Global Optimization in Hilbert Space

Boris HOUSKA[1] and Benoît CHACHUAT[2]

[1]School of Information Science and Technology, ShanghaiTech University, 319 Yueyang Road, Shanghai, 200031, China.
[2]Centre for Process Systems Engineering (CPSE), Department of Chemical Engineering, Imperial College London, South Kensington Campus, London SW7 2AZ, United Kingdom.

Abstract

This paper proposes a complete-search algorithm for solving a class of non-convex, possibly infinite-dimensional, optimization problems to global optimality. We assume that the optimization variables are in a bounded subset of a Hilbert space, and we determine worst-case run-time bounds for the algorithm under certain regularity conditions of the cost functional and the constraint set. Because these run-time bounds are independent of the number of optimization variables and, in particular, are valid for optimization problems with infinitely many optimization variables, we prove that the algorithm converges to an ϵ-suboptimal global solution within finite run-time for any given termination tolerance $\epsilon > 0$. Finally, we illustrate these results for a problem of calculus of variations.

Keywords

Infinite-dimensional optimization · Complete search · Branch-and-lift · Convergence analysis · Complexity analysis

1 Introduction

Infinite-dimensional optimization problems arise in many research fields, including optimal control [9, 10, 27, 57], optimization with partial differential equations (PDE) embedded [25], and shape/topology optimization [7]. In practice, these problems are often solved approximately by applying discretization techniques; the original infinite-dimensional problem is replaced by a finite-dimensional approximation that can then be tackled using standard optimization techniques. However, the resulting discretized optimization problems may comprise a large number of optimization variables, which grows unbounded as the accuracy of the approximation is refined. Unfortunately, worst-case run-time bounds for complete-search algorithms in nonlinear programming (NLP) scale rather poorly with the number of optimization variables. For instance, the worst-case run-time of spatial branch-and-bound [19, 48] scales exponentially with the number of optimization variables. In contrast, algorithms for solving convex optimization problems in polynomial run-time are known [13, 44], e.g. in linear programming (LP) or convex quadratic programming (QP). While these efficient algorithms enable the solution of very large-scale convex optimization problems, such as structured or sparse problems, in general their worst-case run-time bounds also grow unbounded as the number of decision variables tends to infinity.

Existing theory and algorithms that directly analyze and exploit the infinite-dimensional nature of an optimization problem are mainly found in the field of convex optimization. For the most part, these algorithms rely on duality in convex optimization in order to construct upper and lower bounds on the optimal solution value, although establishing strong duality in infinite-dimensional problems can prove difficult. In this context, infinite-dimensional linear programming problems have been analyzed thoroughly [3]. A variety of algorithms are also available for dealing with convex infinite-dimensional optimization problems, some of which have been analyzed in generic Banach spaces [16], as well as certain tailored algorithms for continuous linear programming [4, 15, 35].

In the field of non-convex optimization, problems with an infinite number of variables are typically studied in a local neighborhood of a stationary point. For instance, local optimality in continuous-time optimal control problems can be analyzed by using Pontryagin's maximum principle [50], and a number of local optimal control algorithms are based on this analysis [8, 14, 54, 57]. More generally, approaches in the classical field of variational analysis [41] rely on local analysis concepts, from which information about global extrema may not be derived in general. In fact, non-convex infinite-dimensional optimization remains an open field of research and, to the best knowledge of the authors, there currently are no generic, complete search algorithms available for solving such problems to global optimality.

This paper asks the question whether a global optimization algorithm can be constructed, whose worst-case run-time complexity is independent of the number of optimization variables thereof, such that this algorithm would remain tractable for infinite-dimensional optimization problems. Clearly, devising such an algorithm may only be possible for a certain class of optimization problems. Interestingly, the fact that the "complexity" or "hardness" of an optimization problem does not necessarily depend on the number of optimization variables has been observed – and it is in fact exploited – in state-of-the-art global optimization solvers for NLP/MINLP, although these observations are still to be analyzed in full detail. For instance, instead of applying a branch-and-bound algorithm in the original space of optimization variables, global NLP/MINLP solvers such as BARON [52, 55] or ANTIGONE [38] proceed by lifting the problem to a higher-dimensional space via the introduction of auxiliary variables from the DAG decomposition of the objective and constraint functions. In a different context, the solution of a lifted problem in a higher-dimensional space has become popular in numerical optimal control, where the so-called multiple-shooting methods often outperform their single-shooting counterparts despite the fact that the former calls for the solution a larger-scale (discretized) NLP problem [9, 10]. This idea that certain optimization problems become easier to solve than equivalent problems in fewer variables is also central to the work on lifted Newton methods [2]. To the best of our knowledge, such behaviors cannot be explained currently with results from the field of complexity analysis, which typically give monotonically increasing worst-case run-time bounds as the number of optimization variables increases. In this respect, these run-time bounds therefore predict the opposite behavior to what can sometimes be observed in practice.

1.1 Problem Formulation

The focus of this paper is on complete search algorithms for solving non-convex optimization problems of the form:

$$\inf_{x \in C} \; F(x)\,, \tag{1}$$

where $F : H \to \mathbb{R}$ and $C \subseteq H$ denote the cost functional and the constraint set, respectively; and the domain H of this problem is a (possibly infinite-dimensional) Hilbert space with respect to the inner product $\langle \cdot, \cdot \rangle : H \times H \to \mathbb{R}$. The theoretical considerations in the paper do not assume a separable Hilbert space, although our various illustrating examples are based on separable spaces.

Definition 1. *A feasible point $x^* \in C$ is said to be an ϵ-suboptimal global solution – or ϵ-global optimum – of (1), with $\epsilon > 0$, if*

$$\forall x \in C, \quad F(x^*) < F(x) + \epsilon\,.$$

We make the following assumptions regarding the geometry of C throughout the paper.

Assumption 1. *The constraint set C is convex, has a nonempty relative interior, and is bounded with respect to the induced norm on H; that is, there exists a constant $\gamma < \infty$ such that*

$$\forall x \in C, \quad \|x\|_H := \sqrt{\langle x, x \rangle} \leq \gamma\,.$$

The main objective of the paper is to develop an algorithm that can locate an ϵ-suboptimal global optimum of Problem (1), in finite run-time for any given accuracy $\epsilon > 0$, provided that F satisfies certain regularity conditions alongside Assumption 1.

Remark 1. Certain infinite-dimensional optimization problems are formulated in a Banach space $(B, \|\cdot\|)$ rather than a Hilbert space, for instance in the field of optimal control of partial differential equations in order to analyze the existence of extrema [25]; that is, the optimization problem (1) becomes

$$\inf_{x \in \hat{C}} \; \hat{F}(x) \tag{2}$$

with $\hat{F} : B \to \mathbb{R}$ and $\hat{C}$ a convex bounded subset of B. Nonetheless, provided that:

1. the Hilbert space $H \subseteq B$ is convex and dense in $(B, \|\cdot\|)$;
2. the function $\hat{F}$ is upper semi-continuous in $\hat{C}$; and
3. the constraint set $\hat{C}$ has a nonempty relative interior;

we may consider Problem (1) with $C := \hat{C} \cap H$ instead of (2), for any ϵ-suboptimal global solution of the former is also an ϵ-suboptimal global solution of (2), and both problems have such ϵ-suboptimal points. Because Conditions 1-3 are often satisfied in practical applications, it is for the purpose of this paper not restrictive to assume that the domain of the optimization variables is indeed a Hilbert space.

1.2 Outline and Contributions

The paper starts by discussing several regularity conditions for sets and functionals defined in a Hilbert space in Sect. 2, based on which complete-search algorithms can be constructed whose run-time is independent of the number of optimization variables. Such an algorithm is presented in Sect. 3 and analyzed in Sect. 4, which constitute the main contributions and novelty. A numerical case study is presented in Sect. 5 in order to illustrate the main results, before concluding the paper in Sect. 6.

Although certain of these algorithmic ideas are inspired by a recent paper of the authors on global optimal control [28], the present paper develops a much more general framework for optimization in Hilbert space. Besides, Sect. 4 derives novel worst-case complexity estimates for the proposed algorithm. We argue that these ideas could help lay the foundations towards new ways of analyzing the complexity of certain optimization problems based on their structural properties rather than their number of optimization variables. Although the run-time estimates for the proposed algorithm remain conservative, they indicate that the complexity of numerical optimization does not necessarily depend on whether the problem is small-scale, large-scale, or even infinite-dimensional.

2 Some Regularity Conditions for Sets and Functionals in Hilbert Space

This section builds upon basic concepts in infinite-dimensional Hilbert spaces in order to arrive at certain regularity conditions for sets and functionals defined in these spaces. Such a focus on Hilbert spaces is motivated by their property that an orthogonal basis $\Phi_0, \Phi_1, \ldots \in H$ can be constructed, such that

$$\forall i, j \in \mathbb{N}, \qquad \frac{1}{\sigma_i} \langle \Phi_i, \Phi_j \rangle \;=\; \delta_{i,j} \;:=\; \begin{cases} 0 & \text{if } i \neq j, \\ 1 & \text{otherwise,} \end{cases}$$

for some scalars $\sigma_0, \sigma_1, \ldots \in \mathbb{R}^{++}$. Having such a basis, we can define the associated projection functions $P_M : H \to H$ for each $M \in \mathbb{N}$ as

$$\forall x \in H, \quad P_M(x) \;:=\; \sum_{k=0}^{M} \frac{\langle x, \Phi_k \rangle}{\sigma_k} \, \Phi_k \; .$$

A natural question arising at this point is what can be said about the distance between an element $x \in H$ and its projection $P_M(x)$ for a given $M \in \mathbb{N}$.

Definition 2. *We call $D(M, x) \;:=\; \| \, x - P_M(x) \, \|_H$ the distance between an element $x \in H$ and its projection $P_M(x)$. Moreover, given the constraint set $C \in H$, we define the bound*

$$\overline{D}_C(M) \;:=\; \sup_{x \in C} \, D(M, x) \; .$$

Lemma 1. *Under Assumption 1, the function $\overline{D}_C(M)$ is uniformly bounded from above by γ in $\mathbb{N}$.*

Proof. For each $M \in \mathbb{N}$, we have

$$\left[\overline{D}_C(M)\right]^2 \;=\; \sup_{x \in C} \| x - P_M(x) \|_H^2 \;=\; \sup_{x \in C} \left(\langle x, x \rangle - \sum_{k=0}^{M} \frac{\langle x, \Phi_k \rangle^2}{\sigma_k} \right) \;\leq\; \sup_{x \in C} \langle x, x \rangle \; .$$

The result follows by noting that $\sup_{x \in C} \langle x, x \rangle \leq \gamma^2$ by Assumption 1. □

Although it is uniformly bounded, the function $\overline{D}_C(M)$ does not converge to zero as $M \to \infty$ in infinite-dimensional Hilbert spaces in general. This behavior is illustrated in the following example.

