Math. Program., Ser. A
https://doi.org/10.1007/s10107-017-1215-7

FULL LENGTH PAPER

Global optimization in Hilbert space

Boris Houska[1] · **Benoît Chachuat**[2]

Received: 28 October 2015 / Accepted: 28 November 2017
© The Author(s) 2017. This article is an open access publication

Abstract We propose 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 $\varepsilon > 0$. 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

Mathematics Subject Classification 49M30 · 65K10 · 90C26 · 93B40

1 Introduction

Infinite-dimensional optimization problems arise in many research fields, including optimal control [7,8,24,54], optimization with partial differential equations (PDE)

✉ Benoît Chachuat
b.chachuat@imperial.ac.uk

1 School of Information Science and Technology, ShanghaiTech University, 319 Yueyang Road, Shanghai 200031, China

2 Department of Chemical Engineering, Centre for Process Systems Engineering, Imperial College London, South Kensington Campus, London SW7 2AZ, UK

Published online: 16 December 2017

embedded [22], and shape/topology optimization [5]. 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 poorly with the number of optimization variables. For instance, the worst-case run-time of spatial branch-and-bound [17,44] scales exponentially with the number of optimization variables. In contrast, algorithms for solving convex optimization problems in polynomial run-time are known [11,40], 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 [14], as well as certain tailored algorithms for continuous linear programming [4,13,32].

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 [46], and a number of local optimal control algorithms are based on this analysis [6,12,51,54]. More generally, approaches in the classical field of variational analysis [37] 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 of our knowledge, there currently are no generic complete-search algorithms 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 [49,52] or ANTIGONE [34] 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 [7,8]. 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 $\varepsilon > 0$, if

$$\forall x \in C, \quad F(x^*) \leq F(x) + \varepsilon\,.$$

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

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

Our main objective in this 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 $\varepsilon > 0$, 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 [22]. 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. 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 nonetheless 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 constitutes 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 on global optimal control [25], we develop herein 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 complexity in numerical optimization does not necessarily depend on whether the problem at hand being 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 such spaces. Our focusing on Hilbert space is motivated by the ability to construct an orthogonal basis $\Phi_0, \Phi_1, \ldots \in H$ such that

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

for some scalars $\sigma_0, \sigma_1, \ldots \in \mathbb{R}^{++}$. We make the following assumption throughout the paper:

Assumption 2 The basis functions Φ_k are uniformly bounded with respect to $\|\cdot\|_H$.

Equipped with 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 to ask 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 $C \subseteq H$, we define

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

Lemma 1 *Under Assumption 1, the function $\overline{D}_C : \mathbb{N} \to \mathbb{R}$ is uniformly bounded from above by γ.*

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

$$\left[\overline{D}_C(M)\right]^2 = \sup_{x \in C} \| x - P_M(x) \|_H^2 \;\leq\; \| x \|_H^2 \;.$$

The result follows by Assumption 1. □

Despite being uniformly bounded, the function $\overline{D}_C(M)$ may not converge to zero as $M \to \infty$ in an infinite-dimensional Hilbert space in general. Such lack of convergence 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 $x_k := \Phi_{k+1}$. For all $k \in \mathbb{N}$, we have $P_k(x_k) = 0$, and therefore

$$\overline{D}_C(k) \;\geq\; D(k, x_k) \;=\; \| x_k - P_k(x_k) \|_H = \| x_k \|_H \;=\; 1 \,.$$

□

This behavior is unfortunate because the existence of minimizers to Problem (1) cannot be ascertained without making further regularity assumptions. Moreover, for a

sequence $(x_k)_{k\in\mathbb{N}}$ of feasible points of Problem (1) converging to an infimum, it could be that

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

That is, any attempt to approximate the infimum by constructing a sequence of finite parameterizations of the optimization variable x could in principle be unsuccessful.

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. But instead of making assumptions about the existence, or even the regularity, of the minimizers of Problem (1), we shall impose suitable regularity conditions on the objective function F in (1). In preparation for this analysis, we start by formalizing a particular notion of regularity for the elements of H.

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

$$\lim_{M\to\infty} R_C(M,g)\ =\ 0\quad\text{with}\quad R_C(M,g):=\overline{D}_C(M)D(M,g)\ . \tag{3}$$

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

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

$$\forall M\in\mathbb{N},\quad \sup_{x\in C}\ |\,\langle g, x-P_M(x)\rangle\,|\ \leq\ R_C(M,g)\ . \tag{4}$$

In the particular case of g *being a regular element for* C*, we have*

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

Proof Let $M\in\mathbb{N}$, and consider the optimization problem

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

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 to each other, we have $\langle\Phi_k,w\rangle=0$ for all $k\in\{0,\ldots,M\}$, and it follows that

$$\overline{V}_M\ \leq\ \sup_{w\in H}\ \langle g,w\rangle\quad\text{s.t.}\quad \langle\Phi_0,w\rangle\ =\ \cdots\ =\ \langle\Phi_M,w\rangle\ =\ 0\ ,\quad \|w\|_H\ \leq\ \overline{D}_C(M)\ .$$

