

Global Optimization on an Interval

Thomas A. Weber¹

Received: 26 January 2016 / Accepted: 25 August 2016 / Published online: 20 September 2016
© The Author(s) 2016. This article is published with open access at Springerlink.com

Abstract This paper provides expressions for solutions of a one-dimensional global optimization problem using an adjoint variable which represents the available one-sided improvements up to the interval “horizon.” Interpreting the problem in terms of optimal stopping or optimal starting, the solution characterization yields two-point boundary problems as in dynamic optimization. Results also include a procedure for computing the adjoint variable, as well as necessary and sufficient global optimality conditions.

Keywords Dynamic systems · Global optimality conditions · Optimal stopping

Mathematics Subject Classification 49K05 · 49K15 · 91B06

1 Introduction

The generic nonconcavity of maximization problems generally leads to multiple local optima. Standard optimality conditions tend to be local, and techniques for global optimization are usually algorithmic in nature, restricting the search for the best solution to subsets of the domain. For the simple case where the domain is an interval, a global maximizer of a continuously differentiable function can be found by using techniques from dynamic systems, notably by introducing global information in the form of an adjoint variable. In this manner, we construct expressions for solutions to a global

Communicated by Panos M. Pardalos.

✉ Thomas A. Weber
thomas.weber@epfl.ch

¹ École Polytechnique Fédérale de Lausanne, 1015 Lausanne, Switzerland

optimization problem on an interval, which are directly related to dynamic interpretations in terms of optimal stopping and optimal starting. In addition to providing a full characterization of solutions to a global optimization problem over an interval, the adjoint variable can also be used locally to formulate necessary and sufficient optimality conditions for one-sided subproblems of the original global optimization problem.

1.1 Literature

Following [1], global optimization methods use either deterministic search algorithms (e.g., via gradient methods) or random-sampling procedures. The first type of algorithms consists of schemes for systematic search updates. The Bolzano search finds critical points of a concave objective function via bisection (see, e.g., [2], p. 122). The golden-section search by [3] for unimodal functions increases the efficiency of the bisection method by varying the subdivision using Fibonacci numbers; see also [4].¹ Algorithms based on steepest ascent, such as Newton's method (see, e.g., [7], Ch. 9.5), tend to be greedy and therefore converge to local extrema. Improvements are achieved by using (deterministic) sampling techniques capitalizing on available knowledge about the variation of the function in terms of its Lipschitz constant [8]. The latter can be refined by locally estimating the Lipschitz constant [9], using a quadratic bound [10], or by employing a higher-order approach, e.g., considering additionally the Lipschitz constant for the variation of the gradient [11]. An overview of the second type of algorithms, based on random sampling, is provided by [12], Ch. 4. An alternative Bayesian approach, assuming a probabilistic model of the objective function as a stochastic process, was proposed by [13]. These algorithms amount to numerical techniques, predicated on the assumption that the objective function is expensive to evaluate or nonsmooth, so as to deny the possibility of direct analytical calculations. In breaking with this premise, our goal is to provide insights about the kind of information needed to compute solutions to a global optimization problem as well as their properties, rather than an attempt to improve on the numerical side.

We assume that the underlying objective function is continuously differentiable, and then reduce the solution of the global optimization problem to solving an “adjoint” differential equation. In the spirit of [14], this differential equation performs the somewhat unexpected task of aggregating global information about the available one-sided improvements. Since the adjoint equation has a discontinuous right-hand side, existence and uniqueness of the solution are obtained separately via successive Picard iterations (see, e.g., [15], p. 213), without relying on (here unavailable) Lipschitz constants.

1.2 Outline

The remainder of this paper is organized as follows. Section 2 introduces notation and basic concepts, most notably an auxiliary (adjoint) variable which represents the

¹ For further discussion of one-dimensional search methods, see [5,6].

optimal improvement up to the interval horizon. Section 3 provides expressions for the solutions of a one-dimensional global optimization problem as well as necessary and sufficient global optimality conditions. Section 4 contains several examples to illustrate the results. It also clarifies the equivalence of global optimization with optimal stopping (or starting) problems. Section 5 discusses global optimality conditions and the relationship of the proposed methods to the analysis of optimal control problems. Section 6 concludes.