Example 1. Consider the case that all the basis functions $\Phi_0, \Phi_1, \ldots$ are in the constraint set C, and define the sequence $\{x_k\}_{k \in \mathbb{N}}$ with $x_k := \Phi_{k+1}$. We have

$$\forall k \in \mathbb{N}, \quad 1 \;=\; \| x_k \|_H \;=\; \| x_k - P_k(x_k) \|_H \;=\; D(k, x_k) \;\leq\; \overline{D}_C(k) \, ,$$

and therefore

$$\limsup_{k \to \infty} \; \overline{D}_C(k) \;\geq\; 1 \; .$$

◇

Such a lack of convergence is unfortunate since, without additional regularity assumptions, the existence of minimizers to Problem (1) may not be asserted. Moreover, for a sequence $(x_k)_{k \in \mathbb{N}}$ of feasible points of Problem (1) converging to an infimum, we may have

$$\limsup_{M \to \infty} \, \limsup_{k \to \infty} \; D(M, x_k) \;\neq\; \limsup_{k \to \infty} \, \limsup_{M \to \infty} \; D(M, x_k) \; .$$

In other words, any attempt to approximate the infimum by constructing a sequence of finite parameterizations of the optimization variable x could in principle be unsuccessful. Therefore, a principal aim of the following sections is to develop an optimization algorithm, whose convergence to an ϵ-global optimum of Problem (1) can be certified, yet without assuming anything about the existence, or even the regularity, of the minimizers of Problem (1). Instead, we shall impose a suitable regularity conditions on the objective function F in (1).

In preparation for this analysis, we start by formalizing a notion of regularity for the elements of H.

Definition 3. *An element $a \in H$ is said to be* regular for the constraint set C *if*

$$\lim_{M \to \infty} \mathcal{R}_C(M, a) \;=\; 0 \quad \textit{with} \quad \mathcal{R}_C(M, a) := \overline{D}_C(M) D(M, a) \; . \tag{3}$$

Moreover, we call the function $\mathcal{R}_C(\cdot, a) : \mathbb{N} \to \mathbb{R}^+$ the convergence rate at a on C.

Theorem 1. *For any $a \in H$, we have*

$$\forall M \in \mathbb{N}, \quad \sup_{x \in C} \; | \langle a, x - P_M(x) \rangle | \;\leq\; \mathcal{R}_C(M, a) \; . \tag{4}$$

Moveover, in the case that the element a is regular for C, we have

$$\lim_{M\to\infty} \sup_{x\in C} |\langle a, x - P_M(x)\rangle| \;=\; 0\;.$$

Proof. For any given $M \in \mathbb{N}$, consider the optimization problem

$$\overline{V}_M \;:=\; \sup_{x\in C} \langle a, x - P_M(x)\rangle \;=\; \sup_{x\in C} \langle a, w\rangle\;,$$

where we have introduced the variable $w := x - P_M(x)$ such that

$$\forall x \in C, \quad \|w\|_H \;\leq\; \overline{D}_C(M)\;.$$

Since the functions $\Phi_0, \ldots, \Phi_M$ are orthogonal, we have $\langle \Phi_k, w\rangle \;=\; 0$ for all $k \in \{0, \ldots, M\}$, and it follows that

$$\overline{V}_M \;\leq\; \sup_{w\in H} \langle a, w\rangle \quad \text{s.t.} \quad \langle \Phi_k, w\rangle \;=\; 0\;, \quad \|w\|_H \;\leq\; \overline{D}_C(M)\,.$$

Next, we use duality to obtain

$$\overline{V}_M \;\leq\; \inf_{\lambda\in\mathbb{R}^{M+1}} \sup_{w\in H} \left\langle a - \sum_{k=0}^{M} \lambda_k \Phi_k\,,\, w \right\rangle \quad \text{s.t.} \quad \|w\|_H \;\leq\; \overline{D}_C(M)\;,$$

where $\lambda \in \mathbb{R}^{M+1}$ denotes the multipliers associated with the constraint $\langle \Phi_k, w\rangle = 0$ for $k \in \{0, \ldots, M\}$. Applying the Cauchy-Schwarz inequality then gives

$$\forall \lambda \in \mathbb{R}^{M+1}, \quad \overline{V}_M \;\leq\; \left\| a - \sum_{k=0}^{M} \lambda_k \Phi_k \right\|_H \overline{D}_C(M)\,,$$

and with the particular choice $\lambda_k^* := \frac{\langle a, \Phi_k\rangle}{\sigma_k}$ for each $k \in \{0, \ldots, M\}$, we have

$$\overline{V}_M \;\leq\; \|a - P_M(a)\|_H \;\overline{D}_C(M) \;=\; \mathcal{R}_C(M, a)\;.$$

The optimal value of the minimization problem

$$\underline{V}_M \;:=\; \inf_{x\in H} \langle a, x - P_M(x)\rangle \quad \text{s.t.} \quad c \;\in\; C\;.$$

can be estimated analogously, giving $\underline{V}_M \geq -\mathcal{R}_C(M, a)$, and the result follows. □

The following example establishes the regularity of piecewise smooth functions with a finite number of singularities in the Hilbert space of square-integrable functions with the orthogonal Legendre polynomials as basis functions.

Example 2. We consider the Hilbert space $H \;=\; L_2[0, 1]$ of standard square-integrable functions on the interval $[0, 1]$ equipped with the standard inner product, $\langle f, g\rangle := \int_0^1 f(s)g(s)\mathrm{d}s$, and choose the orthogonal Legendre polynomials on the interval $[0, 1]$ with weighting factors $\sigma_k = \frac{1}{2k+1}$ as the basis functions $(\Phi_k)_{k\in\mathbb{N}}$. Our focus is on piecewise smooth functions $a : [0,1] \to \mathbb{R}$ with a given finite number of singularities, for which we want to establish regularity in the sense of Definition 3 for a bounded constraint set $C \subset L_2[0,1]$.

There exist many results on approximating functions using polynomials, including convergence rate estimates [17]. One such result in [51] shows that any piecewise smooth function $f : [0,1] \to \mathbb{R}$ can be approximated with a polynomial $p_f^M : [0,1] \to \mathbb{R}$ of degree M such that

$$\forall y \in [0,1], \quad \left\| f(y) - p_f^M(y) \right\| \leq K_1 \exp\left(-K_2 M^\alpha d(y)^\beta \right) , \tag{5}$$

for any given $\alpha, \beta > 0$ with either $\alpha < 1$ and $\beta \geq \alpha$, or $\alpha = 1$ and $\beta > 1$; some constants $K_1, K_2 > 0$; and where $d(y)$ denotes the distance to the nearest singularity. In particular, the following convergence rate estimate can be derived using this result in the present example, for any piecewise smooth functions $a : [0,1] \to \mathbb{R}$ with a finite number of singularities:

$$\begin{aligned}\mathcal{R}_C(M,a) \;&=\; \|a - P_M(a)\|_2 \, \overline{D}_C(M) \;=\; \inf_\lambda \left\| a - \sum_{k=0}^M \lambda_k \Phi_k \right\|_2 \overline{D}_C(M) \\ &\overset{\text{(Lemma 1)}}{\leq} \inf_\lambda \left\| a - \sum_{k=0}^M \lambda_k \Phi_k \right\|_2 \gamma \;\leq\; \frac{K}{\sqrt{M}}\end{aligned}$$

for some constant $K < \infty$. In order to establish the very last part of the above inequality, it is enough to consider a function a with a single singularity, e.g., at the mid-point $y = \frac{1}{2}$ and using $\alpha = \beta = \frac{1}{2}$:*

$$\begin{aligned}\inf_\lambda \left\| a - \sum_{k=0}^M \lambda_k \Phi_k \right\|_2 \;&\leq\; \sqrt{\int_0^1 K_1^2 \exp\left(-2K_2 \sqrt{M \left| y - \frac{1}{2} \right|} \right) \mathrm{d}y} \\ &=\; \sqrt{\frac{K_1^2}{\left[K_2 \sqrt{M} \right]^2} + \mathbf{O}\left(\frac{1}{\sqrt{M}} \exp\left(-K_2 \sqrt{M} \right) \right)} \;=\; \mathbf{O}\left(\frac{1}{\sqrt{M}} \right) .\end{aligned} \tag{6}$$

Convergence rate estimates for k-times differentiable and piecewise smooth functions can be obtained in a similar way, using for instance the results in [17, 51]. $\diamond$

A useful generalization of Definition 3 and corresponding corollary of Theorem 1 are given below.

Definition 4. *The set $\mathcal{A} \subseteq H$ is said to be* regular for C *if*

$$\lim_{M\to\infty} \overline{\mathcal{R}}_C(M,\mathcal{A}) \;=\; 0 \quad \textit{with} \quad \overline{\mathcal{R}}_C(M,\mathcal{A}) \;:=\; \sup_{a\in\mathcal{A}} \mathcal{R}_C(M,a) \, .$$

Moreover, we call the function $\overline{\mathcal{R}}_C(\cdot,\mathcal{A}) : \mathbb{N} \to \mathbb{R}^+$ the worst-case convergence rate for the set $\mathcal{A}$ on C.

*We have used the integration formula $\int e^{\sqrt{ax}} \, \mathrm{d}x = \frac{2e^{\sqrt{ax}}(\sqrt{ax} - 1)}{a} + C$ for the integral term in (6).