Next, we use duality to obtain

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

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

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

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

$$\overline{V}_M \leq \|g - P_M(g)\|_H \, \overline{D}_C(M) = R_C(M, g) .$$

The optimal value of the minimization problem

$$\underline{V}_M := \inf_{x \in C} \langle g, x - P_M(x) \rangle$$

can be estimated analogously, giving $\underline{V}_M \geq -R_C(M, g)$, 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 Legendre polynomials as orthogonal 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 we choose the Legendre polynomials on the interval $[0, 1]$ with weighting factors $\sigma_k = \frac{1}{2k+1}$ as orthogonal basis functions $(\Phi_k)_{k \in \mathbb{N}}$. Our focus is on piecewise smooth functions $g : [0, 1] \to \mathbb{R}$ 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 are numerous results on approximating functions using polynomials, including convergence rate estimates [15]. One such result in [48] 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 $g : [0, 1] \to \mathbb{R}$ with a finite number of singularities:

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

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

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

Convergence rate estimates for k-times differentiable and piecewise smooth functions can be obtained in a similar way, using for instance the results in [15,48]. □

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

Definition 4 A set $G \subseteq H$ is said to be *regular for* C if

$$\lim_{M\to\infty} \overline{R}_C(M,G) \;=\; 0 \quad \text{with} \quad \overline{R}_C(M,G) \;:=\; \sup_{g\in G} R_C(M,g) \,.$$

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

Corollary 1 *For any regular set* $G \subseteq H$*, we have*

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

Remark 2 While any subset of the Euclidean space $\mathbb{R}^n$ is trivially regular for a given bounded subset $C \subset \mathbb{R}^n$, only certain subsets/subspaces of an infinite-dimensional Hilbert space happen to be regular. Consider for instance the space of square-integrable functions, $H := L_2[a,b]$, and let G^p be any subset of p-times differentiable functions on $[a,b]$, with uniformly Lipschitz-continuous p-th derivatives. It can be shown—e.g.,

[1] 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).

from the analysis in [27] using the standard trigonometric Fourier basis, or from the results in [55] using the Legendre polynomial basis—that

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

for any bounded constraint set $C \subset L_2[a, b]$, and G^p is thereby regular for C. This leads to a rather typical situation, whereby the stronger the regularity assumptions on the function class, the faster the convergence of the associated worst-case convergence rate $R(\cdot, G^p)$—an increase in the convergence rate order $\log(M) M^{-p-1}$ with p in this instance. In the limit of smooth ($\mathscr{C}^\infty$) functions, it can even be established—e.g., using standard results from Fourier analysis [19,28]—that the convergence rate becomes exponential,

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

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

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

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

$$\left\| x_{0.5} - p_{0.5}^M \right\|_2 \;\leq\; \mathbf{O}\left(M^{-1/2}\right).$$

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

$$\forall \tau \in [0, 1], \quad p_t^M(\tau) \;:=\; p_{0.5}^M\left(\frac{1 - t + \tau}{2}\right).$$

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

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

where the constant $K < \infty$ is independent of t or M, we have $\overline{R}_C(M, G_t) \leq \mathbf{O}\left(M^{-1/2}\right)$.

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

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

is regular in H, and also satisfies $\overline{R}_C(M, G_f) \leq \mathbf{O}\left(M^{-1/2}\right)$. This result can be established by writing the elements in G_f as the product between the piecewise smooth function f and the function x_t, and then approximating the factors separately. □

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

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

$$\forall e \in H, \quad \sup_{x \in C} |F(x+e) - F(x)| \leq L \sup_{g \in G} |\langle g, e \rangle| \,. \tag{7}$$

Remark 3 In the special case of an affine functional F, given by

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

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

Remark 4 Consider the finite-dimensional Euclidean space $\mathbb{R}^n$, a bounded subset $S \subset \mathbb{R}^n$, and a continuously-differentiable function $F : \mathbb{R}^n \to \mathbb{R}$ whose first derivative is bounded in the subset $G \subset \mathbb{R}^n$. By the mean-value theorem, F satisfies

$$\begin{aligned} \forall e \in \mathbb{R}^n, \quad \sup_{x \in S} |F(x+e) - F(x)| &= \sup_{x \in S} \left| \int_0^1 \left\langle \frac{\partial F}{\partial x}(x + \eta e), e \right\rangle \mathrm{d}\eta \right| \\ &\leq \sup_{g \in G} |\langle g, e \rangle| \,. \end{aligned}$$

Thus, any continuously differentiable function with a bounded first derivative is strongly Lipschitz-continuous on any bounded subset of $\mathbb{R}^n$. This result can be generalized to certain classes of functionals in infinite-dimensional Hilbert space. For instance, let $F : H \to \mathbb{R}$ be Fréchet differentiable, such that

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

and let the set of Fréchet derivatives $G := \{DF(x) \mid x \in H\} \subseteq H$ be both bounded and regular on C. Then, F is strongly Lipschitz-continuous on C.

The following two examples investigate strong Lipschitz continuity for certain classes of functionals in the practical space of square-integrable functions with the Legendre polynomials as orthogonal basis functions. The first one (Example 3) illustrates the case of a functional that is not strongly Lipschitz-continuous; the second one (Example 4) identifies a broad class of strongly Lipschitz-continuous functionals defined via the solution of an embedded ODE system. The intention here is to help the reader develop an intuition that strongly Lipschitz-continuous functionals occur naturally in many, although not all, problems of practical relevance.