2 Preliminaries

For any given $T > 0$, consider the global optimization problem²

$$F^* = \max_{t \in [0, T]} F(t), \tag{P}$$

where $F : [0, T] \rightarrow \mathbb{R}$ is a differentiable real-valued objective function with continuous derivative $f : [0, T] \rightarrow \mathbb{R}$. By the Weierstrass theorem (see, e.g., [16], p. 540), problem (P) has a solution, i.e., its solution set $\mathcal{P} \subseteq [0, T]$ is nonempty, and the optimal value F^* is finite. Furthermore, it is well known that any (interior) optimizer $\hat{t} \in]0, T[$ (i.e., excluding the boundary points 0 and T) satisfies the Fermat condition,

$$f(\hat{t}) = 0, \tag{1}$$

but that there may be many points \hat{t} that do not solve (P) but still satisfy $f(\hat{t}) = 0$. For example, if F is equal to a value $\bar{F} < F^*$ on a subinterval, then there is a continuum of such values. We are interested in characterizing the solution(s) to the global optimization problem, as element(s) of $[0, T]$, including the boundaries. For this, we introduce an auxiliary function, also referred to as “adjoint variable,” $x : [0, T] \rightarrow \mathbb{R}$ as the unique solution to the initial-value problem³

$$\dot{x}(s) = \Phi(f(T - s), x(s)), \quad x(0) = 0, \tag{2}$$

for $s \in [0, T]$, where for any $(\hat{t}, \hat{x}) \in \mathbb{R}^2$:

$$\Phi(\hat{t}, \hat{x}) := \begin{cases} \hat{t}, & \text{if } \hat{x} > 0, \\ \max\{0, \hat{t}\}, & \text{if } \hat{x} = 0, \\ 0, & \text{otherwise.} \end{cases}$$

The right-hand side of the differential equation in (2) is discontinuous and generally does not satisfy the Carathéodory conditions (see, e.g., [17], p. 3). Before we establish existence and uniqueness of a solution to the initial-value problem in the space

² The analysis remains unchanged if the domain $[0, T]$ is replaced by any interval $[a, b]$; see Remark 3.2. As usual, the function f is a one-sided derivative at the interval boundaries.

³ Throughout we use the dot-notation for total derivatives, so $\dot{F}(t) \equiv dF(t)/dt \equiv f(t)$.

$\mathcal{W}^{1,1}([0, T])$ of absolutely continuous functions on $[0, T]$ (see Theorem 2.1 below), we provide a useful lower bound.

Lemma 2.1 For any $s \in [0, T]$: $x(s) \geq \max\{0, F(T) - F(T - s)\}$.

Proof The adjoint variable $x(s)$ cannot become negative, since Eq. (2) implies that $\dot{x} \geq 0$ at the boundary of positivity, i.e., whenever $x = 0$. Thus, $x(s) \geq 0$ for all $s \in [0, T]$. We now show that $x(s) \geq F(T) - F(T - s)$. For this, note that the solution to the initial-value problem

$$\dot{z}(s) = f(T - s), \quad z(0) = 0,$$

for $s \in [0, T]$, is of the form

$$z(s) = \int_{T-s}^T f(\theta) \, d\theta = F(T) - F(T - s). \tag{3}$$

Consider the difference $\Delta := x - z$. Then, $\Delta(0) = 0$ and, using the fact that $x(s) \geq 0$, it is

$$\dot{\Delta}(s) = -\min\{0, f(T - s)\} \mathbf{1}_{\{x(s)=0\}} = \max\{0, -f(T - s)\} \mathbf{1}_{\{x(s)=0\}} \geq 0.$$

Thus,

$$\Delta(s) = \int_{T-s}^T \max\{0, -f(\theta)\} \mathbf{1}_{\{x(T-\theta)=0\}} \, d\theta \geq 0, \quad s \in [0, T], \tag{4}$$

which implies that $x(s) \geq z(s)$ for all $s \in [0, T]$. This proves the claim. □

As explained in the next section, the adjoint variable $x(s)$ measures the optimal improvement of the objective value $F(T - s)$ on the interval $[T - s, T]$. Because the comparison set includes the current value of the objective function, the improvement must be nonnegative and has to exceed the difference $F(T) - F(T - s)$, at least weakly.

By Lemma 2.1 any solution x to Eq. (2), if it exists, cannot have negative values on $[0, T]$. Moreover, for any $(\hat{t}, \hat{x}) \in \mathbb{R}^2$:⁴

$$\hat{x} \geq 0 \Rightarrow \Phi(\hat{t}, \hat{x}) = \hat{t} \mathbf{1}_{\{\hat{x}>0\}} + \max\{0, \hat{t}\} \mathbf{1}_{\{\hat{x}=0\}} = \hat{t} - \min\{0, \hat{t}\} \mathbf{1}_{\{\hat{x}\leq 0\}} =: \hat{\Phi}(\hat{t}, \hat{x}).$$

Thus, if we set $\varphi(s) := f(T - s)$ and $\varphi_-(s) := \min\{0, \varphi(s)\}$ for all $s \in [0, T]$, then based on the preceding implication, the initial-value problem in Eq. (2) can be rewritten in the form

$$\dot{x}(s) = \hat{\Phi}(\varphi(s), x(s)) = \varphi(s) - \varphi_-(s) \mathbf{1}_{\{x(s)\leq 0\}}, \quad x(0) = 0, \tag{2'}$$

⁴ For $\hat{x} < 0$, the values $\Phi(\hat{t}, \hat{x})$ and $\hat{\Phi}(\hat{t}, \hat{x})$ may be different. But by Lemma 2.1 the adjoint variable x is nonnegative, so that this case becomes irrelevant for any solution of the initial-value problem (2).