Corollary 1. *For any regular set $\mathcal{A} \subseteq H$, we have*

$$\lim_{M\to\infty} \sup_{\substack{a\in\mathcal{A},\\ x\in C}} |\langle a, x - P_M(x)\rangle| = 0 .$$

Remark 2. While any subset of the Euclidean space $\mathbb{R}^n$ is trivially regular for all bounded subsets $C \subset \mathbb{R}^n$, only certain subsets/subspaces of infinite-dimensional Hilbert spaces may be regular. In the space of square-integrable functions $H := L_2[a, b]$ for instance, the subspace $\mathcal{LC}^p$ of p-times differentiable functions on $[a, b]$, with uniformly Lipschitz continuous p-th derivatives, is regular for any bounded constraint set $C \subset L_2[a, b]$. It can be established – e.g., from the analysis in [30] using the standard trigonometric Fourier basis, or from the results in [58] using the Legendre polynomials – that

$$\overline{\mathcal{R}}_C(M, \mathcal{LC}^p) \leq \mathbf{O}\left(\log(M)M^{-p-1}\right) \leq \mathbf{O}\left(M^{-p}\right) .$$

This leads to a rather typical situation, whereby the stronger the regularity assumptions on the function class, the faster convergence of the associated worst-case convergence rate $\mathcal{R}(\cdot, \mathcal{LC}^p)$—an increase in the convergence rate order $\frac{\log(M)}{M^{p+1}}$ with p in this instance. In the limit of $\mathcal{C}^\infty$ (smooth) functions, it can even be established – e.g., using standard results from Fourier analysis [21, 31] – that the convergence rate becomes exponential,

$$\overline{\mathcal{R}}_C(M, \mathcal{C}^\infty) \leq \mathbf{O}\left(\exp(-\beta M)\right) \quad \text{with } \beta > 0 .$$

Example 2 (Continued)**.** Consider the following set of unit-step functions

$$\mathcal{A} := \{ x_t \,|\, t \in [0, 1]\} \quad \text{with} \quad \forall \tau \in [0, 1],\; x_t(\tau) := \begin{cases} 1 & \text{if } \tau \leq t, \\ 0 & \text{otherwise,} \end{cases}$$

for which we want to establish regularity in the sense of Definition 4. Using earlier results from Example 2, it is known that the function $x_{0.5}$ can be approximated with a sequence of polynomials $p^M_{0.5} : [0, 1] \to \mathbb{R}$ of degree M such that

$$\left\| x_{0.5} - p^M_{0.5} \right\|_2 \leq \mathbf{O}\left(\frac{1}{\sqrt{M}}\right) .$$

Then, for every $t \in [0, 1]$, we can construct the family of polynomials

$$\forall \tau \in [0, 1], \quad p^M_t(\tau) := p^M_{0.5}\left(\frac{1 - t + \tau}{2}\right) .$$

Since the latter satisfy the same property as $x_{0.5}$ that

$$\left\| x_t - p^M_t \right\|_2 \leq \frac{K}{\sqrt{M}} ,$$

where the constant $K > 0$ is independent of t or M, we have $\overline{\mathcal{R}}_C(M, \mathcal{A}) \leq \mathbf{O}\left(\frac{1}{\sqrt{M}}\right)$.

This example can be generalized to other classes of functions. For instance, given any smooth function $f \in L_2[0,1]$, the family of functions

$$\mathcal{A}(f) \ := \ \{ a \in H \,|\, \exists t \in [0,1] : \ a(\tau) = f(\tau) \text{ if } \tau \leq t; a(\tau) = 0 \text{ otherwise}\}$$

is regular in H, and also satisfies $\overline{\mathcal{R}}_C(M,\mathcal{A}) \leq \mathbf{O}\left(\frac{1}{\sqrt{M}}\right)$. This result can be established by writing the functions in $\mathcal{A}(f)$ as the product between the piecewise smooth function f and the function x_t, and then approximating the factors separately. $\diamond$

In the final part of this section, we introduce and illustrate a regularity condition for the cost functional in Problem (1).

Definition 5. *Let C be a bounded subset of H, and let $\mathcal{A}$ be a regular subset of H. The functional $F : H \to \mathbb{R}$ is said to be $\mathcal{A}$-*strongly Lipschitz continuous *on C if there exists a constant $L < \infty$ such that*

$$\forall e \in E_C, \quad \sup_{x \in C} |F(x+e) - F(x)| \ \leq \ L \sup_{a \in \mathcal{A}} |\langle a, e \rangle| \ , \tag{7}$$

where the projection error set $E_C \subseteq H$ is given by

$$E_C \ := \ \{ P_M(x) - x \,|\, x \in C, \ M \in \mathbb{N} \} \ .$$

Remark 3. In the special case that F is a linear functional, given by

$$F(x) \ := \ F_0 + \langle \hat{a}, x \rangle$$

for some regular element $\hat{a} \in H$ and any other element $F_0 \in H$, the condition (7) is trivially satisfied with $L = 1$ and $\mathcal{A} = \{\hat{a}\}$. In this interpretation, the regularity condition (7) essentially provides a means of keeping the nonlinear part of F under control.

Remark 4. We may assume $L = 1$ in Definition 5 without loss of generality, for the set $\mathcal{A}$ can always be rescaled. It is nonetheless natural to retain L in condition (7), since this constant can be interpreted as a generalized Lipschitz constant, under a proper scaling of $\mathcal{A}$. More precisely, the condition (7) is satisfied with any global Lipschitz constant L of F on C by choosing the regular set $\mathcal{A}$ such that

$$\forall e \in E_C, \quad \|e\|_H \ \leq \ \sup_{a \in \mathcal{A}} |\langle a, e \rangle| \ . \tag{8}$$

Unfortunately, the construction of regular sets that would satisfy (8) is generally impractical for infinite-dimensional Hilbert spaces. This is in contrast with the condition (7), which holds for many practical problems, as illustrated in the new few examples.

Example 3. Consider the finite-dimensional Euclidean space $H = \mathbb{R}^n$ and a bounded subset $C \subseteq \mathbb{R}^n$.

By the mean-value theorem, any continuously-differentiable function $F : \mathbb{R}^n \to \mathbb{R}$ satisfies

$$\forall e \in E_C, \quad \sup_{x \in C} |F(x+e) - F(x)| \;=\; \sup_{x\in C} \left| \int_0^1 \left\langle \frac{\partial F}{\partial x}(x+\eta e), e \right\rangle \mathrm{d}\eta \right| \;\leq\; \sup_{g \in \mathcal{G}} |\langle g, e\rangle| \;,$$

where $\mathcal{G} \subseteq \mathbb{R}^n$ denotes the set of gradient values of F on C. Thus, any continuously differentiable function is $\mathcal{G}$-strongly Lipschitz continuous on C.

A generalization of this result to infinite-dimensional Hilbert space is also possible for certain classes of functionals. For instance, any Fréchet-differentiable function $F : H \to \mathbb{R}$ satisfying

$$\forall (x,e) \in C \times E_C, \quad F(x+e) - F(x) \;=\; \int_0^1 \langle DF(x+\eta e), e\rangle \,\mathrm{d}\eta \,,$$

where the set of Fréchet derivatives $\mathcal{G} := \{DF(x) \mid x \in C\} \subseteq H$ is regular, is $\mathcal{G}$-strongly Lipschitz continuous on any bounded subset C of H. $\diamond$

The following two examples investigate strong Lipschitz continuity of certain classes of functionals in the practical space of square-integrable functions with the orthogonal Legendre polynomials as basis functions. The first one (Example 4) illustrates the case of a functional that is not strongly Lipschitz continuous; the second one (Example 5) identifies a broad class of strongly Lipschitz continuous functionals defined via the solution of an embedded ODE system. The objective here is to help the reader develop an intuition that strongly Lipschitz continuous functions occur naturally in many, although not all, problems of practical relevance.

Example 4. We consider the Hilbert space $H = L_2[0,1]$ of square-integrable functions on the interval $[0,1]$ with the standard inner product, and select the basis functions $(\Phi_k)_{k\in\mathbb{N}}$ as the orthogonal Legendre polynomials on the interval $[0,1]$ with weighting factors $\sigma_k = \frac{1}{2k+1}$. We investigate whether the functional F given below is strongly Lipschitz continuous on the set $C := \{x \in L_2[0,1] \mid \forall s \in [0,1],\ |x(s)| \leq 1\}$,

$$\forall x \in L_2[0,1], \quad F(x) \;:=\; \|x\|_2^2 \;=\; \int_0^1 x(s)^2 \,\mathrm{d}s \,.$$

Consider the following family of sets

$$\forall M \in \mathbb{N}, \quad E_M \;:=\; \{P_M(x) - x \mid x \in C\} \;\subseteq\; L_2[0,1] \,,$$

such that $E_C = \cup_{M\in\mathbb{N}} E_M$. If the condition (7) were to hold for some regular set $\mathcal{A}$, we would have by Theorem 1,

$$\lim_{M\to\infty} \sup_{\substack{e\in E_M,\\ x\in C}} |F(x+e) - F(x)| \;\leq\; L \lim_{M\to\infty} \sup_{\substack{e\in E_M,\\ a\in\mathcal{A}}} |\langle a, e\rangle| \;\leq\; L \lim_{M\to\infty} \overline{\mathcal{R}}_C(M,\mathcal{A}) \,,$$

and it would follow from Corollary 1 that

$$\lim_{M\to\infty} \sup_{\substack{e\in E_M,\\ x\in C}} |F(x+e)-F(x)| \;=\; 0\,.$$

However, this leads to a contradiction since we also have

$$\lim_{M\to\infty} \sup_{\substack{e\in E_M,\\ x\in C}} |F(x+e)-F(x)| \;\geq\; \lim_{M\to\infty} \sup_{e\in E_M} \|e\|_2^2 \;=\; \lim_{M\to\infty} \sup_{x\in C} \|x-P_M(x)\|_2^2$$

and $\sup_{x\in C}\|x-P_M(x)\|_2^2 = 1$ for all $M\in\mathbb{N}$. Therefore, the regularity condition (7) may not be satisfied for any choice of the regular set $\mathcal{A}$, and F is not strongly Lipschitz continuous on C. ⋄

Remark 5. The result that the functional F in Example 4 is not strongly Lipschitz continuous on C is not in contradiction with Example 3. Although F is Fréchet differentiable in $L_2[0,1]$, the set of Fréchet derivative functions of F fails to be regular on the bounded set C, as it is too big. Nonetheless, strong Lipschitzness holds for F on the restricted set of uniformly bounded and Lipschitz continuous functions in $L_2[0,1]$ with uniformly bounded Lipschitz constants.

