollary of the stronger finite sample guarantee developed in this paper. The idea to use the framework of RKHSs to construct confidence bands also appears in Csáji and Horváth (2022), where the Paley-Wiener kernel is (exclusively) used. However, their approach requires solving a quadratic optimisation problem for each point x_0 for which the bound is to be calculated. Moreover, their construction only works in the case where the distribution of the regressors (or inputs) is known. In contrast, we derive closed-form expressions for the probabilistic bounds both for the case where the regressors come from a known distribution and the case where they come from an unknown distribution. This makes our data-driven approach readily applicable in a range of situations including simulation.

The rest of this paper is organised as follows: In Section 2, we describe the fundamental properties of RKHSs and the construction of the kernel ridge regression (KRR) estimator. Section 3 contains our key statistical results in the form of probabilistic over- and underestimation guarantees for the cases where the underlying data-generating distribution is known or unknown. In Section 4 we derive our main results by applying these statistical results to develop a novel algorithmic procedure that provides guaranteed bounds for optimal stopping problems. The applicability of our bounding procedure is illustrated in Section 5 using the example of a Bermudan-style put option, which is the canonical optimal stopping problem in finance. Lastly, we provide concluding remarks and directions for future research in Section 6. Appendices A, B, and C contain proofs of lemmas, a summary of the notation used for probability distributions, and supplementary figures, respectively.

2. The nonparametric regression estimate

In this section, we develop the RKHS-based nonparametric regression estimator in preparation for the subsequent derivation of closed-form formulas for the probabilistic guarantees (Section 3), which are a key component of our guaranteed bounding procedure (Section 4). We begin with an overview of the well-developed theory of RKHS to make our work self-contained and state our assumptions. We then present the continuous and the empirical regularisation problem in Subsections 2.2 and 2.3, respectively, with the latter yielding the KRR estimator. The notion of reproducing kernels was originally introduced by Aronszajn (1950), and we refer the interested reader to the textbooks of Wahba (1990); Schölkopf and Smola (2002); Berlinet and Thomas-Agnan (2011) for further details. RKHS-based approaches and kernel methods in particular are well established in artificial intelligence and machine learning, and they have recently started to gain popularity in the OR literature (see, e.g., Ban and Rudin (2019); Bertsimas and Koduri (2022); Schmidt and Pibernik (2024); Sadana et al. (2024)). Appendix A contains all proofs of lemmas.

2.1. Overview of reproducing kernel Hilbert spaces and assumptions

Given some non-empty (but not necessarily compact) set $\mathcal{X} \subset \mathbb{R}^d$, let $k\colon \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ be a function with the following properties:

(P1) Symmetry: $k(x, y) = k(y, x)\, \forall x, y \in \mathcal{X}$;

(P2) Positive semi-definiteness:

$$\sum_{i,j=1}^{m} v_i v_j k(x_i, x_j) \geq 0 \quad \forall m \geq 1, v_i, v_j \in \mathbb{R}, x_i, x_j \in \mathcal{X}.$$

Then k is called a kernel function, Mercer kernel, or simply kernel. In this paper we consider kernels that additionally satisfy assumptions (A1)–(A5) for all $(x, y) \in \mathcal{X} \times \mathcal{X}$:

(A1) Translation invariance: $k(x, y)$ does only depend on $x - y$;

(A2) Boundedness: $0 \leq k(x, y) \leq 1$;

(A3) Normalisation: $k(x, x) = 1$.

Consider the space $\mathcal{H}^{\text{fin}}$ of all real-valued functions f of the form

$$f(\cdot) = \sum_{j=1}^{J} v_j k(x_j, \cdot)\,, \tag{1}$$

where $J < \infty$. Uniqueness of the representation in (1) is guaranteed by the following assumption:

(A4) Linear independence: For any finite collection of distinct points $\{x_1, \ldots, x_J\} \subseteq \mathcal{X}$, the functions $k(x_j, \cdot)$ are linearly independent on $\mathcal{X}$.

On the space of functions on $\mathcal{X}$, which is spanned by the set $\{k(x, \cdot) | x \in \mathcal{X}\}$, an inner product $\langle \cdot, \cdot \rangle_k$ is defined by

$$\langle k(x, \cdot), k(y, \cdot) \rangle_k := k(x, y). \tag{2}$$

Its bilinear extension is used for the space $\mathcal{H}^{\text{fin}}$. If $f(x) = \sum_{i=1}^{I} v_i k(x, y_i)$ and $g(x) = \sum_{j=1}^{J} w_j k(x, z_j)$ are functions in $\text{span}\{k(x,\cdot)|x \in \mathcal{X}\}$, then by linearity

$$\langle f, g\rangle_k = \sum_{i=1}^{I}\sum_{j=1}^{J} v_i w_j k(y_i, z_j)\,.$$

Consequently, the inner product $\langle\cdot,\cdot\rangle_k$ induces a norm $\|\cdot\|_k$ on the real inner product space $\mathcal{H}^{\text{fin}}$, defined by

$$\Big\|\sum_{i=1}^{I} v_i k(\cdot, y_i)\Big\|_k^2 := \sum_{i,j=1}^{I} v_i v_j \langle k(y_i,\cdot), k(y_j,\cdot)\rangle_k = \sum_{i,j=1}^{I} v_i v_j k(y_i, y_j) \geq 0\,,$$

where the non-negativity is a result of the positive semi-definiteness of the kernel k. The completion of $\mathcal{H}^{\text{fin}}$ with respect to the norm $\|\cdot\|_k$ is called the reproducing kernel Hilbert space (RKHS) $\mathcal{H}$. By construction, $\mathcal{H}^{\text{fin}} \subset \mathcal{H}$ and $\mathcal{H}^{\text{fin}}$ is dense in $\mathcal{H}$ (finite linear combinations are dense).

Let $f \in \mathcal{H}^{\text{fin}}$ be a finite linear combination as in (1) and set $g = k(\cdot, x)$. Then

$$\langle f, g\rangle_k = \langle\sum_{i=1}^{I} v_i k(x_i,\cdot), k(\cdot, x)\rangle_k = \sum_{i=1}^{I} v_i \langle k(x_i,\cdot), k(\cdot, x)\rangle_k = \sum_{i=1}^{I} v_i k(x_i, x) = f(x)\,. \tag{3}$$

By density of $\mathcal{H}^{\text{fin}}$ in $\mathcal{H}$, (3) is also valid for $f \in \mathcal{H}$. Thus, the inner product with $k(\cdot, x)$ acts as the evaluation functional in $\mathcal{H}$. Using the Cauchy-Schwartz inequality, one then gets $|f(x)| = |\langle k(x,\cdot), f(\cdot)\rangle_k| \leq \|k(x,\cdot)\|_k \cdot \|f\|_k$. This further implies

$$\sup_x |f(x)| = \|f\|_\infty \leq \|f\|_k\,,$$

since $\|k(x,\cdot)\|_k = \sqrt{k(x,x)} = 1$ holds for all $x \in \mathcal{X}$ by assumption (A3).

Finally, in order to cover a broad class of regression functions, we consider kernels that satisfy the following universality assumption (cf e.g. Sriperumbudur et al. (2011)):

(A5) Universality: The RKHS $\mathcal{H}$ is dense in the space of all continuous functions on $\mathcal{X}$, denoted by C_X, with respect to the supremum norm.

Remark 2.1. *Examples of commonly used functions satisfying properties (P1) and (P2) as well as assumptions (A1–A5) are the sinus kernel* $k(x, y) = \frac{\sin(\pi(x-y))}{\pi(x-y)}$ *(the Paley-Wiener kernel) and the widely used Gaussian kernel* $k(x, y) = \exp(-\alpha\|x - y\|^2)$ *(see Steinwart and Christmann (2008, Corollary 4.58)). The latter kernel, which is a universal kernel (Micchelli et al., 2006) and also know as squared-exponential or Gaussian radial basis function (RBF) kernel (Kanagawa et al., 2018), will also be utilised in our numerical examples in Section 5.*

2.2. The continuous problem and corresponding regularisation operator

For a given probability measure P on $\mathcal{X}$, consider the operator $\mathcal{K}$ on $\mathcal{H}$:

$$(\mathcal{K}f)(x) := \int_{\mathcal{X}} k(x,y)\, f(y)\, P(dy) = \langle k(x,\cdot), f(\cdot)\rangle_{P,2}\,.$$

To emphasise the dependence on the measure, we explicitly write $\langle\cdot,\cdot\rangle_{P,2}$ here. Only when there is no confusion possible, we omit the index P. However, we always keep the indices $\langle\cdot,\cdot\rangle_2$ and $\langle\cdot,\cdot\rangle_k$ to clearly distinguish between the two different inner products that we use. Notice that if P is discrete, then $\mathcal{K}f \in \mathcal{H}^{\text{fin}}$.

By the boundedness of k, the operator $\mathcal{K}$ maps $L^2(P)$ functions to $L^2(P)$ functions and is self-adjoint in L^2: $\langle \mathcal{K}f, g\rangle_2 = \langle f, \mathcal{K}g\rangle_2$. It holds that

$$\|\mathcal{K}f\|_k^2 = \left\| \int_{\mathcal{X}} k(\cdot,y) f(y)\ P(dy)\right\|_k^2 = \iint\limits_{\mathcal{X}\times\mathcal{X}} k(x,y) f(x) f(y)\, P(dx)\, P(dy)\,.$$

There is an important relation between $\langle\cdot,\cdot\rangle_k$ and $\langle\cdot,\cdot\rangle_2$:

Lemma 2.1. *For $f, g \in L^2(P) \cap \mathcal{H}$, the following relation between $\langle\cdot,\cdot\rangle_k$ and $\langle\cdot,\cdot\rangle_2$ holds:*

$$\langle f, g\rangle_{P,2} = \langle \mathcal{K}f, g\rangle_k = \langle f, \mathcal{K}g\rangle_k\,.$$

Since $\mathcal{K}$ is symmetric and positive semidefinite, it has orthonormal (in $\mathcal{H}$) eigenvectors u_i and nonnegative eigenvalues μ_i. If $v = \sum_i \alpha_i u_i$, then $\mathcal{K}v = \sum_i \alpha_i \mu_i u_i$. Notice that some eigenvalues μ_i may be zero and therefore $\mathcal{K}$ may be singular. However, $\mathcal{K} + \lambda I$, with I being the identity and $\lambda > 0$, is always invertible, since for $v = \sum_i \alpha_i u_i$ it holds that

$$[\mathcal{K} + \lambda I]^{-1} v = \sum\nolimits_i \alpha_i \frac{1}{\mu_i + \lambda I} u_i\,.$$

Moreover, as $\|u_i\|_k = 1$, it holds that

$$\|[\mathcal{K} + \lambda I]^{-1}\mathcal{K}v\|_k^2 = \sum\nolimits_i \alpha_i^2 \left(\frac{\mu_i}{\mu_i + \lambda}\right)^2 \le \sum\nolimits_i \alpha_i^2 = \|v\|_k^2\,, \tag{4}$$

and, consequently,

$$\|[\mathcal{K} + \lambda I]^{-1}\mathcal{K}\|_k \le 1\,. \tag{5}$$

The regularisation operator based on $\mathcal{K}$. For every $f \in \mathcal{H}$ and $\lambda > 0$, let $S_\lambda(f)$ be the minimiser in g of the following problem with respect to the underlying distribution P:

Opt (P) $$\min_g\ \|f - g\|_{P,2}^2 + \lambda \|g\|_k^2\,,$$

Since the objective is strictly convex, the solution of problem Opt (P) is unique. Notice the close relation of problem Opt (P) to the notion of the proximal operator (see Parikh and Boyd (2014)).