without affecting its set $\mathcal{R} \subset \mathcal{W}^{1,1}([0, T])$ of solutions. The Sobolev space $\mathcal{W}^{1,1}([0, T])$ contains all absolutely continuous real-valued functions x defined on the domain $[0, T]$ and equipped with the norm $\|\cdot\|_{1,1}$, where

$$\|x\|_{1,1} = \int_0^T (|x(s)| + |\dot{x}(s)|) \, ds. \quad (5)$$

The vector space $\mathcal{W}^{1,1}([0, T])$ is a Banach space, i.e., a complete normed vector space, which means that any Cauchy sequence with elements in the vector space converges (in the $\|\cdot\|_{1,1}$ -norm) to an element of the vector space. The solution set of the initial-value problem (2') is

$$\mathcal{R} := \{x \in \mathcal{W}^{1,1}([0, T]) : \mathbf{P}x = x\},$$

where the operator $\mathbf{P} : \mathcal{W}^{1,1}([0, T]) \rightarrow \mathcal{W}^{1,1}([0, T])$ maps any absolutely continuous function x on $[0, T]$ to a function $\mathbf{P}x$, with

$$(\mathbf{P}x)(s) := \int_0^s \hat{\Phi}(\varphi(\zeta), x(\zeta)) \, d\zeta, \quad s \in [0, T], \quad (6)$$

which (as can be verified) is also an element of $\mathcal{W}^{1,1}([0, T])$. The following result provides existence and uniqueness of a solution to the initial-value problems (2) and (2').

Theorem 2.1 $\mathcal{R} = \{x\}$, i.e., there exists a unique solution $x \in \mathcal{W}^{1,1}([0, T])$ to the initial-value problem (2), and $\mathbf{P}x = x$.

As becomes clear in the proof of the last result (provided in the Appendix), a repeated application of the operator \mathbf{P} to ϕ , where $\phi(s) := \int_0^s \varphi(\zeta) \, d\zeta$ for all $s \in [0, T]$, converges to the unique solution of Eq. (2). That is, when considering the sequence $\sigma := (x_k)_{k=0}^\infty$, with the initial function $x_0 = \phi$ and the Picard iteration $x_{k+1} = \mathbf{P}x_k$ for $k \geq 0$, then $x_k \rightarrow x \in \mathcal{R}$ as $k \rightarrow \infty$. In practice, the convergence of the sequence σ to the adjoint variable $x = \lim_{k \rightarrow \infty} \mathbf{P}^k \phi$ is usually very efficient and takes place within a few iterations; see Fig. 1 for an example.

3 Main Results

Based on the notions introduced in the proof of Lemma 2.1, it is now possible to construct expressions for the solutions of (P), first for the smallest solution t^* , then the largest solution t^{**} , and finally for all solutions in between.

Theorem 3.1 The smallest solution of (P) is

$$t^* = T - \sup\{s \in [0, T] : x(s) = 0\}.$$

Proof By Lemma 2.1 the adjoint variable $x(s) \geq 0$ for all $s \in [0, T]$, and $x(0) = 0$ by the initial condition in Eq. (2). The set $\mathcal{S} := \{s \in [0, T] : x(s) = 0\}$ is nonempty

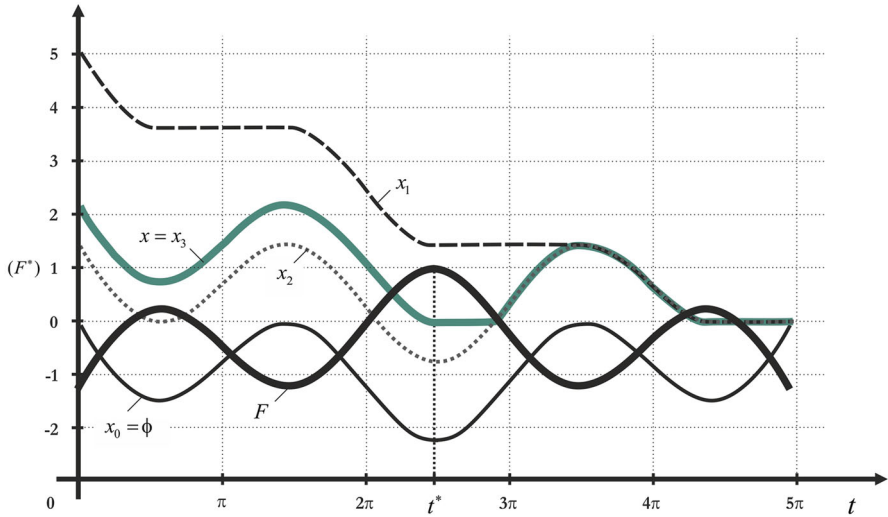


Fig. 1 Computation of x in 3 iterations, for $F(t) = \sin(t) - (t - (5\pi/2))^2/50$ on $[0, 5\pi]$

(because $0 \in \mathcal{S}$), and its supremum, $s^* := \sup \mathcal{S}$, therefore exists and lies in the interval $[0, T]$. Depending on whether or not \mathcal{S} is a singleton, we consider two cases.