Example 5. We again consider the Hilbert space $H = L_2[0,1]$ of square-integrable functions on the interval $[0,1]$ equipped with the standard inner product, and select the basis functions $(\Phi_k)_{k\in\mathbb{N}}$ as the orthogonal Legendre polynomials on the interval $[0,1]$ with weighting factors $\sigma_k = \frac{1}{2k+1}$. Our focus is on the ordinary differential equation (ODE)

$$\forall t\in[0,1],\quad \frac{\partial x}{\partial t}(t,u) = f(x(t,u)) + Bu(t) \quad \text{with} \quad x(0,u)=0\,, \tag{9}$$

where $B\in\mathbb{R}^{n\times n}$ is a constant matrix; and $f:\mathbb{R}^n\to\mathbb{R}^n$, a continuously differentiable and globally Lipschitz continuous function, so that the solution trajectory $x(\cdot,u):[0,1]\to\mathbb{R}^n$ is well-defined for all $u\in L_2[0,1]$. For simplicity, we consider the functional F given by

$$F(u) \;:=\; c^\mathsf{T} x(1,u)\,,$$

for some real vector $c\in\mathbb{R}^n$. Moreover, the constraint set $C\subseteq H$ may be any uniformly bounded function subset here, such as simple uniform bounds of the form

$$C \;:=\; \{u\in L_2[0,1] \mid \forall \tau\in[0,1],\ |u(\tau)|\leq 1\}\,.$$

The following developments aim to establish that F is strongly Lipschitz continuous on C.

By Taylor's theorem, the difference function $\delta(t,u,e) := x(t,u+e)-x(t,u)$ satisfies the differential equation

$$\forall t\in[0,1],\quad \frac{\partial\delta}{\partial t}(t,u,e) \;=\; \Gamma(t,u,e)\delta(t,u,e)+Be(t)$$

with $\delta(0,u,e)=0$ and $\Gamma(t,u,e) := \int_0^1 \frac{\partial f}{\partial x}(x(t,u)+\eta(x(t,u+e)-x(t,u)))\,\mathrm{d}\eta$. Since f is globally

Lipschitz continuous, for any given smooth matrix-valued function $A : [0, 1] \to \mathbb{R}^{n \times n}$, we have

$$\forall (t, u, e) \in [0, 1] \times C \times E_C, \quad \|\Gamma(t, u, e) - A(t)\| \leq \ell_1 \,,$$

for some constant $\ell_1 < \infty$. For a particular choice of A, we can decompose $\delta(t, u, e)$ as the sum $\delta_{\mathrm{l}}(t, e) + \delta_{\mathrm{n}}(t, u, e, \delta_{\mathrm{l}})$ corresponding to the solution of the ODE system

$$\forall t \in [0, 1], \quad \delta_{\mathrm{l}}(t, e) = A(t)\delta_{\mathrm{l}}(t, e) + Be(t) \tag{10}$$

$$\delta_{\mathrm{n}}(t, u, e, \delta_{\mathrm{l}}) = \Gamma(t, u, e)\delta_{\mathrm{n}}(t, u, e, \delta_{\mathrm{l}}) + [\Gamma(t, u, e) - A(t)]\delta_{\mathrm{l}}(t, e) \tag{11}$$

with $\delta_{\mathrm{l}}(0, e) = \delta_{\mathrm{n}}(0, u, e, \delta_{\mathrm{l}}) = 0$. For this decomposition, the left-hand side of (7) satisfies

$$\forall e \in E_C, \quad \sup_{u \in C} |F(u + e) - F(u)| \leq \left| c^{\mathsf{T}} \delta_{\mathrm{l}}(1, e) \right| + \sup_{u \in C} \left| c^{\mathsf{T}} \delta_{\mathrm{n}}(1, u, e) \right| \,.$$

Regarding the linear term δ_{l}, we have

$$\forall s \in [0, 1], \quad c^{\mathsf{T}} \delta_{\mathrm{l}}(s, e) = \langle g_s, e \rangle \tag{12}$$

with

$$\forall t \in [0, 1], \quad g_s(t) := \begin{cases} \int_0^t c^{\mathsf{T}} G(t, \tau) B \, \mathrm{d}\tau & \text{if } t \leq s, \\ 0 & \text{otherwise,} \end{cases}$$

where $G(t, \tau)$ denotes the fundamental solution of the linear ODE (10) such that

$$\forall \tau, t \in [0, 1], \quad \frac{\partial}{\partial t} G(t, \tau) \;=\; A(t) G(t, \tau) \quad \text{with} \quad G(\tau, \tau) = I \,.$$

Since A is smooth, it follows from Example 2 that the set $\mathcal{G} := \{ g_s \mid s \in [0, 1] \}$ is regular in $L_2[0, 1]$ and satisfies

$$\overline{\mathcal{R}}_C(M, \mathcal{G}) \;\leq\; \mathbf{O}\left(\frac{1}{\sqrt{M}} \right) \,.$$

Regarding the nonlinear term δ_{n}, since the function Γ is uniformly bounded, we can apply Gronwall's lemma to the ODE (11) to obtain

$$\begin{aligned} \forall (t, u, e) \in [0, 1] \times C \times E_C, \quad c^{\mathsf{T}} \delta_{\mathrm{n}}(t, u, e, \delta_{\mathrm{l}}) &\leq \ell \exp(\ell) \sup_{s \in [0,1]} |c^{\mathsf{T}} \delta_{\mathrm{l}}(s, e)| \\ &\leq \ell \exp(\ell) \sup_{g \in \mathcal{G}} |\langle g, e \rangle| \,, \end{aligned} \tag{13}$$

for some constant $\ell < \infty$. Finally, combining (12) and (13) shows that F satisfies the condition (7) with $L := 1 + \ell \exp(\ell)$, thus F is $\mathcal{G}$-strongly Lipschitz continuous on C. ⋄

Remark 6. The functional F in the previous example is defined implicitly via the solution of an ODE. The result that such functionals are strongly Lipschitz continuous is particularly significant insofar as the proposed optimization framework will indeed encompass a broad class of optimal control problems. In fact, it turns out that strong Lipschitzness still holds in replacing the constant matrix B in (9) with any matrix-valued continuously differentiable and globally Lipschitz continuous function of $x(t,u)$, thus encompassing quite a general class of nonlinear affine-control systems. In the case of general nonlinear ODEs, however, strong Lipschitzness may be lost unless the constraint set C is further restricted to uniformly bounded and Lipschitz continuous functions in $L_2[0,1]$ with uniformly bounded Lipschitz constants; compare Example 4 and Remark 5.

3 Global Optimization in Hilbert Space using Complete Search

The application of complete-search strategies to infinite-dimensional optimization problems such as (1) calls for an extension of the (spatial) branch-and-bound principle [26] to general Hilbert space. The approach presented in this section differs from branch-and-bound in that the dimension M of the search space is adjusted, as necessary, during the iterations of the algorithm, by using a so-called *lifting* operation – hence the name *branch-and-lift* algorithm. The basic idea is to bracket the optimal solution value of Problem (1) and progressively refine these bounds via this lifting mechanism, combined with traditional branching and fathoming.

Based on the developments in Sect. 2, the following subsections describe methods for exhaustive partitioning in infinite-dimensional Hilbert space (Sect. 3.1) and for computing rigorous upper and lower bounds on given subsets of the variable domain (Sect. 3.2), before presenting the proposed branch-and-lift algorithm (Sect. 3.3).

3.1 Partitioning in Infinite-Dimensional Hilbert Space

Similar to branch-and-bound search, the proposed branch-and-lift algorithm maintains a partition $\mathbb{A} := \{A_1,\ldots,A_k\}$ of finite-dimensional sets $A_1,\ldots,A_k$. This partition is updated through the repeated application of certain operations, including branching and lifting, in order to close the gap between an upper and a lower bound on the global solution value of the optimization problem (1). The following definition is useful in order to formalize these operations:

Definition 6. *With each pair $(M,A)\in\mathbb{N}\times\mathcal{P}(\mathbb{R}^{M+1})$, we associate a subregion $\mathcal{X}_M(A)$ of H given by*

$$\mathcal{X}_M(A) \;:=\; \left\{ x\in C \;\middle|\; \left(\frac{\langle x,\Phi_0\rangle}{\sigma_0},\ldots,\frac{\langle x,\Phi_M\rangle}{\sigma_M}\right)^{\mathsf{T}} \in\; A \right\} .$$

Moreover, we say that the set A is infeasible *if $\mathcal{X}_M(A)=\varnothing$.*

Notice that each subregion $\mathcal{X}_M(A)$ is a convex set if the sets C and A are themselves convex. For practical reasons, we restrict ourselves to compact subsets $A\in\mathcal{S}_M\subseteq\mathcal{P}(\mathbb{R}^{M+1})$ here, where the class of sets $\mathcal{S}_M$ is easily stored and manipulated by a computer. For example, $\mathcal{S}_M$ could be a set of simple interval boxes, polytopes, ellipsoids, etc.

The ability to detect infeasibility of a set $A\in\mathcal{S}_M$ is pivotal for complete search. Under the assumption that the constraint set C is convex (Assumption 1), a certificate of infeasibility can be obtained by considering the convex optimization problem

$$d_C(A) \ := \ \min_{x,y \in H} \|x-y\|_H \quad \text{s.t.} \quad \left(\frac{\langle y, \Phi_0 \rangle}{\sigma_0}, \ldots, \frac{\langle y, \Phi_M \rangle}{\sigma_M} \right)^\intercal \in A \,, \quad x \in C \,. \tag{14}$$

It readily follows from the Cauchy-Schwarz inequality that

$$-\|x-y\|_H \le \langle x, \Phi_k \rangle - \langle y, \Phi_k \rangle \le \|x-y\|_H \ ,$$

for any (normalized) basis function Φ_k, and so $\|x-y\|_H = 0$ implies $\langle x, \Phi_k \rangle = \langle y, \Phi_k \rangle$. Consequently, a set A is infeasible if and only if $d_C(A) > 0$. Because Slater's constraint qualification holds for Problem (14) under Assumption 1, one approach to checking infeasibility to within high numerical accuracy relies on duality for computing lower bounds on the optimal solution value $d_C(A)$—similar in essence to the infinite-dimensional convex optimization techniques in [4, 16]. For the purpose of this paper, our focus is on a general class of non-convex objective functionals F, whereas the constraint set C is assumed to be convex and have a simple geometry in order to avoid numerical issues in solving feasibility problems of the form (14). We shall therefore assume, from this point onwards, that infeasibility can be verified with high numerical accuracy for any set $A \in \mathcal{S}_M$.