Lemma 2.2. *Using the relation $g = \mathcal{K}w$ one gets for the solution of problem Opt (P) that $w = [\mathcal{K} + \lambda I]^{-1} f$ and $S_\lambda(f) = [\mathcal{K} + \lambda I]^{-1}\mathcal{K}f$.*

By (4), the regularisation operator S_λ is contracting in $\mathcal{H}$ since $\|[\mathcal{K} + \lambda I]^{-1}\mathcal{K}f\|_k \leq \|f\|_k$. If $\mathcal{K}$ was invertible, then S_0 would be the identity. However, we can guarantee invertibility of $[\mathcal{K} + \lambda I]$ only for $\lambda > 0$. In the limit $\lambda \to 0$, the following result holds:

Lemma 2.3. *For any $f_0 \in \mathcal{K}\mathcal{H}$, it holds that $S_\lambda(f_0) \to f_0$ in k-norm, and hence also in supremum norm, as λ tends to zero.*

2.3. The empirical problem and kernel ridge regression

Let $(X_1, \ldots, X_n)$ be a set of n independent random variables from the regressor distribution P, i.e. $X_i \sim P$ and $(X_1, \ldots, X_n) \sim P^n$, with P^n being the distribution of the sample points. We assume that P has no atoms. Let $(x_1, \ldots, x_n) \in \mathcal{X}^n$ be the set of realisations of $(X_1, \ldots, X_n)$, drawn independently from P, and let $\hat{P}_n = \frac{1}{n}\sum_i \delta_{x_i}$ be the pertaining (discrete) empirical distribution. Notice that $\hat{P}_n$ is a random object governed by P^n since it depends on the observed data $(x_1, \ldots, x_n)$, which are realisations of $(X_1, \ldots, X_n)$. For a given sample $\boldsymbol{X} = (X_1, \ldots, X_n)$, let $\mathcal{H}_X$ be the linear subspace of $\mathcal{H}$ generated by $(k(\cdot, X_1), \ldots, k(\cdot, X_n))$. This is a finite-dimensional subspace of $\mathcal{H}$ and any function $\sum_{i=1}^n v_i k(x, X_i)$ is characterised by the coefficients $v = (v_1, \ldots, v_n)^\top$. We assume that the sample points are distinct and therefore, by assumption (A4), the coefficients are unique.

We distinguish two different cases:

- Fixed design: The distribution of regressors X_i is known such that the Wasserstein distance $d_W(P, \hat{P}_n)$ can be computed.
- Random design: The regressors represent an i.i.d. sample from an unknown distribution P.

We use the same regression estimate in the two cases, but the closed-form formulas for the probabilistic guarantees that we derive will involve slightly different terms. For fixed design, $(X_1, \ldots, X_n) \sim P^n$ typically represents n i.i.d. samples from P, but in some situations the regressors X_i may be chosen deterministically in order to minimise $d_W(P, \hat{P}_n)$. We refer to the book of Pflug and Pichler (2014) for algorithms to solve the so-called optimal quantisation problem, which refers to constructing a discrete distribution that minimises the Wasserstein distance to a given P. Their book also deals with the problem of computing $d_W(P, \hat{P}_n)$ when $(X_1, \ldots, X_n)$ is a random draw from P^n.

The empirical version of $\mathcal{K}$ is called $\hat{\mathcal{K}}_n$, i.e.,

$$(\hat{\mathcal{K}}_n f)(x) := \int_{\mathcal{X}} k(x, y)\, f(y)\, \hat{P}_n(dy) = \frac{1}{n}\sum_{j=1}^n k(x, X_j) f(X_j) = \langle k(x, \cdot), f(\cdot)\rangle_{\hat{P}_n, 2}$$

and

$$\|\hat{\mathcal{K}}_n f\|_k^2 = \frac{1}{n^2}\sum_{i=1}^n \sum_{j=1}^n f(X_i) f(X_j) k(X_i, X_j) = \iint_{\mathcal{X}\times\mathcal{X}} k(x, y) f(x) f(y)\, \hat{P}_n(dx)\, \hat{P}_n(dy)\,.$$

The operator $\hat{\mathcal{K}}_n$ is self-adjoint and maps $\mathcal{H}_X$ to $\mathcal{H}_X$. Introducing the $n \times n$ symmetric positive semi-definite matrix K_n with entries $K_n(i,j) := \frac{1}{n}k(X_i, X_j)$, one sees that

$$\left\| \sum_{i=1}^{n} v_i k(\cdot, X_i) \right\|_k^2 = n \cdot v^\top K_n v \,. \tag{6}$$

The matrix nK_n is called *Gram matrix* or *kernel matrix*. The operator $\hat{\mathcal{K}}_n$ acts on the coefficients in the following way:

$$\hat{\mathcal{K}}_n \Big(\sum_{i=1}^{n} v_i \cdot k(x, X_i) \Big) = \sum_{i=1}^{n} \left(K_n v\right)_i \cdot k(x, X_i) \,. \tag{7}$$

Similarly, $[K_n + \lambda I]^{-1} \hat{\mathcal{K}}_n \big(\sum_{i=1}^{n} v_i \cdot k(x, X_i) \big) = \sum_{i=1}^{n} \left([K_n + \lambda I]^{-1} K_n v \right)_i \cdot k(x, X_i)$, and analogous to (5), we get

$$\| [K_n + \lambda I]^{-1} K_n \|_k \leq 1 \,. \tag{8}$$

The empirical regularisation operator based on $\hat{\mathcal{K}}_n$. The following result for the empirical distribution can be obtained in a similar way as the solution of problem Opt (P), and by evoking the famous *Representer Theorem*, see Schölkopf et al. (2001).

Theorem 2.1 (Representer Theorem)**.** *The function $\hat{g}$ that solves the regularisation problem*

$$\text{Opt } (\hat{P}_n) \qquad \min_{g \in \mathcal{H}} \|f - g\|_{\hat{P}_n,2}^2 + \lambda \cdot \|g\|_k^2 \,,$$

where $\| \cdot \|_{\hat{P}_n,2}^2$ is the square loss and $\lambda > 0$ is a regularisation constant controlling the smoothness of g, is unique and can be found in $\mathcal{H}_X$. It has the representation $\hat{g} = \sum_{i=1}^{n} v_i k(\cdot, X_i)$, where the vector $v = (v_1, \ldots, v_n)^\top$ satisfies $v = [K_n + \lambda I]^{-1} \phi$, with $\phi = (f(X_1), \ldots, f(X_n))^\top$.

The regularisation problem Opt $(\hat{P}_n)$ with respect to the empirical distribution is an optimisation problem over the potentially infinite dimensional function space $\mathcal{H}$. However, Theorem 2.1 states that the solution of this regularised empirical risk minimisation problem can be found in the finite dimensional space $\mathcal{H}_X$. It is given as a weighted sum of so-called feature vectors $k(\cdot, X_i)$. Note that while the form of the obtained RKHS estimator $\hat{g}$ is identical to the well-known kernel ridge regression (KRR) estimator, we only permit kernel functions k which lead to a RKHS that is dense in the space of continuous functions (see Assumption (A5)). Nevertheless, to ensure consistency with the vast majority of the kernel methods literature (e.g. see Kanagawa et al. (2018) and the discussion therein), in the following we refer to the function $\hat{g}$ as the KRR estimator.

Lemma 2.4. *Assume that f and k are bounded Lipschitz functions with Lipschitz constants $L(f)$ and $L(k)$, respectively. Then, it holds that*

$$\|\mathcal{K}f - \hat{\mathcal{K}}_n f\|_{P,2} \leq \|\mathcal{K}f - \hat{\mathcal{K}}_n f\|_\infty \leq c_L(f,k) d_W(P, \hat{P}_n) \,,$$

where $d_W(P, \hat{P}_n)$ *is the Wasserstein distance defined by*

$$d_W(P, \hat{P}_n) := \inf \left\{ \iint \|x - y\| \, \pi(dx, dy) : \pi \text{ is a bivariate distribution with marginals } P \text{ and } \hat{P}_n \right\}$$

and $c_L(f, k)$ *is the Lipschitz constant of* $y \mapsto k(x, y)f(y)$*, with* $c_L(f, k) \leq L(f) + L(k)\|f\|_\infty$*.*

Let the true regression function f_0 be continuous on $\mathcal{X}$. Given our assumptions, f_0 has a representation of the form

$$f_0(x) = (\mathcal{K}w_0)(x) = \int_{\mathcal{X}} k(x, y)w_0(y) \, P(dy) \, .$$

We consider a standard "signal plus noise" (or additive error) model and, for simplicity, restrict ourselves now to the univariate case ($d = 1$). So, for each one-dimensional $X_i, i = 1, \ldots, n$, we observe a Y_i such that $Y_i = f_0(X_i) + \varepsilon_i$, where ϵ_i denotes random noise. We suppose the standard assumptions that $\mathbb{E}[\varepsilon_i] = 0, \mathbb{E}[\varepsilon_i^2] \leq \sigma^2$ for all $i = 1, \ldots, n$, and that the ε_i's are conditionally independent given the X_i (cf. e.g., the discussion in Kanagawa et al. (2018)). Note that we require the variances of the ϵ_i's to only be bounded rather than known exactly. Let $(y_1, \ldots, y_n) \in \mathbb{R}^n$ be the realisation of the set of random variables $(Y_1, \ldots, Y_n)$. The set of pairs $(x_i, y_i)_{i=1}^n \subset \mathcal{X} \times \mathbb{R}$, which is typically referred to as training data or training set (of observations), can then be used to obtain an estimate of the regression function f_0. Set $\boldsymbol{Y} = (Y_1, \ldots, Y_n)^\top$ and $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_n)^\top$.