Case 1: $\mathcal{S} = \{0\}$. Since $x(s) > 0$ for all $s \in]0, T]$, by Eq. (2) it is

$$x(s) = \int_0^s f(T - \vartheta) d\vartheta = \int_{T-s}^T f(\theta) d\theta > 0, \quad s \in]0, T].$$

Thus, for any $t \in [0, T[$, by setting $s = T - t$, one obtains

$$F(t) = F(T) - \int_t^T f(\theta) d\theta = F(T) - x(T - t) < F(T).$$

Since $s^* = 0$, this implies that $t^* = T - s^* = T$ solves (P).

Case 2: $\mathcal{S} \supsetneq \{0\}$. Let $\hat{s} \in]0, T]$ such that $x(\hat{s}) = 0$. Thus, $\hat{s} \in \mathcal{S}$ and $s^* \geq \hat{s} > 0$. By Eqs. (3) and (4) the difference

$$\Delta(s) = x(s) - z(s) = \int_{T-s}^T \max\{0, -f(\theta)\} \mathbf{1}_{\{x(T-\theta)=0\}} d\theta$$

is nondecreasing in s . Now consider the optimal value of the global optimization problem (P) subject to the additional constraint that $t \in [T - \hat{s}, T]$, so

$$\hat{F}^*(\hat{s}) := \max_{t \in [T-\hat{s}, T]} F(t) = \max_{t \in [T-\hat{s}, T]} \left\{ F(T) - \int_t^T f(\theta) d\theta \right\} \leq F^*. \quad (7)$$

Then by virtue of Eq. (3) and the nonnegativity of x it is

$$\begin{aligned} \hat{F}^*(\hat{s}) &= \max_{t \in [T-\hat{s}, T]} \{F(T) - z(T - t)\} \\ &= F(T) + \max_{t \in [T-\hat{s}, T]} \{\Delta(T - t) - x(T - t)\} \\ &\leq F(T) + \max_{t \in [T-\hat{s}, T]} \{\Delta(T - t)\}. \end{aligned}$$

By the monotonicity of $\Delta(s)$, alluded to earlier, the maximum on the right-hand side is achieved for $t = T - \hat{s}$. Since by assumption $x(\hat{s}) = 0$, it is $\Delta(\hat{s}) = x(\hat{s}) - z(\hat{s}) = -z(\hat{s})$. Furthermore, by Eq. (3), $-z(\hat{s}) = F(T - \hat{s}) - F(T)$, so that $\hat{F}^*(\hat{s}) \leq F(T - \hat{s})$. But the value on the right-hand side of the preceding inequality can be attained in the maximization of F over the interval $[T - \hat{s}, T]$ in Eq. (7) by choosing $t = T - \hat{s}$, which implies

$$\hat{F}^*(\hat{s}) = F(T - \hat{s}^*).$$

Using again the monotonicity of $\Delta(s)$, for any $\hat{s}' \in \mathcal{S}$ with $\hat{s}' \geq \hat{s}$, one obtains $\hat{F}^*(\hat{s}') \geq \hat{F}^*(\hat{s})$, whence

$$\hat{F}^*(\hat{s}) \leq \sup_{s \in \mathcal{S}} \hat{F}^*(s) = \hat{F}^*(s^*) = F(T - s^*).$$

We therefore know that

$$F(T - s^*) = \max_{t \in [T-s^*, T]} F(t), \tag{8}$$

and $x(s) > 0$ for all $s \in]s^*, T]$. Thus, $\hat{\mathcal{S}} := \{s \in [s^*, T] : x(s) = 0\}$ is a singleton: $\hat{\mathcal{S}} = \{s^*\}$. Analogous to Case 1, one can conclude that the maximum of F on the interval $[0, T - s^*]$ is attained at the upper end of the domain, so

$$F(T - s^*) = \max_{t \in [0, T-s^*]} F(t). \tag{9}$$

Combining Eqs. (8) and (9), the solution to the global optimization problem (P) is therefore $t^* = T - s^*$, and

$$F^* = F(T) + \Delta(T - t^*) = F(t^*),$$

which completes the proof. □

Remark 3.1 By substituting $s = T - t$ in Theorem 3.1, the smallest solution to the global optimization problem (P) can also be written in the form

$$t^* = \inf \{t \in [0, T] : x(T - t) = 0\}.$$

Accordingly, the optimal value of (P) is

$$F^* = F(t^*) = F(T) + \int_{t^*}^T \max\{0, -f(\theta)\} \mathbf{1}_{\{x(T-\theta)=0\}} d\theta.$$