Example 3 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 orthogonal basis functions $(\Phi_k)_{k\in\mathbb{N}}$ as the 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 family of sets defined by

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

If the condition (7) were to hold for some bounded and regular set G, we would have by Theorem 1 that

$$\sup_{\substack{e\in E_M,\\ x\in C}} |F(x+e) - F(x)| \ \leq \ L \sup_{\substack{e\in E_M,\\ g\in G}} |\langle g, e\rangle| \ = \ L \sup_{\substack{x\in C,\\ g\in G}} |\langle g, x - P_M(x)\rangle| \, ,$$

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

$$\forall M \in \mathbb{N}, \quad \sup_{\substack{e\in E_M,\\ x\in C}} |F(x+e) - F(x)| \ \geq \ \sup_{e\in E_M} F(e) \ = \ \sup_{x\in C} \|x - P_M(x)\|_2^2 \ = \ 1 \, .$$

Therefore, the regularity condition (7) may not be satisfied for any bounded and regular set G, and F is not strongly Lipschitz-continuous on C. □

Remark 5 The result that the functional F in Example 3 is not strongly Lipschitz-continuous on C is not in contradiction with Remark 4. Although F is Fréchet differentiable in $L_2[0,1]$, the corresponding set G of the Fréchet derivatives of F is indeed unbounded.

Example 4 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 orthogonal basis functions $(\Phi_k)_{k\in\mathbb{N}}$ as the 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)

$$\forall t \in [0,1], \quad \frac{\partial x}{\partial t}(t,u) = f(x(t,u)) + Bu(t) \quad \text{with} \quad x(0,u) = 0 \, , \tag{8}$$

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, 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 defect $\delta(t, u, e) := x(t, u + e) - x(t, u)$ satisfies the differential equation

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

with $\delta(0, u, e) = 0$ and $\Lambda(t, u, e) := \int_0^1 \frac{\partial f}{\partial x}(x(t, u) + \eta\delta(t, u, e))\,\mathrm{d}\eta$. The right-hand-side function f being globally Lipschitz-continuous, we have for any given smooth matrix-valued function $A : [0, 1] \to \mathbb{R}^{n\times n}$,

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

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

$$\forall t \in [0, 1], \qquad \delta_{\mathrm{l}}(t, e) = A(t)\delta_{\mathrm{l}}(t, e) + Be(t) \tag{9}$$

$$\delta_{\mathrm{n}}(t, u, e, \delta_{\mathrm{l}}) = \Lambda(t, u, e)\delta_{\mathrm{n}}(t, u, e, \delta_{\mathrm{l}}) + [\Lambda(t, u, e) - A(t)]\delta_{\mathrm{l}}(t, e) \tag{10}$$

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

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

Regarding the linear term δ_{l} first, we have

$$\forall s \in [0, 1], \quad c^{\mathsf{T}}\delta_{\mathrm{l}}(s, e) = \langle g_s, e\rangle \tag{11}$$

with

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

where $\Gamma(t,\tau)$ denotes the fundamental solution of the linear ODE (9) such that

$$\forall (\tau,t) \in [0,1]^2, \quad \frac{\partial}{\partial t}\Gamma(t,\tau) \;=\; A(t)\Gamma(t,\tau) \quad \text{with} \quad G(\tau,\tau) = I \;.$$

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

$$\overline{R}_C(M,G) \;\le\; \mathbf{O}\left(M^{1/2}\right) .$$

Regarding the nonlinear term δ_n, since the function Λ is uniformly bounded, applying Gronwall's lemma to the ODE (10) gives

$$\begin{aligned} \forall (t,u,e) \in [0,1] \times C \times H, \quad c^\mathsf{T}\delta_\mathrm{n}(t,u,e,\delta_\mathrm{l}) \;&\le \ell \exp(\ell) \sup_{s\in[0,1]} |c^\mathsf{T}\delta_\mathrm{l}(s,e)| \\ &\le \ell \exp(\ell) \sup_{g\in G} |\langle g, e\rangle| \,, \end{aligned} \tag{12}$$

for some constant $\ell < \infty$. Finally, combining (11) and (12) shows that F satisfies the condition (7) with $L := 1 + \ell \exp(\ell)$, thus F is 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 as well as problems in the calculus of variations. In fact, it turns out that strong Lipschitzness still holds in replacing the constant matrix B in (8) 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. Strong Lipschitzness could nevertheless be recovered by restricting condition (7) in Definition 5 as

$$\forall e \in E_C, \quad \sup_{x\in C} |F(x+e) - F(x)| \;\le\; L \sup_{g\in G} |\langle g, e\rangle| \;,$$

with the projection error set $E_C := \{P_M(x) - x \mid x \in C,\ M \in \mathbb{N}\} \subset H$, and also restricting the constraint set C to only contain uniformly bounded and Lipschitz-continuous functions in $L_2[0,1]$ with uniformly bounded Lipschitz constants.

We close this section with a brief analysis of the relationship between strong and classical Lipschitzness in infinite-dimensional Hilbert space.