A *branching* operation subdivides any set $A \in \mathcal{S}_M$ in the partition $\mathbb{A}$ into two compact subsets $A_\mathrm{l}, A_\mathrm{r} \in \mathcal{S}_M$ such that $A_\mathrm{l} \cup A_\mathrm{r} \supseteq A$, thereby updating the partition as

$$\mathbb{A} \leftarrow \mathbb{A} \setminus \{A\} \cup \{A_\mathrm{l}, A_\mathrm{r}\} \,.$$

On the other hand, a *lifting* operation essentially lifts any set $A \in \mathcal{S}_M$ into a higher-dimensional space under the function $\Gamma_M : \mathcal{S}_M \to \mathcal{S}_{M+1}$, defined such that

$$\forall A \in \mathcal{S}_M, \quad \mathcal{X}_M(A) \ \subseteq \ \mathcal{X}_{M+1}(\Gamma_M(A)) \,.$$

The question as to defining the higher-order coefficient $\langle x, \Phi_{M+1} \rangle$ in such a lifting is related to the so called *moment problem* that asks the question under which conditions on a sequence $(a_k)_{k \in \{1,\ldots,N\}}$, named moment sequence, can we find an associated element $x \in H$ with $a_k = \frac{\langle x, \Phi_k \rangle}{\sigma_k}$ for each $k \in \{1, \ldots, N\}$. Classical examples of such moment problems are Stieltjes', Hamburger's, and Legendre's moment problems [1]. Here, we adopt the modern standpoint on moment problems using convex optimization [33, 46], by considering the following optimization subproblems:

$$\underline{a}_{M+1}(A) \ \le \ \min_{x \in \mathcal{X}_M(A)} \frac{\langle x, \Phi_{M+1} \rangle}{\sigma_{M+1}} \quad \text{and} \quad \overline{a}_{M+1}(A) \ \ge \ \max_{x \in \mathcal{X}_M(A)} \frac{\langle x, \Phi_{M+1} \rangle}{\sigma_{M+1}} \,. \tag{15}$$

Although both optimization problems in (15) are convex if A and C are convex, they remain infinite-dimensional, and thus intractable in general. Obtaining lower and upper bounds $\underline{a}_{M+1}(A), \overline{a}_{M+1}(A)$ is nonetheless straightforward under Assumption 1. In case no better approach is available, one can always use

$$\underline{a}_{M+1}(A) := -\frac{\gamma}{\sigma_{M+1}} \quad \text{and} \quad \overline{a}_{M+1}(A) := \frac{\gamma}{\sigma_{M+1}} \ ,$$

which follows readily from the Cauchy-Schwarz inequality and the property that $\|\Phi_{M+1}\|_H = 1$. As already mentioned in the introduction of the paper, a variety of algorithms are now available for tackling convex infinite dimensional problems both efficiently and reliably [4, 16], which could provide tighter bounds in practical applications.

A number of remarks are in order:

Remark 7. The idea to introduce a lifting operation to enable partition in infinite-dimensional function space was originally introduced by the authors in a recently publication [28], focusing on global optimization of optimal control problems. One principal contribution in the present paper is a generalization of these ideas to global optimization in any Hilbert space, by identifying a set of sufficient regularity conditions on the cost functional and constraint set for the resulting branch-and-lift algorithms to converge to an ϵ-global solution in finite run-time.

Remark 8. Many recent optimization techniques for global optimization are based on the theory of positive polynomials and their associated linear matrix inequality (LMI) approximations [33, 49], which are also originally inspired by moment problems. Although these LMI techniques may be applied in the practical implementation of the aforementioned lifting operation, they are not directly related to the branch-and-lift algorithm that is developed in the following sections. An important motivation for moving away from the generic LMI framework is that the available implementations scale quite poorly with the number of optimization variables, due to the combinatorial increase of the number of monomials in the associated multivariate polynomial. Therefore, a direct approximation of the cost function F with multivariate polynomials would conflict with our primary objective to develop a global optimization algorithm whose worst-case run-time does not depend on the number of optimization variables.

3.2 Strategies for Upper and Lower Bounding of Functionals

Besides partitioning, the efficient construction of tight upper and lower bounds on the global solution value of (1) for given subregions of H is key in a practical implementation of branch-and-lift. Here, we shall call $L_M, U_M : \mathcal{S}_M \to \mathbb{R}$ lower- and upper-bounding functions of the functional F such that

$$\forall A \in \mathcal{S}_M, \quad L_M(A) \;\leq\; \inf_{x \in \mathcal{X}_M(A)} \; F(x) \;\leq\; U_M(A)\,. \tag{16}$$

A simple approach to constructing these lower and upper bounds relies on the following decomposition:

1. In a first step, we compute bounds $L^0_M(A)$ and $U^0_M(A)$ on the finite-dimensional approximation of F as

$$\forall A \in \mathcal{S}_M, \quad L^0_M(A) \;\leq\; \inf_{a \in A} \; F\left(\sum_{i=0}^{M} a_i \Phi_i \right) \;\leq\; U^0_M(A)\,. \tag{17}$$

   Clearly, it depends on the particular expression of F how to determine such bounds in practice. In the case that F is factorable, various arithmetics can be used to propagate bounds through a DAG of the function, such as interval arithmetic [40], McCormick relaxations [11, 37] or Taylor model arithmetic [12, 47]. Moreover, if the expression of F is embedding a dynamic system described by differential equations, validated bounds can be obtained by using a variety of set-propagation techniques as described, e.g., in [29, 34, 42, 53, 56]; see also [24, 32].

2. The second step involves computing a bound $\Delta_M(A)$ on the approximation errors such that

$$\forall A \in \mathcal{S}_M, \quad \left| \inf_{x \in \mathcal{X}_M(A)} F(x) - \inf_{a \in A} F\left(\sum_{i=0}^{M} a_i \Phi_i\right) \right| \leq \Delta_M(A)\,. \tag{18}$$

In the case that F is $\mathcal{G}$-strongly Lipschitz continuous on C, we can always take $\Delta_M(A) := L\,\overline{\mathcal{R}}_C(M,\mathcal{G})$, where the constant L satisfies the condition (7). Naturally, better bounds may be derived by exploiting a particular structure or expression of F.

By construction, the lower-bounding function $L_M(A) := L_M^0(A) - \Delta_M(A)$ and the upper-bounding function $U_M(A) := U_M^0(A) + \Delta_M(A)$ trivially satisfy (16). Moreover, when the set $A \in \mathcal{S}_M$ is infeasible—see related discussion in Sect. 3.1—we may set $\Delta_M(A) = L_M(A) = U_M(A) = \infty$.

At this point, we consider the following assumption in anticipation of the convergence analysis in Sect. 4:

Assumption 2. *The functional F in Problem (1) is $\mathcal{G}$-strongly Lipschitz continuous on C, where $\mathcal{G} \in H$ is a regular set for C and with corresponding constant $L < \infty$ in (7). Moreover, F is Lipschitz continuous with respect to the norm $\|\cdot\|_H : H \to \mathbb{R}^+$ with corresponding constant $K < \infty$, and the basis functions Φ_k are uniformly bounded with respect to $\|\cdot\|_H$.*

Remark 9. Under Assumption 2, any pair $(M, A) \in \mathbb{N} \times \mathcal{S}_M$ is such that

$$\forall a, a' \in A, \quad \left| F\left(\sum_{k=0}^{M} a_k \Phi_k\right) - F\left(\sum_{k=0}^{M} a'_k \Phi_k\right) \right| \leq K \sum_{k=0}^{M} |a_k - a'_k|\, \|\Phi_k\| \leq K'\, d_1(A)$$

with $K' := K \sup_{k \in \mathbb{N}} \|\Phi_k\|_H$ and $d_1(A) := \sum_{i=0}^{M} \sup_{a,a' \in A} |a_i - a'_i|$. It follows that

$$\forall (M, A) \in \mathbb{N} \times \mathcal{S}_M, \quad U_M(A) - L_M(A) \leq K'\, d_1(A) + 2\, L\, \overline{\mathcal{R}}_C(M, \mathcal{G})\,,$$

and therefore the gap $U_M(A) - L_M(A)$ can be made arbitrarily small under Assumption 2 by choosing a sufficiently large order M and a sufficiently small diameter for the set A. This observation will be exploited systematically for the convergence analysis in Sect. 4.

Remark 10. An alternative upper bound $U_M(A)$ in (16) may be computed more directly by solving the following nonconvex optimization problem to local optimality,

$$\min_{a \in A} F\left(\sum_{k=0}^{M} a_k \Phi_k\right) \quad \text{s.t.} \quad \sum_{k=0}^{M} a_k \Phi_k \in C\,. \tag{19}$$

Without further constraint qualifications or other assumptions on C, however, it is not hard to contrive examples where $\sum_{k=0}^{M} a_k \Phi_k \notin C$ for all $a \in H$ and all $M \in \mathbb{N}$. This upper-bounding approach can nonetheless be combined with another bounding approach based on set arithmetics in order to prevent convergence issues; e.g., use the solution value of (19) as long as it provides a bound that is smaller than $U_M^0(A) + \Delta_M(A)$.

3.3 Branch-and-Lift Algorithm

The foregoing considerations on partitioning and bounding in Hilbert space can be combined in Algorithm 1 for solving infinite-dimensional optimization problems to ϵ-global optimality.

Algorithm 1: Branch-and-lift algorithm for global optimization in Hilbert space

---

Input: Termination tolerance $\epsilon > 0$; Lifting parameter $\varrho > 0$

Initialization:

1. Set $M = 0$ and $\mathbb{A} = \{A_0\}$ with $A_0 \in \mathcal{S}_0$, $A_0 \supseteq \{\langle x, \Phi_0 \rangle \mid x \in C\}$

Repeat:

2. Select a set $A \in \mathbb{A}$
3. Compute upper and lower bounds, $L_M(A) \leq \inf_{x \in \mathcal{X}_M(A)} F(x) \leq U_M(A)$
4. Apply a fathoming operation
5. If the condition $\min_{A \in \mathbb{A}} U_M(A) - \min_{A \in \mathbb{A}} L_M(A) \leq \epsilon$ is satisfied, stop
6. If the condition $U_M(A) - L_M(A) \leq 2(1+\varrho)\Delta_M(A)$ holds for all $A \in \mathbb{A}$, apply a lifting operation and set $M \leftarrow M + 1$
7. Apply a branching operation, and return to step 2

Output: An ϵ-suboptimal solution of Problem (1)

A number of remarks are in order:

- Regarding initialization, the branch-and-lift iterations starts with $M = 0$. A possible way of initializing the partition $\mathbb{A} = \{A_0\}$ is by noting that
$$\{\langle x, \Phi_0 \rangle \mid x \in C\} \subseteq \left[-\frac{\gamma}{\sigma_0}, \frac{\gamma}{\sigma_0}\right]$$
under Assumption 1.