To develop finite sample results with respect to the true underlying probability distribution, a careful distinction is required between the different probability distributions that exist in our setting. In addition to those introduced at the beginning of this subsection, here $\bar{P}$ denotes the bivariate distribution of the pair of random variables (X, Y), $\bar{P}^n$ denotes the joint distribution of the random sample $((X_1, Y_1), \ldots, (X_n, Y_n))$, and $\bar{P}_X^n$ denotes the conditional sample distribution given the regressors $X_1, \ldots, X_n$. Notice that the distribution P (P^n) introduced at the beginning of this subsection then corresponds to the marginal distribution of component X (vector $(X_i)_{i=1}^n$) with respect to $\bar{P}$ ($\bar{P}^n$). For a real-valued random output variable Y and a random input variable X from the joint distribution $\bar{P}$, the regression function $f_0(x)$ at any point $x \in \mathcal{X}$ is given by the conditional expectation of Y given value x of the input X, i.e. $f_0(x) = \mathbb{E}[Y|X = x]$. It is important to note that no assumption is made about the distribution of $Y|X$, denoted by $\bar{P}_X$. Table B.2 in Appendix B summarises all the different symbols used for probability distributions in this work.

Finally, our KRR estimator $\hat{f}_n \in \mathcal{H}_X \subseteq \mathcal{H}^{\text{fin}} \subset \mathcal{H}$, which is a nonparametric regression estimate of f_0, is then given by

$$\hat{f}_n(x) = \frac{1}{n} \sum_{i=1}^{n} \left([K_n + \lambda I]^{-1} \boldsymbol{Y} \right)_i \cdot k(x, X_i) \, . \tag{9}$$

3. Uniform confidence bands for regression functions

In this section we address the statistical problem of constructing non-asymptotic uniform confidence bands for the true but unknown regression function. The closed-form formulas for the

probabilistic over- and underestimation guarantees that we derive here are a key component of our guaranteed bounding procedure for optimal stopping problems (Section 4). We begin with the fixed design case, where the underlying regressor distribution P is known.

Theorem 3.1 (Fixed design)**.** *Let $\hat{f}_n(\cdot)$ be the nonparametric regression estimate (9) in the case where the regressors X_i come from a known distribution. Then, for a given error level $0 < \beta < 1$, it holds that*

$$\bar{P}^n \left[\exists x : f_0(x) \geq \hat{f}_n(x) + C \right] \leq \beta \,,$$

where

$$C = \sqrt{\frac{\sigma^2 \operatorname{tr}(M_\lambda)}{n\beta}} + c_L(w_0, k) d_W(P, \hat{P}_n) \left(1 + \frac{1}{2\sqrt{n\lambda}} \right) + \frac{\sqrt{\lambda}}{2\sqrt{n}} \|w_0(\boldsymbol{X})\|_2 \,, \tag{10}$$

with $c_L(\cdot)$ defined as in Lemma 2.4, and $M_\lambda := [K_n + \lambda I]^{-1} K_n [K_n + \lambda I]^{-1}$.

Proof. Dissect the estimated regression function (9) into

$$\begin{aligned} \hat{f}_n(x) &= \frac{1}{n} \sum_{i=1}^{n} \left([K_n + \lambda I]^{-1} \boldsymbol{Y} \right)_i \cdot k(x, X_i) \\ &= \hat{m}_n(x) + \hat{R}_{\lambda,n}(x) \,, \end{aligned}$$

where

$$\hat{m}_n(x) = \frac{1}{n} \sum_{i=1}^{n} \left([K_n + \lambda I]^{-1} ((\mathcal{K} w_0)(\boldsymbol{X})) \right)_i \cdot k(x, X_i)$$

represents the main part (with $(\mathcal{K} w_0)(\boldsymbol{X}) = ((\mathcal{K} w_0)(X_1), \ldots, (\mathcal{K} w_0)(X_n))^\top$ and

$$\hat{R}_{\lambda,n}(x) = \frac{1}{n} \sum_{i=1}^{n} \left([K_n + \lambda I]^{-1} \varepsilon \right)_i \cdot k(x, X_i)$$

represents the error part. Define

$$\bar{f}_{0,n}(x) := (\hat{\mathcal{K}}_n w_0)(x) = \int_{\mathcal{X}} k(x, y) w_0(y) \hat{P}_n(dy) = \frac{1}{n} \sum_{i=1}^{n} k(x, X_i) \cdot (w_0(X_i)) \,,$$

Then $\bar{f}_{0,n} \in \mathcal{H}_X \subseteq \mathcal{H}^{\text{fin}} \subset \mathcal{H}$. Using the triangle inequality and the fact that $\|\cdot\|_\infty \leq \|\cdot\|_k$, we get

$$\|\hat{f}_n - f_0\|_\infty = \|\hat{m}_n + \hat{R}_{\lambda,n} - f_0 + \bar{f}_{0,n} - \bar{f}_{0,n}\|_\infty \leq \|\bar{f}_{0,n} - f_0\|_\infty + \|\bar{f}_{0,n} - \hat{m}_n\|_k + \|\hat{R}_{\lambda,n}\|_k \,. \tag{11}$$

We now derive upper bounds for each of the three terms on the right hand side of (11).

(i) By Lemma 2.4, we know that

$$\|f_0 - \bar{f}_{0,n}\|_\infty \leq c_L(w_0, k) \cdot d_W(P, \hat{P}_n) \,. \tag{12}$$

(ii) For the second summand, we get

$$
\begin{aligned}
&\|\bar{f}_{0,n} - \hat{m}_n\|_k \\
&= \left\| \bar{f}_{0,n} - \frac{1}{n}\sum_{i=1}^{n} \big([K_n + \lambda I]^{-1}((\mathcal{K}w_0)(\boldsymbol{X}))\big)_i \cdot k(\cdot, X_i) \right\|_k \\
&= \left\| \frac{1}{n}\sum_{i=1}^{n} k(\cdot, X_i) \cdot \big(w_0(\boldsymbol{X}) - [K_n + \lambda I]^{-1}(\mathcal{K}w_0)(\boldsymbol{X})\big)_i \right\|_k \\
&= \left\| \frac{1}{n}\sum_{i=1}^{n} k(\cdot, X_i)\big([K_n + \lambda I]^{-1}[K_n + \lambda I]w_0(\boldsymbol{X})\big)_i - ([K_n + \lambda I]^{-1}(\mathcal{K}w_0)(\boldsymbol{X}))_i \right\|_k \\
&\leq \lambda \left\| \frac{1}{n}\sum_{i=1}^{n} k(\cdot, X_i) \cdot \big([K_n + \lambda I]^{-1} w_0(\boldsymbol{X})\big)_i \right\|_k \\
&\quad + \left\| \frac{1}{n}\sum_{i=1}^{n} k(x, X_i) \cdot \big([K_n + \lambda I]^{-1}\left(K_n \cdot w_0(\boldsymbol{X}) - (\mathcal{K}w_0)(\boldsymbol{X})\right)\big)_i \right\|_k \\
&\leq \frac{\lambda}{\sqrt{n}} \sqrt{w_0(\boldsymbol{X})^\top M_\lambda w_0(\boldsymbol{X})} + \frac{1}{\sqrt{n}} \sqrt{\left(K_n \cdot w_0(\boldsymbol{X}) - ((\mathcal{K}w_0)(\boldsymbol{X}))\right)^\top M_\lambda \left(K_n \cdot w_0(\boldsymbol{X}) - (\mathcal{K}w_0)(\boldsymbol{X})\right)} \\
&\leq \frac{1}{\sqrt{n}} \sqrt{\mu^{\max}(M_\lambda)}\left(\lambda \|w_0(\boldsymbol{X})\|_2 + \|K_n \cdot w_0(\boldsymbol{X}) - (\mathcal{K}w_0)(\boldsymbol{X})\|_2\right) \\
&\leq \frac{1}{2\sqrt{n\lambda}} \left(\lambda \|w_0(\boldsymbol{X})\|_2 + c_L(w_0, k) d_W(P, \hat{P}_n)\right) .
\end{aligned}
$$

Here we used Lemma 2.4 and the fact that $(\hat{\mathcal{K}}_n w_0)(\boldsymbol{X}) = K_n \cdot w_0(\boldsymbol{X})$ by (7). We also used the fact that the eigenvalues of M_λ are $\mu_i/(\mu_i + \lambda)^2$, where μ_i are the eigenvalues of K_n. Notice that the maximal value of $\mu \mapsto \mu/(\mu + \lambda)^2$ is $1/(4\lambda)$.

(iii) For including the random design case in the calculation of the bound, we consider here the conditional expectation $\mathbb{E}_X$ (with respect to $\bar{P}_X^n$) given the sample $\boldsymbol{X} = (X_1, \ldots, X_n)$, but remark that for the fixed design case, this is just the normal expectation (with respect to $\bar{P}^n$). We get that

$$
\begin{aligned}
\mathbb{E}_X\left[\left\|\hat{R}_{\lambda,n}(\cdot)\right\|_k^2\right] &= \mathbb{E}_X\left[\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\left([K_n + \lambda I]^{-1}\varepsilon\right)_i \cdot k(X_i, X_j) \cdot \left([K_n + \lambda I]^{-1}\varepsilon\right)_j\right] \\
&= \frac{1}{n}\mathbb{E}_X\left[\varepsilon^\top [K_n + \lambda I]^{-1} K_n [K_n + \lambda I]^{-1}\varepsilon\right] \\
&= \frac{1}{n}\mathbb{E}_X\left[\sum_{i=1}^{n}\sum_{j=1}^{n}\varepsilon_i M_\lambda(i,j)\varepsilon_j\right] \leq \frac{\sigma^2}{n}\operatorname{tr}(M_\lambda) ,
\end{aligned}
$$

where we used (6) as well as the conditional independence of the error variables ε_i and their bounded variances. By the generalised Markov inequality of second order, we then get (for

$\bar{P}_X^n$ or $\bar{P}^n$) that

$$\bar{P}_X^n\big[\|\hat{R}_{\lambda,n}(\cdot)\|_k \geq \eta\big] \leq \mathbb{E}_X\big[\|\hat{R}_{\lambda,n}(\cdot)\|_k^2\big]/\eta^2 \leq \frac{\sigma^2}{n\eta^2}\,\mathrm{tr}(M_\lambda)\,. \tag{13}$$