Lemma 2 *Every strongly Lipschitz-continuous functional $F : H \to \mathbb{R}$ on C is also Lipschitz-continuous on C.*

Proof Let G be a bounded and regular subset of H on C such that the condition (7) is satisfied. Since G is bounded, there exists a constant constant $\alpha < \infty$ such that $\sup_{g \in G} \|g\|_H \leq \alpha$. Applying the Cauchy–Schwarz inequality to the right-hand side of (7) gives

$$\forall e \in H, \quad \sup_{x \in C} |F(x+e) - F(x)| \leq L\,\alpha\, \|e\|_H \ ,$$

and so F is Lipschitz-continuous on C. □

Remark 7 With regularity of the set G alone, i.e. without boundedness, the condition (7) may not imply Lipschitz-continuity, or even continuity of F. As a counter-example, let $G := \text{span}\,(\Phi_0, \Phi_1, \ldots, \Phi_N)$ be the subspace spanned by the first N basis functions in the infinite-dimensional Hilbert space H. It is clear that G is regular on any bounded set $C \subset H$ since $\overline{R}_C(M, G) = 0$ for all $M \geq N$. Now, let the functional $F : H \to \mathbb{R}$ given by

$$F(x) := \begin{cases} 0 & \text{if } \langle \hat{g}, x \rangle \leq 0 \\ 1 & \text{otherwise} \end{cases}$$

for some $\hat{g} \in G$. For every $(x, e) \in C \times H$, we have

$$|F(x+e) - F(x)| \leq \left\{ \begin{array}{l} 0 \text{ if } \langle \hat{g}, e \rangle = 0 \\ 1 \text{ otherwise} \end{array} \right\} \leq \left\{ \begin{array}{ll} 0 & \text{if } P_N(e) = 0 \\ \infty & \text{otherwise} \end{array} \right\} = \sup_{g \in G} |\langle g, e \rangle| \ .$$

Therefore, despite being discontinuous, the condition (7) is indeed satisfied.

Remark 8 In general, Lipschitz-continuity does not imply strong Lipschitz-continuity in an infinite-dimensional Hilbert space. A counter-example is easily contrived for the functional $F : L_2[0, 1] \to \mathbb{R}$ given by

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

Although this functional is Lipschitz-continuous, it can be shown by a similar argument as in Example 3 that it is not strongly Lipschitz-continuous.

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 [23] 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 $\mathscr{A} := \{A_1, \ldots, A_k\}$ 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 $(M, A) \in \mathbb{N} \times \mathscr{P}(\mathbb{R}^{M+1})$, we associate a subregion $X_M(A)$ of H given by

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

Moreover, we say that the set A is *infeasible* if $X_M(A) = \varnothing$.

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

The ability to detect infeasibility of a set $A \in \mathbb{S}^{M+1}$ 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

$$d_C(A) := \min_{x,y \in H} \|x - y\|_H \quad \text{s.t.} \quad \left(\frac{\langle y, \Phi_0 \rangle}{\sigma_0}, \ldots, \frac{\langle y, \Phi_M \rangle}{\sigma_M} \right)^{\mathsf{T}} \in A \,, \quad x \in C. \tag{13}$$

It readily follows from the Cauchy–Schwarz inequality that

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

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 (13) 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 [4,14]. 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 to have a simple geometry in order to avoid numerical issues in solving feasibility problems of the form (13). We shall therefore assume, from this point onwards, that infeasibility can be verified with high numerical accuracy for any set $A \in \mathbb{S}^{M+1}$.

A *branching* operation subdivides any set $A \in \mathbb{S}^{M+1}$ in the partition $\mathscr{A}$ into two compact subsets $A_\mathrm{l}, A_\mathrm{r} \in \mathbb{S}^{M+1}$ such that $A_\mathrm{l} \cup A_\mathrm{r} \supseteq A$, thereby updating the partition as

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

On the other hand, a *lifting* operation essentially lifts any set $A \in \mathbb{S}^{M+1}$ into a higher-dimensional space under the function $\Gamma_M : \mathbb{S}^{M+1} \to \mathbb{S}^{M+2}$, defined such that

$$\forall A \in \mathbb{S}^{M+1}, \quad X_M(A) \subseteq X_{M+1}(\Gamma_M(A))\,.$$

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 [30,42], by considering the following optimization subproblems:

$$\underline{a}_{M+1}(A) \leq \min_{x\in X_M(A)} \frac{\langle x, \Phi_{M+1}\rangle}{\sigma_{M+1}} \quad \text{and} \quad \overline{a}_{M+1}(A) \geq \max_{x\in X_M(A)} \frac{\langle x, \Phi_{M+1}\rangle}{\sigma_{M+1}}. \tag{14}$$

Although both optimization problems in (14) are convex when 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 [4,14], which could provide tighter bounds in practical applications.

A number of remarks are in order:

Remark 9 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 [25], 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 10 Many recent optimization techniques for global optimization are based on the theory of positive polynomials and their associated linear matrix inequality (LMI) approximations [30,45], 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. Thereafter, functions $L_M, U_M : \mathbb{S}^{M+1} \to \mathbb{R}$ such that

$$\forall A \in \mathbb{S}^{M+1}, \quad L_M(A) \;\leq\; \inf_{x \in X_M(A)} F(x) \;\leq\; U_M(A)\,, \tag{15}$$

shall be call lower- and upper-bounding functions of the functional F, respectively. A simple approach to constructing these lower and upper bounds relies on the following two-step decomposition:

1. Compute bounds $L_M^0(A)$ and $U_M^0(A)$ on the finite-dimensional approximation of F as

$$\forall A \in \mathbb{S}^{M+1}, \quad L_M^0(A) \;\leq\; \inf_{a \in A} F\left(\sum_{i=0}^{M} a_i \Phi_i\right) \;\leq\; U_M^0(A)\,. \tag{16}$$

   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, including interval arithmetic [36], McCormick relaxations [9,33], and Taylor/Chebyshev model arithmetic [10,43,47]. 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 [26,31,38,50,53]; or via hierarchies of LMI relaxations as in [21,29].