In the foregoing derivations, the nonnegative adjoint variable $x(T - t)$, defined as the solution to the initial-value problem (2), measures the possible cumulative improvement of a solution in the interval $[t, T]$ relative to the current value $F(t)$. The smallest solution of (P) is the smallest t^* for which no improvement of the objective can be obtained on the interval $[t^*, T]$, so $x(T - t^*) = 0$ in particular. Alternatively, one can determine the largest solution t^{**} of (P) by measuring cumulative improvements over $F(t)$ on the interval $[0, t]$. For this, consider the unique solution to the initial-value problem

$$\dot{y}(t) = \Phi(-f(t), y(t)), \quad y(0) = 0, \tag{10}$$

for $t \in [0, T]$. Analogous to the iterative procedure for the solution of the initial-value problem (2) in Sect. 2, it is possible to obtain the (co-)adjoint variable y by successive approximation, $\lim_{k \rightarrow \infty} \hat{\mathbf{P}}^k \hat{\Phi} = y$, where the operator $\hat{\mathbf{P}} : \mathcal{W}^{1,1}([0, T]) \rightarrow \mathcal{W}^{1,1}([0, T])$ maps any absolutely continuous function y on $[0, T]$ to an absolutely continuous function $\hat{\mathbf{P}}y$, with

$$(\hat{\mathbf{P}}y)(t) := \int_0^t \hat{\Phi}(-f(\theta), y(\theta)) d\theta, \quad t \in [0, T], \tag{11}$$

just as the operator \mathbf{P} in Eq. (6), and where $\hat{\Phi}(t) := -\int_0^t f(\theta) d\theta = F(0) - F(t)$. As with Eq. (2'), corresponding to Eq. (2), there exists an equivalent formulation for the initial-value problem (10) for the computation of y ,

$$\dot{y}(t) = \hat{\Phi}(-f(t), y(t)) = -f(t) + f_+(t) \mathbf{1}_{\{y(t) \leq 0\}}, \quad y(0) = 0, \tag{10'}$$

where $f_+(t) := \max\{0, f(t)\}$ for $t \in [0, T]$.

Corollary 3.1 *The largest solution of (P) is $t^{**} = \sup\{t \in [0, T] : y(t) = 0\}$.*

Proof For any $s \in [0, T]$, let $G(s) := F(T - s)$. Then, any solution to the global optimization problem

$$G^* = \max_{s \in [0, T]} G(s) \tag{P'}$$

is also a solution of (P). Moreover, by Theorem 3.1 the smallest solution s^* of (P') is equal to T minus the largest solution t^{**} of (P). Mirroring the objective function from F to G also mirrors the corresponding derivatives from f to g , in the sense that

$$g(s) := \dot{G}(s) = -\dot{F}(T - s) = -f(T - s),$$

for all $s \in [0, T]$. A (unique) solution y to the initial-value problem (2), applied to the primitives of the mirrored global optimization problem (P') (with the independent variable s suitably replaced by t), satisfies

$$\dot{y}(t) = \Phi(g(T - t), y(t)) = \Phi(-f(t), y(t)), \quad y(0) = 0,$$

for $t \in [0, T]$. The latter corresponds to the initial-value problem (10). By Theorem 3.1, the smallest solution of (P') is $s^* = T - \sup\{t \in [0, T] : y(t) = 0\}$, so that the largest solution of (P) becomes

$$t^{**} = T - s^* = \sup\{t \in [0, T] : y(t) = 0\},$$

which concludes the proof. □

The two preceding results together characterize the uniqueness of a solution to the global optimization problem.

Corollary 3.2 *A solution of (P) is unique if and only if*

$$\sup\{s \in [0, T] : x(s) = 0\} + \sup\{t \in [0, T] : y(t) = 0\} = T.$$

Proof The result follows immediately by setting $t^* = t^{**}$ in Theorem 3.1 and Corollary 3.1. □

Intuitively, a solution t^* of (P) is unique if and only if the length of the largest interval for zero cumulative improvement (of the objective function F) to the right of t and the length of the largest interval for zero cumulative improvement to the left of t add up to the length T of the domain $[0, T]$ at $t = t^*$.

Remark 3.2 Consider the (slightly) “generalized” global optimization problem

$$H^* = \max_{\hat{t} \in [a, b]} H(\hat{t}), \tag{P''}$$

featuring a continuously differentiable real-valued objective function H , defined on the interval $[a, b]$, where a, b are any given real numbers such that $a < b$. While (P'') seems more general than (P), it can be reduced to the latter by maximizing $F(t) := H(a + t)$ on the interval $[0, T]$ (for t) with $T := b - a$, just as in the original optimization problem (P). Any solution t^* of (P) directly corresponds to a solution \hat{t}^* of (P'') via translation, $\hat{t}^* = t^* - a$.

It is possible to generalize the representation of the solutions in Theorem 3.1 and Corollary 3.1 to cases where the global optimization problem has more than 2 solutions. Indeed, if (P) has any finite number of solutions, all solutions can be found recursively.