Finally, choosing

$$\eta = \sqrt{\frac{\sigma^2 tr(M_\lambda)}{n\beta}}$$

and putting the pieces (i)–(iii) together, we get that

$$\begin{aligned}
&\bar{P}^n\left[\exists x : f_0(x) \geq \hat{f}_n(x) + C\right] \\
&\leq \bar{P}^n\left[\|\hat{f}_n - f_0\|_\infty \geq C\right] \\
&\leq \bar{P}^n\left[\|\bar{f}_{0,n} - f_0\|_\infty + \|\bar{f}_{0,n} - \hat{m}_n\|_k + \|\hat{R}_{\lambda,n}\|_k \geq C\right] \\
&\leq \bar{P}^n\left[c_L(w_0,k)d_W(\hat{P}_n,P)\left(1+\frac{1}{2\sqrt{n\lambda}}\right) + \frac{\sqrt{\lambda}}{2\sqrt{n}}\|w_0(\boldsymbol{X})\|_2 + \|\hat{R}_{\lambda,n}\|_k \geq C\right] \\
&= \bar{P}^n\left[\|\hat{R}_{\lambda,n}\|_k \geq \eta\right] \leq \frac{\sigma^2}{n\eta^2}\,\mathrm{tr}(M_\lambda) \;= \beta\,.
\end{aligned}$$

□

Corollary 3.1 (Weak consistency)**.** *Let the regressors X_i be chosen according to a deterministic algorithm, in such a way that $d_W(P,\hat{P}_n) \to 0$ as $n \to \infty$. Then, it holds that $C = C(n) \xrightarrow{n\to\infty} 0$ in Theorem 3.1, for any $0 < \beta < 1$. Thus, the regression estimate $\hat{f}_n(x)$ uniformly converges to the true regression function $f_0(x)$ (in probability), as $n \to \infty$.*

We now turn to the random design case. Here the distance $d_W(P,\hat{P}_n)$ is unknown and a random quantity. Its quantiles, however, can be bounded by existing concentration inequalities, such as the following one established by Fournier and Guillin (2015, Theorem 2):

$$\bar{P}^n\left[d_W(P,\hat{P}_n) \geq \gamma\right] \leq c_1 \exp\left(-c_2 n\gamma^2\right)\,, \tag{14}$$

where the regressors X_i, $i = 1,\ldots,n$, are an i.i.d. sample from P with bounded support $\mathcal{X}$, and c_1 and c_2 are known constants. Note that in the random design case there is an additional uncertainty source stemming from the sampling of the design, which typically makes the confidence regions larger than in the fixed design case. Nevertheless, a similar result to above can be established:

Theorem 3.2 (Random design)**.** *Consider the nonparametric regression estimate $\hat{f}_n(\cdot)$ of the true (but unknown) regression function $f_0(\cdot)$, as given in* (9)*, where the regressors X_i result from a random sampling procedure from an unknown marginal distribution P. Then, for a given error level $0 < \beta < 1$, it holds that*

$$\bar{P}^n\left[\exists x : f_0(x) \geq \hat{f}_n(x) + C^*\right] \leq \beta\,,$$

where C^* is a positive root (in C) of the equation

$$c_1 \exp\left(-c_2 n \frac{\frac{C}{2} - \frac{\sqrt{\lambda}}{2\sqrt{n}}\|w_0(\boldsymbol{X})\|_2}{\left(1+\frac{1}{2\sqrt{n\lambda}}\right) c_L(w_0,k)}\right) + \frac{4\sigma^2}{nC^2}\operatorname{tr}(M_\lambda) - \beta = 0\,, \tag{15}$$

with constants c_1, c_2 following from (14).

Proof. We may repeat steps (i)- (iii) of the proof of Theorem 3.1. In the summarising step, we use the following bound

$$\begin{aligned}
&\bar{P}_X^n\left[\exists x : f_0(x) \geq \hat{f}_n(x) + C\right] \\
&\leq \bar{P}_X^n\left[c_L(w_0,k) d_W(P,\hat{P}_n)\left(1+\frac{1}{2\sqrt{n\lambda}}\right) + \frac{\sqrt{\lambda}}{2\sqrt{n}}\|w_0(\boldsymbol{X})\|_2 \geq \frac{C}{2}\right] + \bar{P}_X^n\left[\|\hat{R}_{\lambda,n}\|_k \geq \frac{C}{2}\right] \\
&= \bar{P}_X^n\left[d_W(P,\hat{P}_n) \geq \frac{\frac{C}{2} - \frac{\sqrt{\lambda}}{2\sqrt{n}}\|w_0(\boldsymbol{X})\|_2}{\left(1+\frac{1}{2\sqrt{n\lambda}}\right) c_L(w_0,k)}\right] + \bar{P}_X^n\left[\|\hat{R}_{\lambda,n}\|_k \geq \frac{C}{2}\right] \\
&\leq c_1 \exp\left(-c_2 n \frac{\frac{C}{2} - \frac{\sqrt{\lambda}}{2\sqrt{n}}\|w_0(\boldsymbol{X})\|_2}{\left(1+\frac{1}{2\sqrt{n\lambda}}\right) c_L(w_0,k)}\right) + \frac{4\sigma^2}{nC^2}\operatorname{tr}(M_\lambda)\,,
\end{aligned} \tag{16}$$

where the last inequality follows from (14) (for the first term) and (13) (for the second term). Setting the right hand side in (16) equal to the target confidence level β and solving the equation in C then gives the desired result. Since the conditional probability given $\boldsymbol{X}$ is smaller than β for all samples $\boldsymbol{X}$, the inequality also holds true unconditionally (i.e. with respect to $\bar{P}^n$). □

Remark 3.1. *The upper bound estimates obtained above are valid for all possible regressors and not only for those that were observed. This becomes important for procedures where the regression function is evaluated at freshly generated sample points (e.g. in an out-of-sample test).*

While the above results are formulated in terms of an upper bound, we have, in fact, developed a non-asymptotic uniform (or simultaneous) confidence band $\hat{f}_n(\cdot) \pm C$ with worst-case coverage probability (or confidence level) $1-\beta$:

Corollary 3.2 (Uniform confidence band)**.** *For a given error level $0 < \beta < 1$ and corresponding constant C defined in* (10) *(or* (15) *for the random design case), it holds that*

$$\bar{P}^n\left[\hat{f}_n(x) - C \leq f_0(x) \leq \hat{f}_n(x) + C \ \forall x\right] \geq 1-\beta\,.$$

To numerically illustrate the constant-width confidence bands proposed in Corollary 3.2 and the effect of varying sample sizes, Appendix C contains an application to a regression problem.

4. Guaranteed bounds for optimal stopping problems

In this section, we utilise the non-asymptotic uniform confidence band presented in the previous section to develop a novel procedure that provides (probabilistically) guaranteed upper and lower bounds on the true optimal value of stopping problems.

4.1. Problem formulation

In its standard form, optimal stopping problems feature a single binary decision variable (to stop or to wait), as well as one controllable state (with state space $\{\textit{waiting}, \textit{stopped}, \textit{expired}\}$) and one information state variable (e.g., the price of an underlying asset). Both state variables together represent the necessary information to make a decision and calculate the expected discounted payoff (or reward) when stopping the stochastic reward process, assuming it has not previously been stopped or become *expired*. Consider a discrete-time setting with a finite time horizon T and a set $\mathcal{T} := \{1, \ldots, T\}$ of admissible stopping times. Let $\{S_0, S_1, \ldots, S_T\}$ be a real-valued stochastic process that describes the information available at all times $t \in \mathcal{T}$. The value at time $t = 0$, S_0, is assumed to be known. Furthermore, let r be the periodic discount rate and denote the payoff (reward) for stopping when in state S_t at time t by $c_t(S_t)$.

The goal is to find the optimal stopping time $\tau^* \in \mathcal{T}$ that results in the maximised expected discounted payoff $V_0(S_0)$, that is:

$$V_0(S_0) = \sup_{\tau \in \mathcal{T}} \mathbb{E}\left[e^{-r\tau} c_\tau(S_\tau) | S_0\right] . \tag{17}$$

It is well-known (see, e.g., Glasserman (2004, Chapter 8)) that an optimal stopping time τ^*, which achieves the supremum in (17), exists in the finite horizon discrete-time setting. Hence, for consistency with the notation used throughout this paper, we may replace the sup in (17) with a max operator. For notational simplicity, we suppress the discount factor in our presentation of algorithmic procedures, but note that the formulations can be generalised straightforwardly to account for discounting.

It is also well known that the optimal stopping policy and, ultimately, the optimal value of problem (17) can be determined recursively using backward stochastic dynamic programming (SDP). Let $V_t(S_t)$ denote the value of the optimal stopping problem when in (information) state S_t at time $t \in \mathcal{T}$. Then, at the terminal time $t = T$, the value of the stopping problem is given by:

$$V_T(S_T) = c_T(S_T) . \tag{18}$$

Proceeding in a backward recursive manner, the value at times $t = T-1, T-2, \ldots, 1$ is given by:

$$V_t(S_t) = \max\left\{c_t(S_t), \mathbb{E}[V_{t+1}(S_{t+1})|S_t]\right\} . \tag{19}$$

Finally, assuming without loss of generality that no decision can be made at time $t = 0$, the optimal

value of the stopping problem is given by:

$$V_0(S_0) = \mathbb{E}[V_1(S_1)|S_0]\,. \tag{20}$$

Based on Bellman's principle of optimality, the values given by (17) and (20) coincide. Assuming $c_T(\cdot) \geq 0$, it is clear that $V_0(S_0) \geq 0$ holds for all S_0, meaning that the optimal value of the stopping problem is non-negative.

4.2. The simulation-and-regression approach

Let $(S_0, S_1^{(j)}, \ldots, S_T^{(j)})$, $j = 1, 2, \ldots, n$, be n independent (simulated) sample paths from the process (S_t). The empirical distribution of the set of sample realizations at time t, $\left(S_t^{(1)}, S_t^{(2)}, \ldots, S_t^{(n)}\right)$, is thus a sampled approximation of the (continuous) distribution of the random variable S_t. The key idea behind simulation-and-regression-based approaches is to approximate the (unknown) continuation function[3] within a backward recursive procedure, by regressing realised continuation values on (basis) functions of realised scenario values at the preceding time step. In particular, the function $\Phi_t(s) = \mathbb{E}[V_{t+1}(S_{t+1})|S_t = s]$ is approximated by a regression estimate based on the n available data pairs of the form (X_i, Y_i) with $X_i = S_t^{(i)}$ and $Y_i = V_{t+1}(S_{t+1}^{(i)})$. For each state, its associated value is then determined by comparing the estimated expected continuation value with the reward from immediate stopping. Backward iteration eventually yields an approximation of the value $V_0(S_0)$, which (under certain assumptions) can be shown to converge to true value as the number of sampled paths of the scenario process (S_t) tends to infinity. For details, we refer the reader to the excellent book of Glasserman (2004) and to the classical papers of Carriere (1996); Tsitsiklis and Van Roy (2001); Longstaff and Schwartz (2001).