2. Compute a bound $\Delta_M(A)$ on the approximation errors such that

$$\forall A \in \mathbb{S}^{M+1}, \quad \left| \inf_{x \in X_M(A)} F(x) - \inf_{a \in A} F\left(\sum_{i=0}^{M} a_i \Phi_i \right) \right| \leq \Delta_M(A) \,. \tag{17}$$

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

By construction, the lower-bounding function $L_M(A) := L^0_M(A) - \Delta_M(A)$ and the upper-bounding function $U_M(A) := U^0_M(A) + \Delta_M(A)$ trivially satisfy (15). Moreover, when the set $A \in \mathbb{S}^{M+1}$ is infeasible—see related discussion in Sect. 3.1—we may set $\Delta_M(A) = L_M(A) = U_M(A) = \infty$.

We state the following assumptions in anticipation of the convergence analysis in Sect. 4.

Assumption 3 The cost functional F in Problem (1) is strongly Lipschitz-continuous on C, with the condition (7) holding for the constant $L < \infty$ and the bounded regular subset $G \subset H$.

Remark 11 Under Assumption 3, Lemma 2 implies that

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

for a Lipschitz constant $L' \geq L \, \sup_{g \in G} \|g\|_H$. Thus, if Assumption 2 is also satisfied, any pair $(M, A) \in \mathbb{N} \times \mathbb{S}^{M+1}$ is such that

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

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

$$\forall (M, A) \in \mathbb{N} \times \mathbb{S}^{M+1}, \quad U_M(A) - L_M(A) \leq K \, d_1(A) + 2\, L\, \overline{R}_C(M, G) \,,$$

and therefore the gap $U_M(A) - L_M(A)$ can be made arbitrarily small under Assumption 3 by choosing a sufficiently large order M and a sufficiently small diameter for the set A. This result will be exploited systematically in the convergence analysis in Sect. 4.

Remark 12 An alternative upper bound $U_M(A)$ in (15) may be computed more directly by solving the following nonconvex optimization problem to local optimality,

$$\min_{a \in A} \; F\left(\sum_{k=0}^{M} a_k \Phi_k\right) \quad \text{s.t.} \quad \sum_{k=0}^{M} a_k \Phi_k \in C\,. \tag{18}$$

Without further assumptions on the orthogonal basis functions $\Phi_0, \Phi_1, \ldots$ and on the constraint set C, however, it is not hard to contrive examples where $P_M(x) \notin C$ for all $x \in C$ and all $M \in \mathbb{N}$; that is, contrive examples where the upper bound (18) does not converge as $M \to \infty$. This upper-bounding approach could nonetheless be combined with another bounding approach based on set arithmetics in order to prevent convergence issues; e.g., use the solution value of (18) 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 $\varepsilon > 0$; Lifting parameter $\rho > 0$
Initialization:

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

Repeat:

2. Select a set $A \in \mathscr{A}$
3. Compute upper and lower bounds, $L_M(A) \leq \inf\limits_{x \in X_M(A)} F(x) \leq U_M(A)$
4. Apply a fathoming operation
5. If the condition $\min\limits_{A \in \mathscr{A}} U_M(A) - \min\limits_{A \in \mathscr{A}} L_M(A) \leq \varepsilon$ is satisfied, stop
6. If the condition $U_M(A) - L_M(A) \leq 2(1+\rho)\Delta_M(A)$ holds for all $A \in \mathscr{A}$, 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 $\mathscr{A} = \{A_0\}$ 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 $A \in \mathscr{A}$ from the partition if

$$L_M(A) = \infty \quad \text{or} \quad \exists\, A' \in \mathscr{A} : \quad L_M(A) > U_M(A')\,.$$

– The main idea behind the lifting condition defined in Step 6 of Algorithm 1, namely

$$\forall A \in \mathscr{A}, \quad U_M(A) - L_M(A) \leq 2(1+\rho)\Delta_M(A)\,, \tag{19}$$

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 $\mathscr{A}$–in Algorithm 1, so all the subsets in $\mathscr{A}$ 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 $G \subseteq H$ be a regular set for C, and define the inverse function $\overline{R}_C^{-1}(\cdot, G) : \mathbb{R}^{++} \to \mathbb{N}$ by

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

The following result is a direct consequence of the lifting condition (19) in the branch-and-lift algorithm:

Lemma 3 *Let Assumption 3 hold, and suppose that finite bounds $L_M^0(A)$, $U_M^0(A)$ and $\Delta_M(A)$ satisfying (16)–(17) can be computed for any feasible pair $(M, A) \in \mathbb{N} \times \mathbb{S}^{M+1}$. Then, the number of lifting operations in a run of Algorithm 1 as applied to Problem (1) is at most*

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

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,

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

for a certain feasible set $A \in \mathscr{A}$. If the lifting condition (19) were to hold for A, then it would follow from (16)–(17) that

$$\varepsilon - 2\Delta_{\overline{M}}(A) \;<\; U^0_{\overline{M}}(A) - L^0_{\overline{M}}(A) \;\leq\; 2\rho\Delta_{\overline{M}}(A)\,.$$