- Besides the branching and lifting operations introduced earlier in Sect. 3.1, fathoming in Step 4 of Algorithm 1 refers to the process of discarding a given set $A \in \mathbb{A}$ from the partition if
$$L_M(A) = \infty \quad \text{or} \quad \exists\, A' \in \mathbb{A}: \; L_M(A) > U_M(A')\,.$$

- The main idea behind the lifting condition defined in Step 6 of Algorithm 1, namely
$$\forall A \in \mathbb{A}, \quad U_M(A) - L_M(A) \leq 2(1+\varrho)\Delta_M(A)\,, \tag{20}$$

is that a subset A should be lifted to a higher-dimensional space whenever the approximation error $\Delta_M(A)$ due to the finite parameterization becomes of the same order of magnitude as the current optimality gap $U_M(A) - L_M(A)$. The aim here is to apply as few lifts as possible, since it is preferable to branch in a lower dimensional space. The convergence of the branch-and-lift algorithm under this lifting condition is examined in Sect. 4 below. Notice also that a lifting operation is applied globally – that is, to all parameter subsets in the partition $\mathbb{A}$ – in Algorithm 1, so all the subsets in $\mathbb{A}$ share the same parameterization order at any iteration. In a variant of Algorithm 1, one could also imagine a family of subsets that would have different parameterization orders by applying the lifting condition locally instead.

- Finally, it will be established in the following section that, upon termination and under certain assumptions, Algorithm 1 returns an ϵ-suboptimal solution of Problem (1). In particular, Assumption 1 rules out the possibility of an infeasible solution.

4 Convergence Analysis of Branch-and-Lift

This section investigates the convergence properties of the branch-and-lift algorithm (Algorithm 1) developed previously. It is convenient to introduce the following notation in order to conduct the analysis:

Definition 7. *Let $\mathcal{G} \in H$ be a regular set for C, and define the inverse function $\overline{\mathcal{R}}_C^{-1}(\cdot, \mathcal{G}) : \mathbb{R}^{++} \to \mathbb{N}$ by*

$$\forall \epsilon > 0, \quad \overline{\mathcal{R}}_C^{-1}(\epsilon, \mathcal{G}) \;:=\; \min_{M \in \mathbb{N}} \; M \;\; \text{s.t.} \;\; \overline{\mathcal{R}}_C(M, \mathcal{G}) \leq \epsilon \, .$$

A direct consequence of the lifting condition (20) in the branch-and-lift algorithm is the following:

Lemma 2. *Let F be $\mathcal{G}$-strongly Lipschitz continuous, where $\mathcal{G} \in H$ is a regular set for C and with corresponding constant $L < \infty$ in (7). Suppose that finite bounds $L_M^0(A)$, $U_M^0(A)$ and $\Delta_M(A)$ satisfying (17)-(18) can be computed for any feasible pair $(M, A) \in \mathbb{N} \times \mathcal{S}_M$. Then, the number of lifting operations in a run of Algorithm 1 as applied to Problem (1) is at most*

$$\overline{M} := \overline{\mathcal{R}}_C^{-1}\left(\frac{\epsilon}{2(\varrho+1)L} \, , \, \mathcal{G} \right) \, ,$$

regardless of whether or not the algorithm terminates finitely.

Proof. Assume that $M = \overline{M}$ in Algorithm 1, and that the termination condition is not yet satisfied; that is,

$$U_{\overline{M}}(A) - L_{\overline{M}}(A) > \epsilon$$

for a certain feasible set $A \in \mathbb{A}$. If the lifting condition (20) were to hold for A, then it would follow from (17)-(18) that

$$\epsilon - 2\Delta_{\overline{M}}(A) \;<\; U_{\overline{M}}^0(A) - L_{\overline{M}}^0(A) \;\leq\; 2\varrho\Delta_{\overline{M}}(A) \, .$$

Moreover, F being $\mathcal{G}$-strongly Lipschitz continuous on C by assumption, we would have

$$\overline{\mathcal{R}}_C(\overline{M}, \mathcal{G}) \ > \ \frac{\epsilon}{2(\varrho+1)L} \ .$$

This is a contradiction, since $\overline{\mathcal{R}}_C(\overline{M}, \mathcal{G}) \leq \frac{\epsilon}{2(\varrho+1)L}$ by Definition 7. □

Besides having a finite number of lifting operations, the convergence of Algorithm 1 can be established if the elements of a partition can be made arbitrarily small after applying a finite number of subdivisions.

Definition 8. *A partitioning scheme is said to be* exhaustive *if, given any dimension $M \in \mathbb{N}$, any tolerance $\eta > 0$, and any bounded initial partition $\mathbb{A} = \{A_0\}$, we have*

$$\mathrm{diam}\,(\mathbb{A}) \ := \ \max_{A \in \mathbb{A}} \ \mathrm{diam}\,(A) \ < \ \eta \, ,$$

after finitely many subdivisions, where $\mathrm{diam}\,(A) := \sup_{a,a' \in A} \ \|a - a'\|$. Moreover, we denoted by $\Sigma(\eta, M)$ an upper bound on the corresponding number of subdivisions in an exhaustive scheme.

The following theorem provides the main convergence result for the proposed branch-and-lift algorithm.

Theorem 2. *Let Assumptions 1 and 2 hold, and suppose that finite bounds $L_M^0(A)$, $U_M^0(A)$ and $\Delta_M(A)$ satisfying (17)-(18) can be computed for any feasible pair $(M, A) \in \mathbb{N} \times \mathcal{S}_M$. If the partitioning scheme is exhaustive, then Algorithm 1 terminates after at most $\overline{\Sigma}$ iterations, with*

$$\overline{\Sigma} \ \leq \ \max_{0 \leq M \leq \overline{M}} \ \Sigma\left(\frac{\epsilon\varrho}{K'(\varrho+1)}, M \right) \ . \tag{21}$$

Proof. By Lemma 2, the maximal number M of lifting operations during a run of Algorithm 1 is finite, such that $M \ \leq \ \overline{M}$. Therefore, the lifting condition (20) may not be satisfied for any feasible subset $A \in \mathbb{A}$, and we have

$$\Delta_M(A) \ \leq \ \frac{U_M^0(A) - L_M^0(A)}{2\,\varrho} \, .$$

Since $L_M(A) = L_M^0(A) - \Delta_M(A)$ and $U_M(A) = U_M^0(A) + \Delta_M(A)$, it follows that the termination condition $U_M(A) - L_M(A) \leq \epsilon$ is satisfied if

$$U_M^0(A) - L_M^0(A) \ \leq \ \frac{\varrho\epsilon}{\varrho+1} \ .$$

By Assumption 2, we have

$$U_M^0(A) - L_M(A) \ \leq \ K' \, \mathrm{diam}\,(\mathbb{A}) \ ,$$

with K' as defined in Remark 9, and the termination condition is thus satisfied if

$$\mathrm{diam}\,(\mathbb{A}) \ \leq \ \frac{\epsilon\varrho}{K'(\varrho+1)} \, .$$

This latter condition is met after at most $\Sigma\left(\frac{\epsilon\varrho}{K'(\varrho+1)}, M\right)$ iterations under the assumption that the partitioning scheme is exhaustive. □

Remark 11. In the case that the sets $A \in \mathbb{A}$ are simple interval boxes and the lifting process is implemented per (15), we have

$$\forall k \in \{0,\ldots,M\}, \quad [\,\underline{a}_k(A), \overline{a}_k(A)\,] \;\subseteq\; \left[-\frac{\gamma}{\sigma_k}, \frac{\gamma}{\sigma_k}\right] .$$

Therefore, one can always subdivide these boxes in such a way that the condition $\mathrm{diam}\,(\mathbb{A}) \leq \eta$ is satisfied after at most $\Sigma(\eta, M)$ subdivisions, with

$$\Sigma(\eta, M) \;:=\; \prod_{k=0}^{M} \left\lceil \frac{\gamma}{\eta\,\sigma_k} \right\rceil \;\in\; \mathbb{N} ,$$

for any given dimension M. In particular, $\Sigma(\eta, M)$ is monotonically increasing in M, and (21) simplifies to

$$\overline{\Sigma} \;\leq\; \Sigma\left(\frac{\epsilon\varrho}{K'(\varrho+1)}, \overline{M}\right) .$$

It should be clear, at this point, that the worst-case estimate $\overline{\Sigma}$ given in Theorem 2 may be extremely conservative, and the performance of Algorithm 1 could be much better in practice. Nonetheless, a key property of this estimate $\overline{\Sigma}$ is that it is independent of the actual nature or the number of optimization variables in Problem (1), be it a finite-dimensional or even an infinite-dimensional optimization problem. As already pointed in the introduction of the paper, this result is quite remarkable since available runtime estimates for standard convex and non-convex optimization algorithms do not enjoy this property. On the other hand, $\overline{\Sigma}$ is dependent on:

- the bound γ on the constraint set C;
- the Lipschitz constants K and L of the cost functional F;
- the uniform bound $\sup_k \|\Phi_k\|_H$ and the scaling factors σ_k of the chosen orthogonal functions Φ_k; and
- the lifting parameter ϱ and the termination tolerance ϵ in Algorithm 1.

All these dependencies are illustrated in the following example.

Example 6. Consider the space of square-integrable functions $H := L_2[-\pi, \pi]$, for which it has been established in Remark 2 that the subspace $\mathcal{LC}^p$ of p-times differentiable functions on $[-\pi, \pi]$ is regular, with convergence rate $\overline{\mathcal{R}}_C(M, \mathcal{LC}^p) \leq \alpha M^{-p}$ for some constant $\alpha < \infty$. On choosing the standard trigonometric Fourier basis, such that $\sigma_k = \pi$ are constant scaling factors and $K' := K \sup_k \|\Phi_k\|_2 = K$, and doing the partitioning using simple interval boxes as in Remark 11, a worst-case iteration count can be obtained as

$$\overline{\Sigma} \;=\; \left(\left\lceil \frac{\gamma K(\varrho+1)}{\pi\,\varrho\,\epsilon}\right\rceil\right)^{\left\lceil (2\alpha(\varrho+1)L/\epsilon)^{\frac{1}{p}}\right\rceil} \;\leq\; \exp\left(\mathbf{O}\left((1/\epsilon)^{\frac{1}{p}}\,\log(1/\epsilon)\right)\right)\,.$$

Now, if the global minimizer of Problem (1) is known to be a smooth function – instead of a more general p-times differentiable functions – the convergence rate can be expected to be of the form $\mathcal{R}(M,\mathcal{C}^\infty) = \alpha\exp(-\beta M)$, and Theorem 2 then predicts a worst-case iteration count as

$$\overline{\Sigma} \;\leq\; \exp\left(\mathbf{O}\left((\log(1/\epsilon))^2\right)\right)\,,$$

which is much more favorable. $\diamond$

5 Numerical Case Study

We consider the Hilbert space $H := L^2[0,T]$ of square-integrable functions on the interval $[0,T]$, here with $T = 10$. Our focus is on the following nonconvex, infinite-dimensional optimization problem

$$\begin{aligned} \inf_{x\in L_2[0,T]} F(x) &:= \int_0^T \left[\left(\int_0^T f_1(t-t')x(t')\,\mathrm{d}t'\right)^2 - \left(\int_0^T f_2(t-t')x(t')\,\mathrm{d}t'\right)^2\right]\mathrm{d}t \\ \text{s.t.}\;\; x\in C &:= \{\, x\in H \mid \forall t\in[0,T],\; |x(t)|\leq 1\}\,, \end{aligned} \tag{22}$$

with the functions f_1 and f_2 given by

$$\forall t\in\mathbb{R},\quad f_1(t) = \frac{t}{2}\left(\sin\left(\frac{\pi t}{2T}\right)+1\right) \quad \text{and} \quad f_2 = \frac{\partial f_1}{\partial t}\,.$$

Notice the symmetry in the optimization problem (22), as $F(x) = F(-x)$ and $x \in C$ iff $-x \in C$. Thus, if x^* is a global solution point of (22), then $-x^*$ is also a global solution point.