Corollary 3.3 *If $\mathcal{P} = \{t_1, \dots, t_N\} \subset [0, T]$ (with $t^* = t_1 < \dots < t_N = t^{**}$) is a complete set of $N > 2$ distinct solutions of (P), then all solutions (between the smallest and the largest) are*

$$t_k = \check{T} - \sup\{s \in [0, \check{T} - t_{k-1}] : \check{x}(s) = 0\}, \quad k \in \{2, \dots, N - 1\}, \tag{12}$$

where \check{x} is the unique solution of the initial-value problem (2) with T replaced by $\check{T} := t^{**}$.

Proof Note first that necessarily the optimal value of (P) is such that $F^* = F(t_k)$ for all $k \in \{1, \dots, N\}$. Consider now any solution $t_k \in (t^*, t^{**})$ for $k \in \{2, \dots, N - 1\}$, obtained by the recursion in Eq. (12). Since $[0, \check{T}]$ is a subset of $[0, T]$, the point t_k also solves the “generalized” global optimization problem (P’) on the interval $[a, b] = [t_k, \check{T}]$. Moreover, by Theorem 3.1:

$$t_k = \check{T} - \sup\{s \in [0, \check{T} - t_k] : \check{x}(s) = 0\}.$$

Since $F^* = F(\check{T})$, there exists an $\varepsilon \in]0, \check{T} - t_k[$ so that the right-sided improvement $\check{x}(s)$ is strictly positive for all $s \in]\check{T} - t_k - \varepsilon, \check{T} - t_k[$. But this implies that

$$t_{k+1} = \check{T} - \sup\{s \in [0, \check{T} - t_k[: \check{x}(s) = 0\},$$

which corresponds to the recursion in (12), thus concluding the proof. □

Note that the cardinality of the solution set \mathcal{P} need not be finite. For instance, the objective function F , defined by $F(t) := 1 - (t^2 \sin(1/t))^2$ for $t > 0$, with $F(0) := 0$, is continuously differentiable, and (for $T \geq 1/\pi$) the global optimization problem (P) has the countable solution set $\mathcal{P} = \{t_1, t_2, \dots\}$, where $t_k = 1/(k\pi)$ for all $k \geq 1$. But \mathcal{P} need not even be countable: as an example, any constant objective function, $F(t) \equiv c \in \mathbb{R}$, would produce the continuum $\mathcal{P} = [0, T]$ as solution set of (P), equal to the entire domain.

Remark 3.3 Given $F^* = F(t^*) = F(t^{**})$, the solution set of (P), for any number of solutions, is $\mathcal{P} = \{t \in [t^*, t^{**}] : F(t) \geq F^*\}$, corresponding to the upper contour set of F relative to its globally optimal value F^* on $[0, T]$.

By combining the interpretations of the two adjoint variables x and y as the right-sided and left-sided gains, respectively, it is possible to construct a necessary and sufficient optimality condition to decide whether a given point solves the global optimization problem. For this, we introduce the combined (or “two-sided”) adjoint variable $\lambda(t) := \max\{x(T - t), y(t)\}$.

Theorem 3.2 A point $\hat{t} \in [0, T]$ is a solution of (P) if and only if

$$\lambda(\hat{t}) = 0. \tag{13}$$

Accordingly, the solution set is $\mathcal{P} = \{t \in [0, T] : \lambda(t) = 0\}$.

Proof Consider the set \mathcal{P} of solutions to (P), and let F^* be the optimal value of this global optimization problem.

- (i) *Necessity:* If $\hat{t} \in \mathcal{P}$, then by Remark 3.3 no improvement is possible on the interval $[\hat{t}, T]$, so $x(T - \hat{t}) = 0$ necessarily. Similarly, no improvement is possible on the interval $[0, \hat{t}]$ which implies that $y(\hat{t}) = 0$. Together with the definition of λ , this establishes Eq. (13) as a necessary optimality condition for any element of \mathcal{P} .

- (ii) *Sufficiency*: Consider a point $\hat{t} \in [0, T]$ which satisfies $\lambda(\hat{t}) = 0$. By Lemma 2.1, the adjoint variable x is nonnegative-valued, which—by symmetry—is also true for y . Hence, $x(T - \hat{t}) = y(\hat{t}) = 0$, so neither a right-sided (on $[\hat{t}, T]$) nor a left-sided (on $[0, \hat{t}]$) strict improvement over $F(\hat{t})$ is possible, which implies that $F(\hat{t}) = F^*$. Hence, \hat{t} must be an element of \mathcal{P} .

Based on (i) and (ii), Eq. (13) characterizes any solution of (P), which implies the representation of the solution set \mathcal{P} as the set of roots of $\lambda(t)$, concluding the proof. \square

At any given point t the combined adjoint variable $\lambda(t)$ can be interpreted as the best gain available on the domain $[0, T]$. This implies the following *invariance property*.