It is well known that, depending on the update rule applied in simulation-and-regression-based approaches, the obtained approximation is typically either low-biased (due to the suboptimality of the stopping rule used when updating with realised continuation values) or unpredictably-biased (when updating using estimated continuation values, since the regression-based approximation may under- and/or over-estimate the true continuation function). For example, while approaching the true value $V_0(S_0)$ in the limit as $n \to \infty$, the algorithmic approach of Longstaff and Schwartz (2001) typically yields a low estimator because it underestimates the true value using realised continuation values under a sub-optimal stopping rule. On the other hand, the sign of the bias of the regression-based approach of Tsitsiklis and Van Roy (2001) is typically not predictable as their estimator relies on the quality of the regression functions, whose accuracy depends on the basis functions chosen and which may over- and under-estimate the true continuation function for different values of the sampled stock price paths. Carrying out an out-of-sample analysis by running the regression-based algorithm using a second set of sample paths addresses the in-sample bias but also leads, as noted by Glasserman (2004), to a low-biased estimator of the true but unknown objective value.

[3]The continuation function is the expected profit if the process is not stopped.

4.3. The guaranteed bounding procedure

Following the regression-based approach just described, the statistical results developed in Section 3 can be utilised to design an algorithmic procedure that yields (probabilistically) guaranteed performance bounds on the true optimal value of the stopping problem. The key to obtain guaranteed upper and lower bounds is to uniformly over- and under-estimate, respectively, the true regression function $\Phi_t(\cdot)$ (with a given confidence) in each step of the backward recursion. The procedure's main steps to generate an upper bound can be summarised as follows:

(i) At the terminal time $t = T$, for each path $j \in \{1, 2, \ldots, n\}$, evaluate:

$$\hat{V}_T\big(S_T^{(j)}\big) = c_T\big(S_T^{(j)}\big)\,.$$

(ii) For $t = T-1, T-2, \ldots, 1$, follow the backward recursion:

$$\hat{V}_t\big(S_t^{(j)}\big) = \begin{cases} c_t\big(S_t^{(j)}\big) & \text{if } c_t\big(S_t^{(j)}\big) > \hat{\Phi}_t\big(S_t^{(j)}\big) \\ \hat{\Phi}_t\big(S_t^{(j)}\big) & \text{if } c_t\big(S_t^{(j)}\big) \leq \hat{\Phi}_t\big(S_t^{(j)}\big), \end{cases}$$

where $\hat{\Phi}_t\big(S_t^{(j)}\big)$ is a high-biased approximation (i.e., an over-estimate) of $\mathbb{E}[\hat{V}_{t+1}(S_{t+1})|S_t^{(j)}]$, calculated based on the data pairs $(X_j = S_t^{(j)}, Y_j = \hat{V}_{t+1}\big(S_{t+1}^{(j)}\big))$ according to Theorem 3.1 (or Theorem 3.2 for the random design case). Notice that the constant C (or C^*) appearing in Theorem 3.1 (Theorem 3.2) depends here on the set of realizations at time t, so it typically changes over time.

(iii) Finally, at time $t = 0$, the high-biased estimate of the optimal stopping problem's value is obtained by calculating a simple sample average over approximated time $t = 1$ values using the over-estimated one-step ahead (unconditional) expectation:

$$\hat{V}_0(S_0) = \frac{1}{n}\sum_{j=1}^{n} \hat{V}_1\big(S_1^{(j)}\big)\,.$$

An analogous procedure yields a guaranteed lower bound $\check{V}_0(S_0)$ on the true optimal value $V_0(S_0)$ of the stopping problem. It simply requires replacing $\hat{\Phi}_t\big(S_t^{(j)}\big)$ in step (ii) with its low-biased counterpart $\check{\Phi}_t\big(S_t^{(j)}\big)$, which is obtained by subtracting (instead of adding) the constant C (or C^*) in Theorem 3.1 (Theorem 3.2). The following result summarises the guaranteed confidence associated with bounds resulting from the above approach.

Theorem 4.1 (Guaranteed bounds)**.** *Consider the nonparametric regression-based estimates $\hat{V}_0(S_0)$ and $\check{V}_0(S_0)$ for the true (but unknown) optimal value $V_0(S_0)$, as determined by the above described procedures. Then, for given error levels $0 < \beta_t < 1$ at times $t = 1, \ldots, T-1$, it holds in the worst case of stage-wise dependence, that*

$$\mathbb{P}\left[\check{V}_0(S_0) \leq V_0(S_0) \leq \hat{V}_0(S_0)\right] \geq 1 - \sum_{t=1}^{T-1} \beta_t\,.$$

If, on the other hand, stage-wise independence can be assumed, then the probability on the right-hand side slightly improves to $\prod_{t=1}^{T-1}(1-\beta_t)$ *.*

Proof. The fact that $\hat{V}_0(S_0)$ and $\check{V}_0(S_0)$ represent guaranteed upper and lower bounds, respectively, on the true value $V_0(S_0)$ follows immediately from the fact that the recursive construction of the estimates ensures (probabilistically) guaranteed stage-wise high- and low-biased estimates. □

5. Numerical Example: Bounds for American-style put options

To demonstrate the applicability of our guaranteed bounding approach, we consider the example of valuing an American-style put option. An American put option entitles its holder with the right, but not the obligation, to sell the option's underlying asset at a predetermined price (the "strike price") at any time before maturity T (Hull, 2021). Based on the Fundamental Theorem of Asset Pricing, the fair value of a financial option corresponds to its expected payoff under a so-called risk-neutral probability measure. In the case of an American option, however, the payoff depends on the exercise strategy and determining the option's fair value thus requires determining an optimal exercise strategy. Therefore, the pricing problem boils down to solving an optimal stopping problem.

In practice, there are situations where the potential exercise times are limited to a predetermined discrete set of dates. Such options (which are frequently traded, but mostly over-the-counter) are often referred to as Bermudan options instead of American options. Clearly, the value of a Bermudan option with a large number of potential exercise times approximates the value of the corresponding American option. Given that solving optimal stopping problems analytically is generally not possible, meaning that there are no closed-form pricing formulas available for American-style put options, their value is typically approximated using numerical methods by way of determining the value of a Bermudan-style option with the same parameters and sufficiently frequent exercise dates.

Consider a Bermudan put option with strike price K and potential exercise times $t \in \mathcal{T} := \{1, 2, \ldots, T\}$, with the option expiring at time T (which is the option's maturity date). Further, assume that the price of the single underlying asset, S_t, has been sampled for all times $t \in \mathcal{T}$. Then, at maturity T, the value of an unexercised put option is:

$$V_T(S_T) = \max\{K - S_T, 0\}\,.$$

At each potential exercise time $0 < t < T$, the optionholder compares the immediate exercise value $\max\{K - S_t, 0\}$ with the expected continuation value, $\mathbb{E}\left[V_{t+1}(S_{t+1})|S_t\right]$. The value of the put option given state S_t at time t, $V_t(S_t)$, is thus recursively defined as

$$V_t(S_t) = \max\left\{\max\{K - S_t, 0\}, \mathbb{E}\left[V_{t+1}(S_{t+1})|S_t\right]\right\}\,.$$

In the approximative calculation, the inner conditional expectation is carried out for a finite sample and not for the entire continuous distribution. However, our guaranteed bounds for the conditional expectation must be valid for the underlying distribution and not only for the sample. This is where

our assumption that the unknown regression function lies in the RKHS comes into play as it allows us to make a statement regarding all possible regressor values – not just those that were observed.

In our numerical example, we consider a Bermudan put option in the Black-Scholes model with strike price $K = 105$, final expiration date $T = 3$, annualised risk-free rate $r = 0.02$, annualised volatility $\bar{\sigma} = 0.20$, and three different initial spot prices, $S_0 \in \{100, 105, 110\}$, representing in-the-money, at-the-money and out-of-the-money situations, respectively. Given that the option is exercisable at times $t \in \{1, 2, 3\}$ and assuming that $t = 3$ represents a final maturity date of one year, we simulated $n = 1000$ trajectories of a geometric Brownian motion (gBm) process according to the following discretised risk-neutral stock price dynamics:

$$S_{t+1} = S_t \exp\left\{\left(r - \frac{\bar{\sigma}^2}{2}\right)\Delta + \bar{\sigma}\sqrt{\Delta}\xi_{t+1}\right\},$$

where $\Delta = 1/T$ is the exercise frequency and ξ_{t+1} denotes a standard normal random variable (i.e. $\xi_{t+1} \overset{\text{iid}}{\sim} \mathcal{N}(0,1)$). Figure 1 shows 50 sampled trajectories (a scenario fan) of the stock price process using $S_0 = 100$. To construct our nonparametric regression functions, we employed the squared

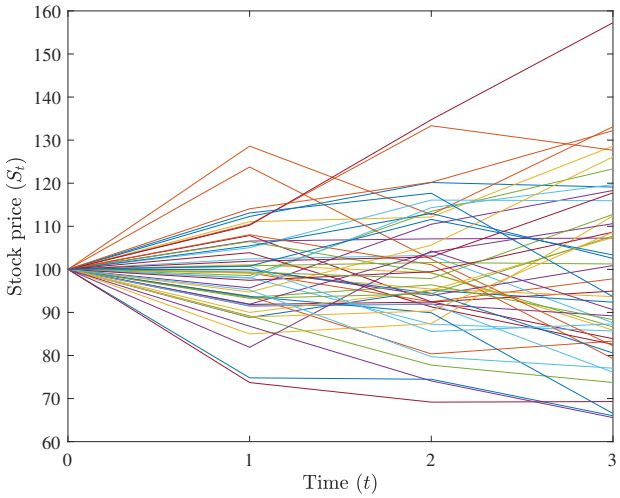


Figure 1: Selection of 50 equally likely trajectories of the discrete-time stock price process for $S_0 = 100$.

exponential kernel with parameter $\alpha = 0.01$ and set $\lambda = 0.01$. Given the set of sampled trajectories of the underlying asset, we can readily apply our proposed bounding procedure of Section 4, with confidence bands calculated according to the fixed design case.