Although it might be possible to apply techniques from the field of variational analysis to determine the set of optimal solutions, our main objective here is to apply Algorithm 1 without exploiting any particular knowledge about the solution set. For this, we use the Legendre polynomials as basis functions in $L^2[0,T]$,

$$\forall i\in\mathbb{N}\quad \Phi_i(t) = (-1)^i\sum_{j=0}^{i}\begin{pmatrix} i \\ j\end{pmatrix}\begin{pmatrix} i+j \\ j\end{pmatrix}\left(-\frac{t}{T}\right)^j,$$

which are orthogonal by construction.

We start by showing that the functional F is $\mathcal{G}$-strongly Lipschitz continuous, with

$$\mathcal{G} \;:=\; \left\{\, f_1^t \;\middle|\; t\in[0,T]\,\right\}\cup\left\{\, f_2^t \;\middle|\; t\in[0,T]\,\right\} \;\subseteq\; H\,,$$

where we use the shorthand notation $f_1^t(\tau) := f_1(t-\tau)$ and $f_2^t(\tau) := f_2(t-\tau)$. For all $x \in L_2[0,T]$

and all $e \in E_C$, we have

$$
\begin{aligned}
|F(x+e)-F(x)| &= \left| \int_0^T \langle f_1^t, x+e\rangle^2 - \langle f_1^t, x\rangle^2 - \langle f_2^t, x+e\rangle^2 + \langle f_2^t, x\rangle^2 \, \mathrm{d}t \right| \\
&= \left| \int_0^T \langle f_1^t, 2x+e\rangle \langle f_1^t, e\rangle - \langle f_2^t, 2x+e\rangle \langle f_2^t, e\rangle \, \mathrm{d}t \right| \\
&\leq L \max\left\{ \sup_{t\in[0,T]} \left|\langle f_1^t, e\rangle\right|, \sup_{t\in[0,T]} \left|\langle f_2^t, e\rangle\right| \right\} = \sup_{g\in\mathcal{G}} |\langle g, e\rangle| \ ,
\end{aligned}
$$

where L is any upper bound on the term

$$
\begin{aligned}
&\int_0^T \left|\langle f_1^t, 2x+e\rangle\right| + \left|\langle f_2^t, 2x+e\rangle\right| \, \mathrm{d}t \\
&\qquad \leq 2\int_0^T \max_{\tau\in[0,T]} \left(\left|f_1^t(\tau)\right| + \left|f_2^t(\tau)\right|\right) \mathrm{d}t + \ 2T \sup_{g\in\mathcal{G}} |\langle g, e\rangle| \\
&\qquad \leq T\left(22 + \frac{\pi}{2} + 2\sup_{g\in\mathcal{G}} |\langle g, e\rangle|\right) \ . \qquad (23)
\end{aligned}
$$

In order to obtain an explicit bound, we need to further analyze the term $\sup_{g\in\mathcal{G}} |\langle g, e\rangle|$. First of all, we have

$$
\overline{D}_C(M) \ \leq \ \gamma \ = \ \sup_{x\in C} \|x\|_2 \ = \ \sqrt{T} \, .
$$

Next, recalling that the Legendre approximation error for any smooth function $g \in L_2[0,T]$ is bounded as

$$
D(M,g) := \|g - P_M(g)\|_2 \ \leq \ \frac{\mu_{M+1}\sqrt{T}}{(M+1)!}\left(\frac{T}{M}\right)^M \quad \text{with} \quad \mu_i := \sup_{\xi\in[0,T]} \left|\frac{\partial^i g}{\partial t^i}(\xi)\right|
$$

for all $M \geq 1$, and working out explicit bounds on the derivatives of the functions f_1^t and f_2^t, we obtain

$$
\begin{aligned}
\forall M \in \mathbb{N}^+, \quad \sup_{g\in\mathcal{G}} D(M,g) &\leq \frac{T^{\frac{3}{2}}}{(M+1)!}\left(\frac{T}{M}\right)^M \left(\frac{1}{2} + \frac{M}{\pi}\right)\left(\frac{\pi}{2T}\right)^M \\
&\leq \frac{3}{4}\frac{T^{\frac{3}{2}}}{(M+1)!}\left(\frac{\pi}{2M}\right)^{M-1} \, .
\end{aligned}
$$

It follows by Theorem 1 that

$$
\sup_{g\in\mathcal{G}} |\langle g, e\rangle| \ \leq \ \overline{\mathcal{R}}_C(M,\mathcal{G}) \ = \ \sup_{g\in\mathcal{G}} \overline{D}_C(M) D(M,g) \ \leq \ \frac{3}{4}\frac{T^2}{(M+1)!}\left(\frac{\pi}{2M}\right)^{M-1} \, .
$$

Combining all the bounds and substituting $T = 10$ shows that the constant $L = 611$ satisfies the condition (23).

Based on the foregoing developments and the considerations in Sect. 3.2, a simple bound $\Delta_M(A)$

on the approximation error satisfying (18) can be obtained as

$$\forall (M,A) \in \mathbb{N}^+ \times \mathcal{S}_M, \quad \Delta_M(A) = \frac{45825}{(M+1)!}\left(\frac{\pi}{2M}\right)^{M-1} .$$

Although rather loose for very small M, this estimate converges quickly to 0 for larger M; for instance, $\Delta_7(A) \leq 2 \cdot 10^{-4}$. Note also that, in a practical implementation, the computation of $\Delta_M(A)$ – and also to validate the generalized Lipschitz constant L – could be automated using computer algebra programs, such as Chebfun (http://www.chebfun.org/) [18] or MC++ (http://omega-icl.bitbucket.org/mcpp/) [39].

With regards to the computation of bounds $L_M^0(A)$ and $U_M^0(A)$ satisfying (17), we note that $F(x)$ can be interpreted as a quadratic form in x,

$$F\left(\sum_{i=0}^{M} a_i \Phi_i\right) \;=\; a^{\mathsf{T}} \mathcal{Q} a \;,$$

with the elements of the matrix $\mathcal{Q}$ given by

$$\forall j,k \in \{0,\ldots,M\}, \quad q_{j,k} \;=\; \int_0^T \left\{ \langle f_1^t, \Phi_j \rangle \langle f_1^t, \Phi_k \rangle - \langle f_2^t, \Phi_j \rangle \langle f_2^t, \Phi_k \rangle \right\} \, \mathrm{d}t \;.$$

Of the available approaches [20, 43, 45] to compute bounds $L_M^0(A)$ and $U_M^0(A)$ such that

$$L_M^0(A) \;\leq\; \min_{a \in A} \, a^{\mathsf{T}} \, \mathcal{Q} \, a \;\leq\; U_M^0(A)$$

for interval boxes $A \subseteq \mathbb{R}^{M+1}$, we use standard LMI relaxation techniques [23] here.

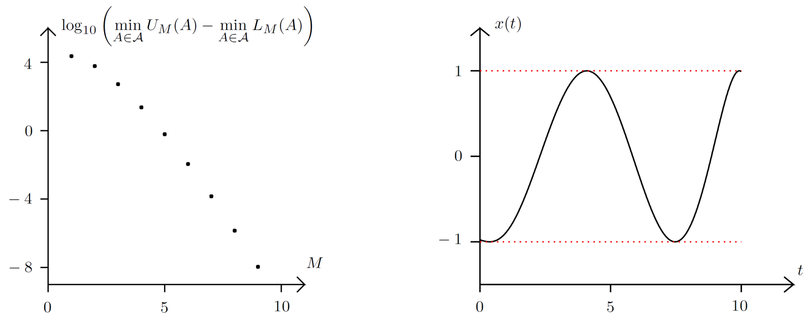

Figure 1: Results of Algorithm 1 applied to Problem (22) for $\epsilon = 10^{-5}$ and $\varrho = 1$. Left: Gap between upper and lower bounds as a function of the lifted subspace dimension M. Right: A globally ϵ-suboptimal solution x.

At this point, we have all the elements needed for implementing Algorithm 1 for Problem (22). On selecting the termination tolerance $\epsilon = 10^{-5}$ and the lifting parameter $\varrho = 1$, Algorithm 1 terminates after less than 100 iterations and applies 8 lifting operations (starting with $M = 1$). The corresponding decrease in the gap between upper and lower bounds as a function of the lifted subspace dimension M – immediately after each lifting operation – is shown on the left plot of Fig. 1. Upon convergence, the infimum of (22) is bracketed as

$$-0.16811 \;\leq\; \inf_{x \in C} \, F(x) \;\leq\; -0.16812 \, ,$$

and a corresponding ϵ-global solution x is reported on the right plot of Fig. 1; the symmetric function $(-x)$ provides another ϵ-global solution for this problem. Overall, this case study demonstrates that the proposed branch-and-lift algorithm is thus capable of solving such non-convex and infinite-dimensional optimization problem to global optimality within reasonable computational effort.