Corollary 3.4 *For any $t \in [0, T]$, it is $\lambda(t) + F(t) = F^*$.*

Combining the last result with the initial conditions in Eqs. (2) and (10) yields an expression of the optimal value of (P) as a function of the adjoint variables evaluated at the interval horizon.

Corollary 3.5 *$x(T) = \lambda(0) = F^* - F(0)$ and $y(T) = \lambda(T) = F^* - F(T)$.*

The aforementioned properties of the adjoint variables reveal an inherent *complementarity*, in the sense that the nonnegative one-sided adjoint variables x and y can only vanish together at a global optimum. In addition, because of the normalization to zero at either interval end, the sum of the one-sided adjoint variables at the boundaries must be equal to the optimal increment of the objective function: $x(T) + y(0) = F^* - F(0)$ and $x(0) + y(T) = F^* - F(T)$.

Remark 3.4 In the global optimality condition (13), one could replace λ by any non-trivial convex combination of x and y (e.g., by $\hat{\lambda} := (x + y)/2$), and Corollary 3.5 would continue to hold. However, as the *upper envelope* of all convex combinations of x and y , the combined adjoint variable $\lambda(t) = F^* - F(t)$ enjoys particular significance in terms of its interpretation as the available global gain relative to the value $F(t)$ at any point $t \in [0, T]$, as stated in Corollary 3.4.

4 Applications

The following examples illustrate the notions and results developed earlier.

Example 4.1 (Multiple Solutions) Consider a 2π -periodic objective function of the form $F(t) := \sin(t)$ on the interval $[0, T]$ for $T = (2N - 1)\pi$, where $N \geq 1$ is a given integer. Equation (2) yields the cumulative improvement of $F(T - s)$ over the interval $[T - s, T]$,

$$x(s) = (1 - \sin(s))\mathbf{1}_{\{s \geq \pi/2\}}, \quad s \in [0, T].$$

By symmetry of the objective function with respect to the midpoint ($T/2$) of the domain, the cumulative improvement of $F(t)$ over the interval $[0, t]$, i.e., the solution to Eq. (10), is

$$y(t) = (1 - \sin(t))\mathbf{1}_{\{t \geq \pi/2\}}, \quad t \in [0, T].$$

Thus, by Theorem 3.1 and Corollary 3.1 one obtains the smallest and the largest solution of (P), respectively: $t^* = T - \sup\{s \in [0, T] : \sin(s) = 1\} = \pi/2$ and $t^{**} = \sup\{t \in [0, T] : \sin(t) = 1\} = (4N - 3)(\pi/2)$. By Corollary 3.2, the solution of (P) is unique if and only if $N = 1$, since then $t^* = t^{**}$. For $N \geq 2$, there are exactly N different solutions: $t_1 = t^*$ and $t_N = t^{**}$, as well as $t_k = (4k - 3)(\pi/2)$ for $k \in \{2, \dots, N - 1\}$, as provided by Corollary 3.3.

Example 4.2 (Monopoly Pricing) A single-product monopolist faces heterogeneous consumers whose highest willingness-to-pay (WTP) for its good is normalized to $T = 1$, without loss of generality. Given a continuous probability density function $h : [0, 1] \rightarrow \mathbb{R}_+$ describing the distribution of consumers' WTP, the aggregate demand for the product at the price t is

$$D(t) = \int_t^1 h(\theta) d\theta, \quad t \in [0, 1].$$

Thus, assuming (for simplicity) zero marginal cost, the monopolist's optimal pricing problem becomes

$$\max_{t \in [0, 1]} \{tD(t)\},$$

which is of the form (P) for $F(t) = tD(t)$ and $f(t) = D(t) - t h(t)$. Fermat's necessary optimality condition (1) yields that at any positive optimal price $t^* \in]0, 1[$, the monopolist would set the marginal revenue f to zero, so $D(t^*) = t^*h(t^*)$.⁵ For a multimodal distribution h , there can be many prices that satisfy this optimality condition. Figure 2 depicts the situation for a bimodal beta-mixture $h(t) = \gamma p_{\alpha_1, \beta_1}(t) + (1 - \gamma) p_{\alpha_2, \beta_2}(t)$, where $\gamma \in [0, 1]$ and $p_{\alpha, \beta}(t) := t^{\alpha-1}(1 - t)^{\beta-1}/B(\alpha, \beta)$ for any $\alpha, \beta > 0$.⁶ In order to derive a necessary and sufficient optimality condition, we use Eqs. (2') and (10') to compute the adjoint variables x and y . Given any price $t \in [0, 1]$, it is best for the monopolist to increase the price if and only if the adjoint variable $x(1 - t) > 0$. And it is best for the monopolist to decrease the price if and only if the (co-)adjoint variable $y(t) > 0$. Hence, as stated in Theorem 3.2 the price $t = t^*$ is globally optimal if and only if $\lambda(t^*) = \max\{x(1 - t^*), y(t^*)\} = 0$; see Fig. 2. Furthermore, following Corollary 3.4 and Corollary 3.5 the combined adjoint variable $\lambda(t)$, at any price $t \in [0, 1]$, is equal to the distance of the profit $F(t)$ to its optimal value F^* .