The closed-form formulas for our probabilistic bounds involve several constants. However, it is important to point out that determining the bounds numerically does not require knowledge of the exact values of these constants but only upper bounds for their values. In practical applications upper bounds can often be inferred from structural properties such as side information about the problem at hand and problem-dependent features. If this is not possible, experiments may be undertaken to obtain a reliable overestimate of the constants. For the present application, we applied the following procedure: Given the relation $f_0 = \mathcal{K}w_0$, we first sampled the stock price paths and determined the corresponding continuation values to estimate both the matrix K_n and the

KRR estimator $\hat{f}_n$, before determining an estimate for w_0 by setting $w_0 \approx K_n^{-1}\hat{f}_n$. To determine an upper bound for $\|w_0(\boldsymbol{X})\|_2$, which is the quantity needed for our guaranteed bounds on the optimal objective value, we repeated this procedure 100 times and used a value larger than the largest observation plus twice the sample standard deviation. In a similar fashion, to obtain a conservative overestimate for the Lipschitz constant $c_L(w_0, k)$ we observed the largest difference in values of $\hat{f}_n$ over a fine grid and then repeated this procedure for each of the 100 generated scenario fans. While an upper bound for the noise variance σ can be obtained in a similar way, it is also possible here to use theoretical considerations in order to determine an analytic upper bound for σ in terms of the volatility $\bar{\sigma}$ of the stock price process, which is the driving source of uncertainty.

The resulting guaranteed bounds on the true value of the Bermudan put option, considering three different levels of both overall confidence and initial stock price, are shown in Table 1. For a rough

Table 1: Guaranteed bounds – lower bound (LB) and upper bound (UB) – for Bermudan put option prices considering three different overall confidence levels, $(1-\beta_1-\beta_2)$, and initial stock prices, S_0.

Spot price	Binomial lattice	Confidence 80% ($\beta_1=\beta_2=0.10$)			Confidence 90% ($\beta_1=\beta_2=0.05$)			Confidence 95% ($\beta_1=\beta_2=0.025$)		
S_0	American	LB	UB	Δ_{rel}[†]	LB	UB	Δ_{rel}	LB	UB	Δ_{rel}
100	9.900	9.412	10.196	7.69	9.281	10.365	10.46	9.102	10.623	14.32
105	7.465	6.882	7.852	12.35	6.737	8.058	16.39	6.545	8.350	21.62
110	5.527	4.965	5.968	16.81	4.799	6.183	22.38	4.577	6.487	29.44

[†] Relative performance gap (in %): Δ_{rel}=(UB-LB)/UB*100.

comparison, the table also gives the estimated (non guaranteed) fair value of the corresponding American-style option, which was determined using a standard Binomial lattice with 1000 steps. Notice, however, that the price for a Bermudan option is always lower than the price for the corresponding American option (since the exercise times are restricted), so their prices should not be compared directly[4]. It can be seen that the gap between low- and high-biased option prices or, in other words, the difference between lower bound (LB) and upper bound (UB), increases in both the initial stock price, S_0, and in the overall required level of confidence, $(1-\beta_1-\beta_2)$. The former effect is largely due to the more frequent use of the very conservatively over and underestimated continuation functions, as immediate exercise becomes less likely the higher the stock price relative to the option's strike price.

By contrast, the widening of the performance gap in the required overall confidence level can be attributed to the fact that a higher probability for the true value $V_0(S_0)$ being in $[\check{V}_0(S_0), \hat{V}_0(S_0)]$ necessitates wider confidence bands at every possible exercise time, given the recursive construction of our uniform over- and underestimates. Indeed, as can be seen from Figure 2, which illustrates

[4]More generally, it should be emphasised that since there are no other constructions of guaranteed bounds for stopping problems in the existing literature – numerical values resulting from any other available approach are only valid with respect to the sampled stock price paths, so do not come with a probabilistic guarantee –, there exist no benchmarks that would allow for a meaningful comparison.

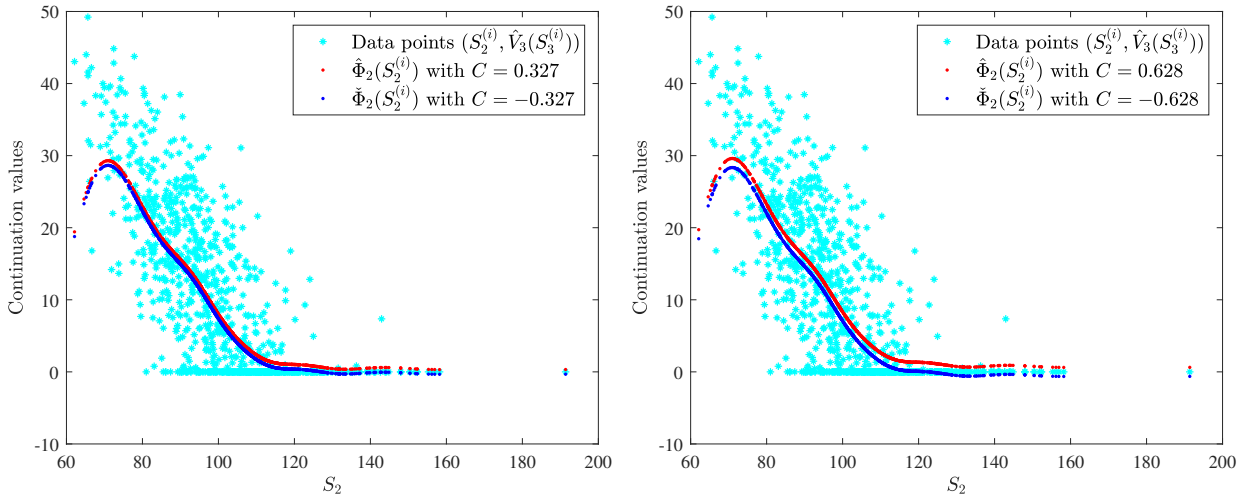
high- and low-biased continuation function approximations and the underlying data points at times 2 and 1 for an 80% (left) and a 95% confidence level (right), the estimated continuation functions and underlying data on the right are lying above those on the left. The larger upward (downward) shift of the high-biased (low-biased) continuation functions for the higher overall confidence level can be seen in Figure 2b and Figure 2c on close inspection, but it is clearly apparent visually in Figure 2a. Note that while the latter figure displays the high- and low-biased nonparametric regression functions for each overall confidence level, plotting both functions together with data at time 1 in one figure is not practical since the underlying data points for high- and low-biased function approximations are different at that point. This is because the underlying data themselves are (partly) based on over- and underestimated continuation values at time 2, given the estimates' recursive construction. Finally, note that while our kernel-based approach does not rely on structural properties – such as the Bermudan option's theoretical lower-bound value of zero[5] –, accounting for these can be expected to greatly improve the quality of our guaranteed bounds.

6. Conclusion

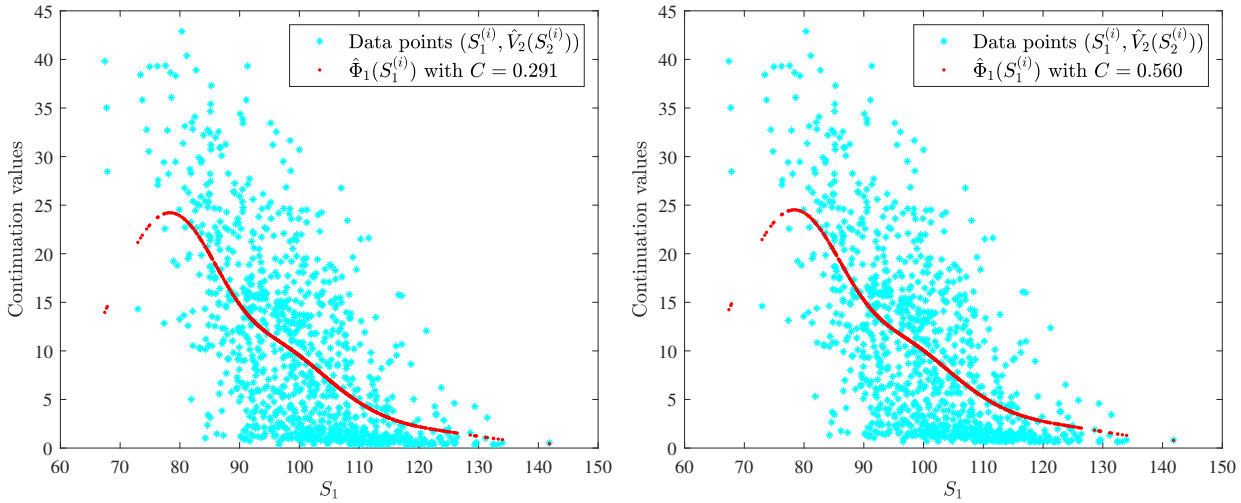
This paper presents for the first time an approach for obtaining guaranteed bounds for optimal stopping problems using non-asymptotic uniform confidence bands based on a KRR estimator. Nonparametric regression estimates require an assumption about the space of admissible functions: Without such an assumption, weak consistency cannot be guaranteed for the sample size tending to infinity. In the literature, one can find assumptions regarding a bounded second or higher derivative or regarding the boundedness of the Fourier transform. Such assumptions require the bound to be known. In contrast, the main assumption of the present paper is that the unknown regression function lies in the considered RKHS. This assumption is relatively weak, since these RKHS functions are dense in the space of continuous functions. Our proposed approach does not require the knowledge of a boundedness parameter for the considered function space, but we did require that the independent variable for the regression problem is univariate (in the considered optimal stopping problem, we accounted for a single stochastic process, or factor). An extension to multivariate regressors seems possible, but this is left for future work. If the observations are noise free, then a quite substantial simplification would take place: Both the error limit β as well as the parameter σ could be set to zero in our formulas. In this case, increasing the number of stages would not worsen the quality of the estimates and a much larger number of stages could easily be treated. However, a model with noise free observations may be unrealistic in certain applications.

We have illustrated the applicability of our proposed bounding procedure for optimal stopping problems by valuing a Bermudan-style put option. Optimal stopping problems feature a single binary decision variable, such as whether or not to exercise a put option. While such problems require the calculation of just one regression function per time stage, the presented methodology for

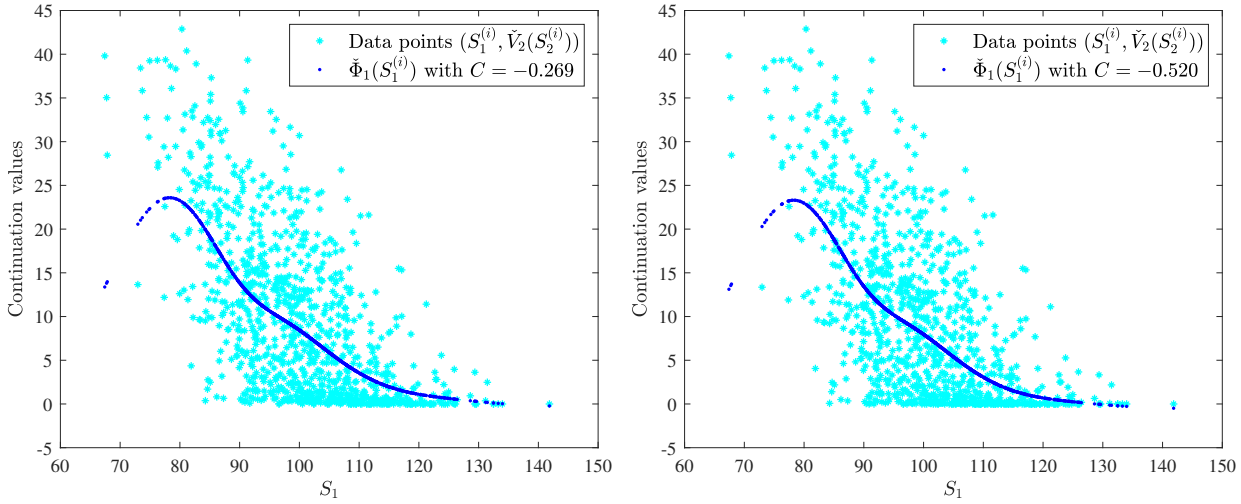
[5]Our low-biased continuation values $\check{\Phi}_2(S_2)$ and $\check{\Phi}_1(S_1)$ may become marginally negative for high stock prices S_2 and S_1, respectively.