Moreover, F being strongly Lipschitz-continuous on C by Assumption 3, we would have

$$\overline{R}_C(\overline{M}, G) \;>\; \frac{\varepsilon}{2(\rho+1)L}\,.$$

This is a contradiction, since $\overline{R}_C(\overline{M}, G) \leq \frac{\varepsilon}{2(\rho+1)L}$ 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 $\mathscr{A} = \{A_0\}$, we have

$$\operatorname{diam}(\mathscr{A}) \;:=\; \max_{A\in\mathscr{A}} \operatorname{diam}(A) \;<\; \eta\,,$$

after finitely many subdivisions, where $\operatorname{diam}(A) := \sup_{a,a'\in A} \|a - a'\|$. 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, 2 and 3 *hold, and suppose that finite bounds* $L^0_M(A)$, $U^0_M(A)$ *and* $\Delta_M(A)$ *satisfying* (16)–(17) *can be computed for any feasible pair* $(M, A) \in \mathbb{N} \times \mathbb{S}^{M+1}$. *If the partitioning scheme is exhaustive, then Algorithm* 1 *terminates after at most* $\overline{\Sigma}$ *iterations, where*

$$\overline{\Sigma} \;\leq\; \max_{0\leq M\leq\overline{M}} \Sigma\left(\frac{\varepsilon\rho}{K(\rho+1)}, M\right), \quad \textit{with}\;\; K := L \sup_{k\in\mathbb{N}} \|\Phi_k\|_H\,. \tag{20}$$

Proof By Lemma 3, 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 (19) may not be satisfied for any feasible subset $A \in \mathscr{A}$, and we have

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

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 $U_M(A) - L_M(A) \leq \varepsilon$ is satisfied if

$$U_M^0(A) - L_M^0(A) \;\leq\; \frac{\rho\varepsilon}{\rho+1}\,.$$

By Assumptions 2 and 3 and Remark 11, we have

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

and the termination condition is thus satisfied if

$$\mathrm{diam}\,(\mathscr{A}) \;\leq\; \frac{\varepsilon\rho}{K(\rho+1)}\,.$$

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

Remark 13 In the case that the sets $A \in \mathscr{A}$ are simple interval boxes and the lifting process is implemented per (14), we have

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

Therefore, one can always subdivide these boxes in such a way that the condition $\mathrm{diam}\,(\mathscr{A}) \leq \eta$ 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 (20) simplifies to

$$\overline{\Sigma} \;\leq\; \Sigma\left(\frac{\varepsilon\rho}{K(\rho+1)}, \overline{M}\right).$$

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 run-time 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 5 Consider the space of square-integrable functions $H := L_2[-\pi, \pi]$, for which it has been established in Remark 2 that any subset G^p of p-times differentiable functions with uniformly Lipschitz-continuous p-th derivatives on $[-\pi, \pi]$ is regular, with convergence rate $\overline{R}_C(M, G^p) \leq \alpha M^{-p}$ 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 13, a worst-case iteration count can be obtained as

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

Furthermore, if the global minimizer of Problem (1) happens to be a smooth ($\mathscr{C}^\infty$) function, the convergence rate can be expected to be of the form $R(M, G^\infty) = \alpha \exp(-\beta M)$, and Theorem 2 then predicts a worst-case iteration count as

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

which is much more favorable. □

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

$$\inf_{x \in L_2[0,T]} F(x) := \int_0^T \left[\left(\int_0^T f_1(t - t') x(t')\, \mathrm{d}t' \right)^2 - \left(\int_0^T f_2(t - t') x(t')\, \mathrm{d}t' \right)^2 \right] \mathrm{d}t$$
$$\text{s.t. } x \in C := \{ x \in H \mid \forall t \in [0, T],\ |x(t)| \leq 1 \} , \tag{21}$$

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 (21), as $F(x) = F(-x)$ and $x \in C$ if and only if $-x \in C$. Thus, if x^* is a global solution point of (21), 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 $L_2[0,T]$,

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

which are orthogonal by construction.

We start by showing that the functional F is strongly Lipschitz-continuous, with the bounded regular subset G in condition (7) taken as

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

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 $e \in H$, we have

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

where L is any upper bound on the term

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

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

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

It follows by Theorem 1 that

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

Combining all the bounds and substituting $T = 10$ shows that the constant $L = 611$ satisfies the condition (22).

Based on the foregoing developments and the considerations in Sect. 3.2, a simple bound $\Delta_M(A)$ on the approximation error satisfying (17) can be obtained as

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

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/) [16] or MC++ (https://github.com/omegaicl/mcpp) [35].

With regards to the computation of bounds $L_M^0(A)$ and $U_M^0(A)$ satisfying (16), we note that $F(x)$ can be interpreted as a quadratic form in x,

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

with the elements of the matrix 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 [18,39,41] to compute bounds $L_M^0(A)$ and $U_M^0(A)$ such that