6 Conclusions

This paper has presented a complete search algorithm, called branch-and-lift, for global optimization of problems with a non-convex cost functional and a bounded and convex constraint sets defined on a Hilbert space. A key contribution is the determination of run-time complexity bounds for branch-and-lift that are independent of the number of variables in the optimization problem, as long as the cost functional is strongly Lipschitz continuous with respect to a regular subset of that Hilbert space. The corresponding convergence conditions are satisfied for a large class of practically relevant optimal control problems with embedded ODE or PDE systems. In particular, the complexity analysis in this paper implies that branch-and-lift can be applied to solve potentially non-convex and infinite-dimensional optimization problems without needing a-priori knowledge about the existence or regularity of minimizers, as the run-time bounds solely depend on the structural and regularity properties of the cost functional as well as the underlying Hilbert space and the geometry of the constraint set. This might pave the way for a new complexity analysis of optimization problems, whereby the “complexity” or “hardness” of a problem does not necessarily depend on their number of optimization variables. In order to demonstrate that these algorithmic ideas and complexity analysis are not of pure theoretical interest only, the practical applicability of branch-and-lift has been illustrated with a numerical case study for a problem of calculus of variations. The case study of an optimal control problem in [28] provides another illustration.

Acknowledgements This paper is based upon work supported by the Engineering and Physical Sciences Research Council (EPSRC) under Grant EP/J006572/1, National Natural Science Foundation of China (NSFC) under Grant 61473185, and ShanghaiTech University under Grant F-0203-14-012. Financial support from Marie Curie Career Integration Grant PCIG09-GA-2011-293953 and from the Centre of Process Systems Engineering (CPSE) of Imperial College is gratefully acknowledged.

References

[1] N.I. Akhiezer. The classical moment problem and some related questions in analysis. Translated by N. Kemmer, Hafner Publishing Co., New York, 1965.

[2] J. Albersmeyer and M. Diehl. The Lifted Newton Method and its Application in Optimization. SIAM Journal on Optimization, 20(3):1655–1684, 2010.

[3] E.J. Anderson and P. Nash. Linear Programming in infinite-dimensional spaces. Wiley, 1987.

[4] D. Bampou, D. Kuhn. Polynomial Approximations for Continuous Linear Programs. SIAM Journal on Optimization, 22(2):628–648, 2012.

[5] R. Beals Analysis: An Introduction. Cambridge University Press, New York, 2004.

[6] R.E. Bellman Dynamic Programming. Princeton University Press, Princeton, 1957.

[7] M.P. Bendsøe, O. Sigmund. Topology Optimization: Theory, Methods, and Applications. Springer, Berlin, 2004.

[8] J.T. Betts. Practical Methods for Optimal Control Using Nonlinear Programming. 2nd edn, Advances in Design and Control Series. SIAM, Philadelphia, 2010.

[9] L.T. Biegler. Solution of dynamic optimization problems by successive quadratic programming and orthogonal collocation. Computers & Chemical Engineering, 8:243–248, 1984.

[10] H.G. Bock and K.J. Plitt. A multiple shooting algorithm for direct solution of optimal control problems. In: Proceedings 9th IFAC World Congress Budapest, pages 243–247. Pergamon Press, 1984.

[11] A. Bompadre, A. Mitsos. Convergence rate of McCormick relaxations. Journal of Global Optimization, 52(1)::1–28, 2012.

[12] A. Bompadre, A. Mitsos, A., B. Chachuat. Convergence analysis of Taylor and McCormick-Taylor models. Journal of Global Optimization, 57(1):75–114, 2013.

[13] S. Boyd and L. Vandenberghe. Convex Optimization. University Press, Cambridge, 2004.

[14] A.E. Bryson, Y. Ho. Applied Optimal Control. Hemisphere, Washington, 1975.

[15] R. Buie and J. Abrham. Numerical solutions to continuous linear programming problems. Zeitschrift für Operations Research, 17(3):107-117, 1973.

[16] O. Devolder, F. Glineur, Y. Nesterov. Solving Infinite-Dimensional Optimization Problems by Polynomial Approximation In: Recent Advances in Optimization and its Applications in Engineering, pp. 31-40, Springer, Berlin Heidelberg, 2010.

[17] Z. Ditzian, V. Totik. Moduli of smoothness. Springer-Verlag, 1987.

[18] T.A. Driscoll, N. Hale, L.N. Trefethen. Chebfun Guide, Pafnuty Publications, Oxford, 2014.

[19] C.A. Floudas. Deterministic Global Optimization: Theory, Methods, and Applications. Kluwer, Dordrecht, 1999.

[20] M.X. Goemans, D.P. Williamson. Improved approximation algorithms for Maximum Cut and satisfiability problems using semidefinite programming. Journal of the ACM, 42(6):1115–1145, 1995.

[21] D. Gottlieb, C.W. Shu. On the Gibbs phenomenon and its resolution. SIAM Reviews, 39(4):644–668, 1997.

[22] L. Grüne, W. Semmler. Using dynamic programming with adaptive grid scheme to solve nonlinear dynamic models in economics. Comput. Econ. Finance, 99, 2002.

[23] D. Henrion, S. Tarbouriech, D. Arzelier. LMI Approximations for the Radius of the Intersection of Ellipsoids: A Survey. Journal of Optimization Theory & Applications, 108(1):1–28, 2001.

[24] D. Henrion and M. Korda. Convex computation of the region of attraction of polynomial control systems. IEEE Transactions on Automatic Control, 59(2):297–312, 2014.

[25] M. Hinze, R. Pinnau, M. Ulbrich and S. Ulbrich. Optimization with PDE constraints, Springer-Verlag, 2009.

[26] R. Horst, H. Tuy. Global Optimization: Deterministic Approaches. 3rd edn. Springer-Verlag, Berlin, Germany, 1996.

[27] B. Houska, H.J. Ferreau, M. Diehl. ACADO Toolkit—An open-source framework for automatic control and dynamic optimization. Optimal Control Applications & Methods 32:298–312, 2011.

[28] B. Houska, B. Chachuat. Branch-and-Lift Algorithm for Deterministic Global Optimization in Nonlinear Optimal Control. Journal of Optimization Theory & Applications, 162(1):208–248, 2014.

[29] B. Houska, M.E. Villanueva, B. Chachuat. Stable Set-Valued Integration of Nonlinear Dynamic Systems using Affine Set Parameterizations. SIAM Journal on Numerical Analysis, 53(5):2307–2328, 2015.

[30] D. Jackson. The theory of Approximation. AMS Colloquium Publication, Volume XI, New York 1930.

[31] Y. Katznelson. An introduction to harmonic analysis. 2nd edn, Dover Publications, New York, 1976.

[32] M. Korda, D. Henrion, C.N. Jones. Convex computation of the maximum controlled invariant set for polynomial control systems. SIAM Journal on Control & Optimization, 52(5):2944–2969, 2014

[33] J.B. Lasserre. Moments, Positive Polynomials and Their Applications. Imperial College Press, 2009.

[34] Y. Lin, M.A. Stadtherr. Validated solutions of initial value problems for parametric ODEs. Applied Numerical Mathematics, 57(10):1145–1162, 2007.

[35] X. Luo and D. Bertsimas. A new algorithm for state-constrained separated continuous linear programs. SIAM Journal on Control & Optimization, 37:177–210, 1998.

[36] R. Luss. Optimal control by dynamic programming using systematic reduction in grid size. International Journal of Control, 51(5):995–1013, 1990.

[37] G.P. McCormick. Computability of global solutions to factorable nonconvex programs: Part I–Convex underestimating problems. Mathematical Programming, 10:147–175, 1976.

[38] R. Misener, C.A. Floudas. ANTIGONE: Algorithms for continuous/integer global optimization of nonlinear equations. Journal of Global Optimization, 59(2-3):503–526.

[39] A. Mitsos, B. Chachuat, P.I. Barton. McCormick Based Relaxations of Algorithms. SIAM Journal on Optimization, 20:573–601, 2009.

[40] R.E. Moore. Methods and applications of interval analysis. SIAM Studies in Applied Mathematics, SIAM, 1979.

[41] B.S. Mordukhovich. Variational Analysis and Generalized Differentiation I: Basic Theory. Springer, 2006.

[42] M. Neher, K.R. Jackson, N.S. Nedialkov. On Taylor model based integration of ODEs. SIAM Journal on Numerical Analysis, 45:236–262, 2007.

[43] A. Nemirovski, C. Roos, T. Terlaky. On maximization of quadratic form over intersection of ellipsoids with common center. Mathematical Programming, 86(3):463–473, 1999.

[44] Y. Nesterov, A. Nemirovskii. Interior-Point Polynomial Methods in Convex Programming. SIAM, 1994.

[45] Y. Nesterov. Semidefinite relaxation and non-convex quadratic optimization. Optimization Methods & Software, 12:1–20, 1997.

[46] Y. Nesterov. Squared functional systems and optimization problems. In High Performance Optimization, pp. 405–440, Kluwer Academic Publishers, 2000.

[47] A. Neumaier. Taylor forms – Use and limits. Reliable Computing, 9(1):43–79, 2002.

[48] A. Neumaier. Complete search in continuous global optimization and constraint satisfaction. Acta Numerica, 13:271–369, 2004.

[49] P.A. Parrilo. Polynomial games and sum of squares optimization. In: Proceedings of the 45th IEEE Conference on Decision & Control, pp. 2855–2860, San Diego (CA), 2006.

[50] L.S. Pontryagin, V.G. Boltyanskii, R.V. Gamkrelidze, E.F. Mishchenko. The Mathematical Theory of Optimal Processes. Wiley, New York, 1962.

[51] E.B. Saff, V. Totik. Polynomial approximation of piecewise analytic functions. Journal of the London Mathematical Society, 39(2):487–498, 1989.

[52] N.V. Sahinidis. A general purpose global optimization software package. Journal of Global Optimization, 8(2):201–205, 1996.

[53] J.K. Scott, B. Chachuat, P.I. Barton. Nonlinear convex and concave relaxations for the solutions of parametric ODEs. Optimal Control Applications & Methods, 34(2):145–163, 2013.

[54] O. von Stryk, R. Bulirsch. Direct and indirect methods for trajectory optimization. Annals of Operations Research, 37:357–373, 1992.

[55] M. Tawarmalani, N.V. Sahinidis. A polyhedral branch-and-cut approach to global optimization. Mathematical Programming, 103(2):225–249, 2005.

[56] M.E. Villanueva, B. Houska, B. Chachuat. Unified Framework for the Propagation of Continuous-Time Enclosures for Parametric Nonlinear ODEs. Journal of Global Optimization, 62(3):575–613, 2015.

[57] R. Vinter. Optimal Control. Springer, 2010.

[58] H. Wang, S. Xiang. On the convergence rates of Legendre approximation. Mathematics of Computation, 81(278):861–877, 2012.