(a) Realised and approximated (high- and low-biased) continuation values at time $t = 2$.



(b) High-biased realisations and approximation of continuation values at time $t = 1$.



(c) Low-biased realisations and approximation of continuation values at time $t = 1$.

Figure 2: Nonparametric regression functions and underlying data for overall confidence levels of 80% (left) and 95% (right) and a put option that is in the money at time 0 (with $S_0 = 100$).

over- and underestimation of a regression function is quite general. Indeed, it could be adapted for decision problems with a finite number of possible actions. This, however, would require regression estimates for all feasible actions in a given state of the system, which would increase the complexity. On a related note, the simple example that was used to illustrate our proposed methodology involved only three decision stages. Again, this is not a fundamental restriction of our data-driven approach. Every additional stage, however, would make the total estimation error increase, so obtaining tight bounds might then require larger sample sizes. Although all nonparametric regression estimates exhibit some boundary effects, such boundary effects only had a minor impact on the computed bounds in our example. Note that for the studied option pricing problem, the quality of the decision depends on the accuracy of the estimate in the region where the option is in-the-money. Lastly, improvements in performance may be achieved through hyperparameter tuning (e.g. kernel type, model-specific choices, and regularisation constant) and confidence bands whose width (size) varies locally according to the observed data, rather than being constant for all possible inputs. Future work will explore the construction of such variable-width confidence bands and the development of automated hyperparameter tuning strategies, which can be expected to lead to significantly improved performance bounds, especially for larger problem instances.

Acknowledgements

This work was supported by University College London under the Global Engagement funding scheme. The authors are grateful for the comments of participants at the 33rd European Conference on Operational Research (EURO) in Copenhagen and the European Conference on Stochastic Optimization and Computational Management Science (ECSO-CMS 2024) in Stockholm. Comments from the editor and three anonymous referees have greatly improved the paper. All remaining errors are the authors' own.

Appendix A. Proofs of Lemmas

Proof of Lemma 2.1

Let $f(\cdot) = \int_{\mathcal{X}} k(\cdot, x) v(x)\, Q(dx)$. Then

$$\begin{aligned}
\langle f, \mathcal{K} g \rangle_k &= \langle \int_{\mathcal{X}} k(\cdot, x) v(x)\, Q(dx), \int_{\mathcal{X}} k(\cdot, y) g(y)\, P(dy) \rangle_k \\
&= \iint_{\mathcal{X}\times\mathcal{X}} \langle k(\cdot, x), k(\cdot, y) \rangle_k v(x)\, Q(dx) g(y) P(dy) \\
&= \int_{\mathcal{X}} \left[\int_{\mathcal{X}} k(x, y) v(x)\, Q(dx) \right] g(y)\, P(dy) \\
&= \int_{\mathcal{X}} f(y) g(y)\, P(dy) = \langle f, g \rangle_{P,2}\,.
\end{aligned}$$

□

Proof of Lemma 2.2

By Lemma 2.1 we get that

$$\|\mathcal{K}w\|_k^2 = \langle \mathcal{K}w, \mathcal{K}w\rangle_k = \langle \mathcal{K}w, w\rangle_{P,2}\,.$$

By using $g = \mathcal{K}w$ we have to minimise (in w)

$$\|f - \mathcal{K}w\|_2^2 + \lambda\langle \mathcal{K}w, w\rangle_2 = \|f\|_2^2 - 2\langle f, \mathcal{K}w\rangle_2 + \langle \mathcal{K}w, \mathcal{K}w\rangle_2 + \lambda\langle \mathcal{K}w, w\rangle_2\,.$$

The first order optimality condition for this expression is

$$-2\mathcal{K}f + 2\mathcal{K}^2 w + 2\lambda\mathcal{K}w = 0\,.$$

Hence, $w = [\mathcal{K}^2 + \lambda\mathcal{K}]^{-1}\mathcal{K}f = [\mathcal{K} + \lambda I]^{-1}f$ and $S_\lambda f = [\mathcal{K} + \lambda I]^{-1}\mathcal{K}f\,.$ □

Proof of Lemma 2.3

Set $f_\lambda := S_\lambda f_0$ and let $f_0 = \mathcal{K}w_0$. Since $[\mathcal{K} + \lambda I]^{-1}\mathcal{K} = I - \lambda[\mathcal{K} + \lambda I]^{-1}$, using Lemma 2.2 we get

$$\begin{aligned} f_\lambda(x) &= (S_\lambda f_0)(x) \\ &= \langle k(x,\cdot), [\mathcal{K} + \lambda I]^{-1}\mathcal{K}^2 w_0\rangle_k \\ &= \langle k(x,\cdot), \mathcal{K}w_0 - \lambda[\mathcal{K} + \lambda I]^{-1}\mathcal{K}w_0\rangle_k \\ &= \langle k(x,\cdot), \mathcal{K}w_0\rangle_k - \lambda\langle k(x,\cdot), [\mathcal{K} + \lambda I]^{-1}\mathcal{K}w_0\rangle_k \\ &= f_0(x) - \lambda([\mathcal{K} + \lambda I]^{-1}\mathcal{K}w_0)(x)\,. \end{aligned}$$

From (5) we have that $\|[\mathcal{K} + \lambda I]^{-1}\mathcal{K}\|_k \le 1$. Therefore

$$\|f_0 - f_\lambda\|_\infty \le \|f_0 - f_\lambda\|_k \le \lambda\|[\mathcal{K} + \lambda I]^{-1}\mathcal{K}w_0\|_k \le \lambda\|w_0\|_k\,.$$

□

Proof of Lemma 2.4

The main assertion follows from the well known fact that

$$\left|\int h(y)P(dy) - \int h(y)Q(dy)\right| \le L(h)d_W(P,Q)\,.$$

To see that $c_L(f,k) \le L(f) + L(k)\|f\|_\infty$, first notice that for the Lipschitz constant L of a product of two functions h_1 and h_2 (with Lipschitz constants $L(h_1)$ and $L(h_2)$, respectively) it holds that $L(h_1 \cdot h_2) \le L(h_1) \cdot \|h_2\|_\infty + L(h_2) \cdot \|h_1\|_\infty$. Now, using the fact that $\|k\|_\infty = 1$ gives the desired result. □

Appendix B. Notation for probability distributions

Table B.2 lists and describes the different symbols used in the main text for the various probability distributions.

Symbol	Interpretation	Notation
$\bar{P}$	joint (bivariate) distribution of (X, Y), $(x, y) \in \mathcal{X} \times \mathbb{R}$	$(X, Y) \sim \bar{P}$
$\bar{P}_X$	conditional distribution of output Y given input X	$Y\|X \sim \bar{P}_X$
P	marginal (univariate) distribution of X w.r.t. $\bar{P}$	$X \sim P$
$\bar{P}^n$	joint distribution of $(X_i, Y_i)_{i=1}^n$, $(x_i, y_i)_{i=1}^n \subset \mathcal{X} \times \mathbb{R}$	$((X_1, Y_1), \ldots, (X_n, Y_n)) \sim \bar{P}^n$
$\bar{P}_X^n$	conditional distribution of sample given all X_i	$(Y_1, \ldots, Y_n\|X_1, \ldots, X_n) \sim \bar{P}_X^n$
P^n	marginal distribution of $(X_i)_{i=1}^n$ w.r.t. $\bar{P}^n$	$(X_1, \ldots, X_n) \sim P^n$
$\hat{P}_n$	empirical distribution associated with P^n	$\hat{P}_n[X = x] = \frac{1}{n} \sum_{i=1}^n \delta_{x_i}(x)$

Note: $x \in \mathcal{X}$ and $y \in \mathbb{R}$ are the realisations of random variables $X \colon \Omega \to \mathcal{X}$ and $Y \colon \Omega \to \mathbb{R}$, respectively, at sample point $\omega \in \Omega$, with sample space Ω, such that $x = X(\omega)$ and $y = Y(\omega)$. Equally, $x_i = X_i(\omega)$ and $y_i = Y_i(\omega)$ are the realisations of random variables X_i and Y_i, respectively, with $(X_i, Y_i) \sim \bar{P}$.

Table B.2: Notation for probability distributions.