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

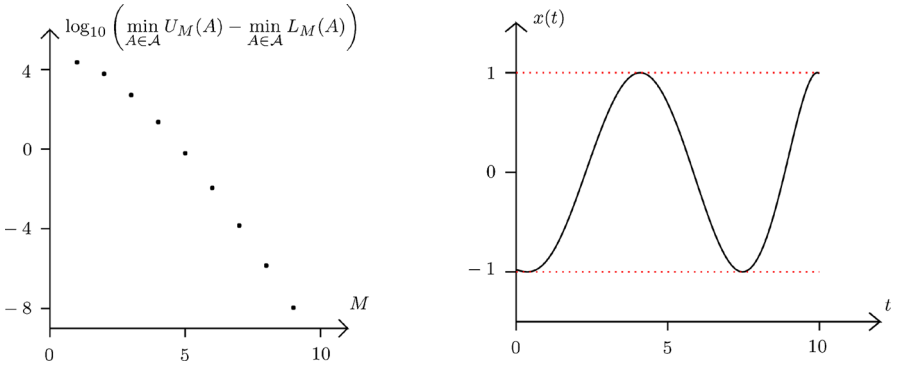


Fig. 1 Results of Algorithm 1 applied to Problem (21) for $\varepsilon = 10^{-5}$ and $\rho = 1$. Left gap between upper and lower bounds as a function of the lifted subspace dimension M. Right a globally ε-suboptimal solution x

for interval boxes $A \subseteq \mathbb{R}^{M+1}$, we use standard LMI relaxation techniques [20] here.

At this point, we have all the elements needed for implementing Algorithm 1 for Problem (21). On selecting the termination tolerance $\varepsilon = 10^{-5}$ and the lifting parameter $\rho = 1$, 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 (21) is bracketed as

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

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, provided that the cost functional is strongly Lipschitz-continuous with respect to a regular and bounded subset of that Hilbert space. The corresponding convergence conditions are satisfied for a large class of practically relevant problems in calculus of variations and optimal control. 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 with-

out 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 could 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 [25] 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. The authors would like to thank Co-Editor Dr. Sven Leyffer for his constructive comments about minimality of assumptions for the convergence of branch-and-lift.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

1. Akhiezer, N.I.: The Classical Moment Problem and Some Related Questions in Analysis. Translated by N. Kemmer. Hafner Publishing Co., New York (1965)
2. Albersmeyer, J., Diehl, M.: The lifted Newton method and its application in optimization. SIAM J. Optim. **20**(3), 1655–1684 (2010)
3. Anderson, E.J., Nash, P.: Linear Programming in Infinite-Dimensional Spaces. Wiley, Hoboken (1987)
4. Bampou, D., Kuhn, D.: Polynomial approximations for continuous linear programs. SIAM J. Optim. **22**(2), 628–648 (2012)
5. Bendsøe, M.P., Sigmund, O.: Topology Optimization: Theory, Methods, and Applications. Springer, Berlin (2004)
6. Betts, J.T.: Practical Methods for Optimal Control Using Nonlinear Programming. Advances in Design and Control Series, 2nd edn. SIAM, Philadelphia (2010)
7. Biegler, L.T.: Solution of dynamic optimization problems by successive quadratic programming and orthogonal collocation. Comput. Chem. Eng. **8**, 243–248 (1984)
8. Bock, H.G., Plitt, K.J.: A multiple shooting algorithm for direct solution of optimal control problems. In: Proceedings 9th IFAC World Congress Budapest, pp. 243–247. Pergamon Press, Oxford (1984)
9. Bompadre, A., Mitsos, A.: Convergence rate of McCormick relaxations. J. Glob. Optim. **52**(1), 1–28 (2012)
10. Bompadre, A., Mitsos, A., Chachuat, B.: Convergence analysis of Taylor and McCormick-Taylor models. J. Glob. Optim. **57**(1), 75–114 (2013)
11. Boyd, S., Vandenberghe, L.: Convex Optimization. University Press, Cambridge (2004)
12. Bryson, A.E., Ho, Y.: Applied Optimal Control. Hemisphere, Washington (1975)
13. Buie, R., Abrham, J.: Numerical solutions to continuous linear programming problems. Z. Oper. Res. **17**(3), 107–117 (1973)
14. Devolder, O., Glineur, F., Nesterov, Y.: Solving infinite-dimensional optimization problems by polynomial approximation. In: Diehl, M., Glineur, F., Jarlebring, E., Michiels, W. (eds.) Recent Advances in Optimization and its Applications in Engineering, pp. 31–40. Springer, Berlin Heidelberg (2010)
15. Ditzian, Z., Totik, V.: Moduli of Smoothness. Springer, Berlin (1987)

16. Driscoll, T.A., Hale, N., Trefethen, L.N.: Chebfun Guide. Pafnuty Publications, Oxford (2014)
17. Floudas, C.A.: Deterministic Global Optimization: Theory, Methods, and Applications. Kluwer, Dordrecht (1999)
18. Goemans, M.X., Williamson, D.P.: Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming. J. ACM **42**(6), 1115–1145 (1995)
19. Gottlieb, D., Shu, C.W.: On the Gibbs phenomenon and its resolution. SIAM Rev. **39**(4), 644–668 (1997)
20. Henrion, D., Tarbouriech, S., Arzelier, D.: LMI approximations for the radius of the intersection of ellipsoids: a survey. J. Optim. Theory Appl. **108**(1), 1–28 (2001)
21. Henrion, D., Korda, M.: Convex computation of the region of attraction of polynomial control systems. IEEE Trans. Autom. Control **59**(2), 297–312 (2014)
22. Hinze, M., Pinnau, R., Ulbrich, M., Ulbrich, S.: Optimization with PDE Constraints. Springer, Berlin (2009)
23. Horst, R., Tuy, H.: Global Optimization: Deterministic Approaches, 3rd edn. Springer, Berlin, Germany (1996)
24. Houska, B., Ferreau, H.J., Diehl, M.: ACADO toolkit–an open-source framework for automatic control and dynamic optimization. Optim. Control Appl. Methods **32**, 298–312 (2011)
25. Houska, B., Chachuat, B.: Branch-and-lift algorithm for deterministic global optimization in nonlinear optimal control. J. Optim. Theory Appl. **162**(1), 208–248 (2014)
26. Houska, B., Villanueva, M.E., Chachuat, B.: Stable set-valued integration of nonlinear dynamic systems using affine set parameterizations. SIAM J. Numer. Anal. **53**(5), 2307–2328 (2015)
27. Jackson, D.: The Theory of Approximation, vol. XI. AMS Colloquium Publication, New York (1930)
28. Katznelson, Y.: An Introduction to Harmonic Analysis, 2nd edn. Dover Publications, New York (1976)
29. Korda, M., Henrion, D., Jones, C.N.: Convex computation of the maximum controlled invariant set for polynomial control systems. SIAM J. Control Optim. **52**(5), 2944–2969 (2014)
30. Lasserre, J.B.: Moments, Positive Polynomials and Their Applications. Imperial College Press, London (2009)
31. Lin, Y., Stadtherr, M.A.: Validated solutions of initial value problems for parametric ODEs. Appl. Numer. Math. **57**(10), 1145–1162 (2007)
32. Luo, X., Bertsimas, D.: A new algorithm for state-constrained separated continuous linear programs. SIAM J. Control Optim. **37**, 177–210 (1998)
33. McCormick, G.P.: Computability of global solutions to factorable nonconvex programs: part I-convex underestimating problems. Math. Program. **10**, 147–175 (1976)
34. Misener, R., Floudas, C.A.: ANTIGONE: algorithms for continuous/integer global optimization of nonlinear equations. J. Glob. Optim. **59**(2–3), 503–526 (2014)
35. Mitsos, A., Chachuat, B., Barton, P.I.: McCormick based relaxations of algorithms. SIAM J. Optim. **20**, 573–601 (2009)
36. Moore, R.E.: Methods and Applications of Interval Analysis. SIAM Studies in Applied Mathematics. SIAM, Philadelphia (1979)
37. Mordukhovich, B.S.: Variational Analysis and Generalized Differentiation I: Basic Theory. Springer, Berlin (2006)
38. Neher, M., Jackson, K.R., Nedialkov, N.S.: On Taylor model based integration of ODEs. SIAM J. Numer. Anal. **45**, 236–262 (2007)
39. Nemirovski, A., Roos, C., Terlaky, T.: On maximization of quadratic form over intersection of ellipsoids with common center. Math. Program. **86**(3), 463–473 (1999)
40. Nesterov, Y., Nemirovskii, A.: Interior-Point Polynomial Methods in Convex Programming. SIAM, Philadelphia (1994)
41. Nesterov, Y.: Semidefinite relaxation and non-convex quadratic optimization. Optim. Methods Softw. **12**, 1–20 (1997)
42. Nesterov, Y.: Squared functional systems and optimization problems. In: Frenk, H., Roos, K., Terlaky, T., Zhang, S. (eds.) High Performance Optimization, pp. 405–440. Kluwer Academic Publishers, Dordrecht (2000)
43. Neumaier, A.: Taylor forms—use and limits. Reliab. Comput. **9**(1), 43–79 (2002)
44. Neumaier, A.: Complete search in continuous global optimization and constraint satisfaction. Acta Numer. **13**, 271–369 (2004)
45. Parrilo, P.A.: Polynomial games and sum of squares optimization. In: Proceedings of the 45th IEEE Conference on Decision & Control, pp. 2855–2860. San Diego (CA) (2006)

46. Pontryagin, L.S., Boltyanskii, V.G., Gamkrelidze, R.V., Mishchenko, E.F.: The Mathematical Theory of Optimal Processes. Wiley, New York (1962)
47. Rajyaguru, J., Villanueva, M.E., Houska, B., Chachuat, B.: Chebyshev model arithmetic for factorable functions. J. Glob. Optim. **68**(2), 413–438 (2017)
48. Saff, E.B., Totik, V.: Polynomial approximation of piecewise analytic functions. J. Lond. Math. Soc. **39**(2), 487–498 (1989)
49. Sahinidis, N.V.: A general purpose global optimization software package. J. Glob. Optim. **8**(2), 201–205 (1996)
50. Scott, J.K., Chachuat, B., Barton, P.I.: Nonlinear convex and concave relaxations for the solutions of parametric ODEs. Optim. Control Appl. Methods **34**(2), 145–163 (2013)
51. von Stryk, O., Bulirsch, R.: Direct and indirect methods for trajectory optimization. Ann. Oper. Res. **37**, 357–373 (1992)
52. Tawarmalani, M., Sahinidis, N.V.: A polyhedral branch-and-cut approach to global optimization. Math. Program. **103**(2), 225–249 (2005)
53. Villanueva, M.E., Houska, B., Chachuat, B.: Unified framework for the propagation of continuous-time enclosures for parametric nonlinear ODEs. J. Glob. Optim. **62**(3), 575–613 (2015)
54. Vinter, R.: Optimal Control. Springer, Berlin (2010)
55. Wang, H., Xiang, S.: On the convergence rates of Legendre approximation. Math. Comput. **81**(278), 861–877 (2012)