Appendix C. Supplementary figures

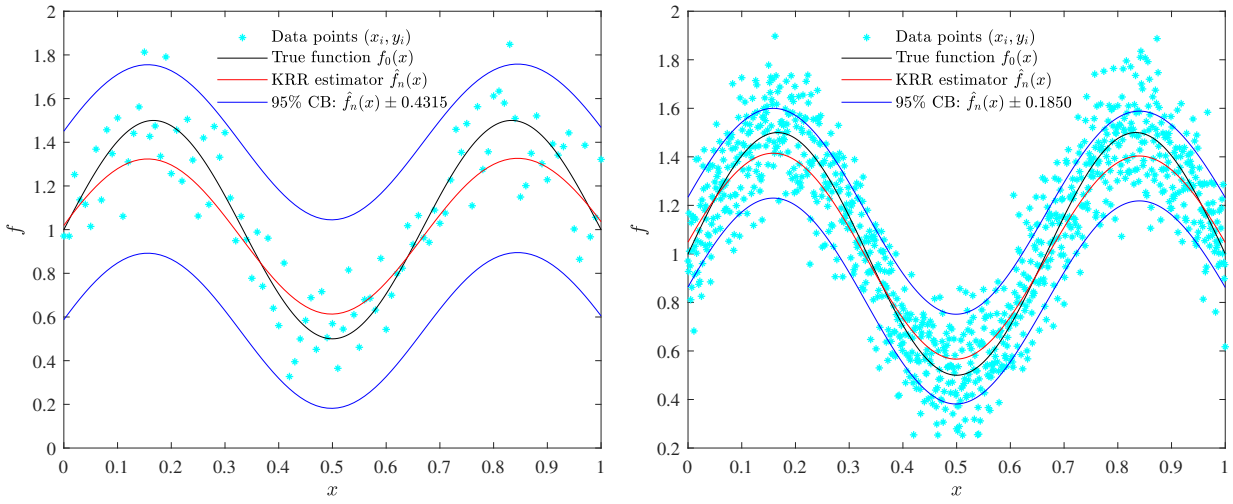
In this appendix, we illustrate the application of the key theoretical result of Section 3 using the numerical example of constructing uniform (or simultaneous) confidence bands in regression. Assume that $X \sim \mathcal{U}(0, 1)$ and consider a regression function of the form $f_0(x) = 1 + 0.5 \cdot \sin(3\pi x)$. The n observed pairs (X_i, Y_i) are i.i.d., where each $Y_i = f_0(X_i) + \epsilon_i$, with $\epsilon_i \overset{\text{iid}}{\sim} \mathcal{N}(0, 0.15^2)$. Given that $P = \mathcal{U}(0, 1)$, the optimal (deterministic) choice of regressors X_i, such that $d_W(P, \hat{P}_n)$ is minimised, is given by $x_i = \frac{i-1}{n-1}, i = 1, \ldots, n$, meaning $\hat{P}_n$ is the discrete uniform distribution supported on $\left\{0, \frac{1}{n-1}, \frac{2}{n-1}, \ldots, 1\right\}$. Thus, $d_W(P, \hat{P}_n) = \frac{1}{n-1}$. To obtain the KRR estimator $\hat{f}_n(x)$, we used a Gaussian kernel with $\alpha = 10$. Lastly, set $\beta = 0.05$. Figures C.3a and C.3b illustrate uniform 95% confidence bands $\hat{f}_n(x) \pm C$ for the true, but typically unknown, regression function $f_0(x)$ considering sample sizes of $n = 101$ and $n = 1001$, respectively.

References

Aronszajn, N., 1950. Theory of reproducing kernels. Transactions of the American Mathematical Society 68, 337–404.

Baardman, L., Cristian, R., Perakis, G., Singhvi, D., Skali Lami, O., Thayaparan, L., 2023. The role of optimization in some recent advances in data-driven decision-making. Mathematical Programming 200, 1–35.

Ban, G.Y., Rudin, C., 2019. The big data newsvendor: Practical insights from machine learning. Operations Research 67, 90–108.



(a) $n = 101$ (with $\lambda = 0.02$). (b) $n = 1001$ (with $\lambda = 0.01$).

Figure C.3: Uniform (or simultaneous) 95% confidence bands $\hat{f}_n(x) \pm C$, centred around the KRR estimator $\hat{f}_n(x)$, for the true regression function $f_0(x)$, using training data $(x_i, y_i), i = 1, \ldots, n,$ where $x_i \in \{0, 1/(n-1), 2/(n-1), \ldots, 1\}$.

Berlinet, A., Thomas-Agnan, C., 2011. Reproducing kernel Hilbert spaces in probability and statistics. Springer Science & Business Media.

Bertsimas, D., Koduri, N., 2022. Data-driven optimization: A reproducing kernel hilbert space approach. Operations Research 70, 454–471.

Bjerve, S., Doksum, K.A., Yandell, B.S., 1985. Uniform confidence bounds for regression based on a simple moving average. Scandinavian Journal of Statistics , 159–169.

Brown, D.B., Smith, J.E., 2022. Information Relaxations and Duality in Stochastic Dynamic Programs: A Review and Tutorial. Foundations and Trends in Optimization 5, 246–339.

Cai, T.T., Low, M., Ma, Z., 2014. Adaptive confidence bands for nonparametric regression functions. Journal of the American Statistical Association 109, 1054–1070.

Carriere, J.F., 1996. Valuation of the early-exercise price for options using simulations and nonparametric regression. Insurance: Mathematics and Economics 19, 19–30.

Ciocan, D.F., Mišić, V.V., 2022. Interpretable optimal stopping. Management Science 68, 1616–1638.

Csáji, B.C., Horváth, B., 2022. Nonparametric, Nonasymptotic Confidence Bands With Paley-Wiener Kernels for Band-Limited Functions. IEEE Control Systems Letters 6, 3355–3360.

David, I., Yechiali, U., 1985. A time-dependent stopping problem with application to live organ transplants. Operations Research 33, 491–504.

Desai, V.V., Farias, V.F., Moallemi, C.C., 2012. Pathwise optimization for optimal stopping problems. Management Science 58, 2292–2308.

Dommel, P., Pichler, A., 2023. Stochastic optimization with estimated objectives. Pure and Applied Functional Analysis 8, 1377–1399.

Dong, Y., 2024. Randomized optimal stopping problem in continuous time and reinforcement learning algorithm. SIAM Journal on Control and Optimization 62, 1590–1614.

Eubank, R.L., Speckman, P.L., 1993. Confidence bands in nonparametric regression. Journal of the American Statistical Association 88, 1287–1301.

Feng, Y., Gallego, G., 1995. Optimal starting times for end-of-season sales and optimal stopping times for promotional fares. Management science 41, 1371–1391.

Fournier, N., Guillin, A., 2015. On the rate of convergence in Wasserstein distance of the empirical measure. Probability theory and related fields 162, 707–738.

Gao, R., 2023. Finite-sample guarantees for wasserstein distributionally robust optimization: Breaking the curse of dimensionality. Operations Research 71, 2291–2306.

Gauffriau, A., Malgouyres, F., Ducoffe, M., 2021. Overestimation learning with guarantees. Preprint arXiv:2101.11717 .

Glasserman, P., 2004. Monte Carlo methods in financial engineering. volume 53. Springer.

Guigues, V., Juditsky, A., Nemirovski, A., 2017. Non-asymptotic confidence bounds for the optimal value of a stochastic program. Optimization Methods and Software 32, 1033–1058.

Hall, P., Horowitz, J., 2013. A simple bootstrap method for constructing nonparametric confidence bands for functions. The Annals of Statistics , 1892–1921.

Hull, J.C., 2021. Options, futures, and other derivatives. 11th ed., Pearson, Harlow, England.

Ibáñez, A., Velasco, C., 2020. Recursive lower and dual upper bounds for bermudan-style options. European Journal of Operational Research 280, 730–740.

Jiang, H., 2019. Non-asymptotic uniform rates of consistency for k-nn regression, in: Proceedings of the AAAI Conference on Artificial Intelligence, pp. 3999–4006.

Kanagawa, M., Hennig, P., Sejdinovic, D., Sriperumbudur, B.K., 2018. Gaussian processes and kernel methods: A review on connections and equivalences. `arXiv:1807.02582`.

Kang, B.K., 2005. Optimal stopping problem with double reservation value property. European Journal of Operational Research 165, 765–785.

Knafl, G., Sacks, J., Ylvisaker, D., 1985. Confidence bands for regression functions. Journal of the American Statistical Association 80, 683–691.

Lederer, A., Umlauft, J., Hirche, S., 2019. Uniform Error Bounds for Gaussian Process Regression with Application to Safe Control. Advances in Neural Information Processing Systems 32, 659–669.

Longstaff, F.A., Schwartz, E.S., 2001. Valuing American options by simulation: a simple least-squares approach. The Review of Financial Studies 14, 113–147.

Maier, S., Pflug, G.C., Polak, J.W., 2020a. Valuing portfolios of interdependent real options under exogenous and endogenous uncertainties. European Journal of Operational Research 285, 133–147.

Maier, S., Polak, J.W., Gann, D.M., 2020b. Valuing portfolios of interdependent real options using influence diagrams and simulation-and-regression: A multi-stage stochastic integer programming approach. Computers & Operations Research 115, 104505.

Micchelli, C.A., Xu, Y., Zhang, H., 2006. Universal kernels. Journal of Machine Learning Research 7.

Nadarajah, S., Margot, F., Secomandi, N., 2017. Comparison of least squares monte carlo methods with applications to energy real options. European Journal of Operational Research 256, 196–204.

Parikh, N., Boyd, S., 2014. Proximal algorithms. Foundations and Trends in Optimization 1, 127–239.

Pflug, G.Ch., Pichler, A., 2014. Multistage stochastic optimization. Springer.

Pichler, A., Liu, R.P., Shapiro, A., 2022. Risk-averse stochastic programming: Time consistency and optimal stopping. Operations Research 70, 2439–2455.

Powell, W.B., 2019. A unified framework for stochastic optimization. European Journal of Operational Research 275, 795–821.

Powell, W.B., 2022. Reinforcement Learning and Stochastic Optimization: A Unified Framework for Sequential Decisions. John Wiley & Sons, Inc.

Sadana, U., Chenreddy, A., Delage, E., Forel, A., Frejinger, E., Vidal, T., 2024. A survey of contextual optimization methods for decision-making under uncertainty. European Journal of Operational Research 320, 271–289.

Schmidt, F.G., Pibernik, R., 2024. Data-driven inventory control for large product portfolios: A practical application of prescriptive analytics. European Journal of Operational Research .

Schölkopf, B., Herbrich, R., Smola, A.J., 2001. A generalized representer theorem, in: Helmbold, D., Williamson, B. (Eds.), Computational Learning Theory, Springer Berlin Heidelberg. pp. 416–426.

Schölkopf, B., Smola, A.J., 2002. Learning with kernels: support vector machines, regularization, optimization, and beyond. The MIT Press.

Sriperumbudur, B.K., Fukumizu, K., Lanckriet, G.R., 2011. Universality, characteristic kernels and rkhs embedding of measures. Journal of Machine Learning Research 12.

Steinwart, I., Christmann, A., 2008. Support vector machines. Springer Science & Business Media.

Sturt, B., 2023. A nonparametric algorithm for optimal stopping based on robust optimization. Operations Research 71, 1530–1557.

Sun, J., Loader, C.R., 1994. Simultaneous confidence bands for linear regression and smoothing. The Annals of Statistics 22, 1328–1345.

Tsitsiklis, J.N., Van Roy, B., 2001. Regression Methods for Pricing Complex American-Style Options. IEEE Transactions on Neural Networks 12, 694–703.

Van Foreest, N.D., Kilic, O.A., 2023. An intuitive approach to inventory control with optimal stopping. European Journal of Operational Research 311, 921–924.

Wahba, G., 1990. Spline models for observational data. Society for Industrial and Applied Mathematics.

Wei, W., Zhu, D., 2022. Generic improvements to least squares Monte Carlo methods with applications to optimal stopping problems. European Journal of Operational Research 298, 1132–1144.