Example 4.3 (Optimal Stopping) Suppose that at any time t , a decision maker has the option to either stick with a given utility stream $u(t)$ or to make an irreversible switch to an alternative utility stream $v(t)$, where both u and v are defined for all times $t \in [0, T]$. In addition, $t = 0$ denotes the present and $t = T > 0$ the relevant time horizon. By considering the utility increment of the default utility stream over the alternative utility stream,

⁵ The Fermat condition corresponds to the well-known monopoly pricing rule (see, e.g., [18], p. 66), which does not guarantee optimality.

⁶ In the numerical example, $(\alpha_1, \beta_1) = (20, 5)$, $(\alpha_2, \beta_2) = (5, 20)$, and $\gamma = 1/4$.

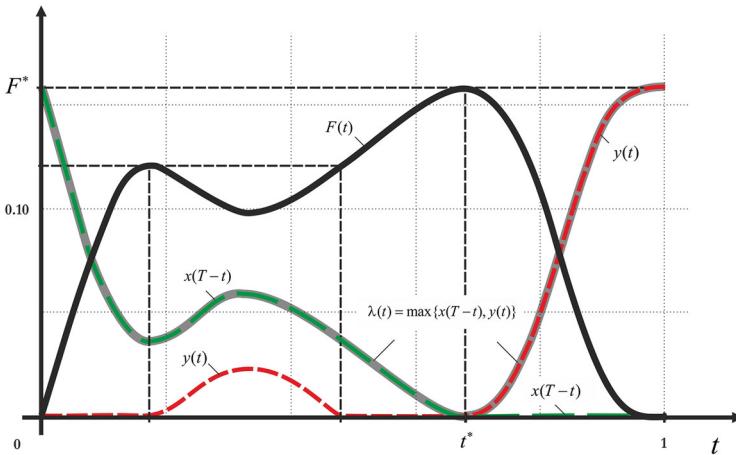


Fig. 2 Objective function $F(t)$ and cumulative one-sided gains $x(t), y(t)$ in Example 4.2

$$\delta(t) := u(t) - v(t), \quad t \in [0, T],$$

the decision maker’s optimal stopping problem can be written in the form

$$\max_{t \in [0, T]} \left\{ \int_0^t e^{-r\theta} u(\theta) d\theta + \int_t^T e^{-r\theta} v(\theta) d\theta \right\} = V_0 + \max_{t \in [0, T]} F(t),$$

where $r \geq 0$ is a given discount rate, $V_0 := \int_0^T e^{-r\theta} v(\theta) d\theta$ is a constant, and

$$F(t) := \int_0^t e^{-r\theta} \delta(\theta) d\theta, \quad t \in [0, T],$$

is the relevant objective function in the global optimization problem (P). Since $F(0) = 0$, the optimal utility increment F^* over the discounted utility V_0 of selecting the outside option immediately must be nonnegative. For all s in the interval $[0, T]$, Eq. (2) with $f(T - s) = e^{-r(T-s)}\delta(T - s)$ yields the incremental utility of following the optimal stopping strategy on the interval $[T - s, T]$, expressed by the adjoint variable $x(s)$. Moreover, the best stopping strategy, once having arrived at t (possibly *suboptimally*, by sticking to the default option), is to stop if and only if $x(T - t) = 0$. Hence, the earliest stopping time t^* must be globally optimal, and $t^* = \inf\{t \in [0, T] : x(T - t) = 0\}$ as already noted in Remark 3.1.

Remark 4.1 The foregoing example shows that a (deterministic) optimal stopping problem can be written in the form (P). The converse also holds: (P) can be interpreted as an optimal stopping problem, given the utility increment $f(t) \equiv \dot{F}(t)$ and a zero

discount rate. Theorem 3.1 addresses this interpretation. By switching the reference point, in the sense that

$$\max_{t \in [0, T]} \left\{ \int_0^t e^{-r\theta} u(\theta) \, d\theta + \int_t^T e^{-r\theta} v(\theta) \, d\theta \right\} = U_0 + \max_{t \in [0, T]} \hat{F}(t),$$

where $U_0 := \int_0^T e^{-r\theta} u(\theta) \, d\theta$ is a constant, the modified objective function

$$\hat{F}(t) := - \int_t^T e^{-r\theta} \delta(\theta) \, d\theta, \quad t \in [0, T],$$

is a translation of the original objective function: $\hat{F}(t) \equiv F(t) + (U_0 - V_0)$. Hence, one can think of (P) as an optimal starting problem. Corollary 3.1 and the cumulative left-sided benefit $y(t)$ in Eq. (10) highlight this interpretation.

5 Perspectives

The representation of solutions to the global optimization problem (P) in Sect. 3 suggests several global optimality conditions and a dynamic-systems interpretation.

5.1 Global Optimality Conditions

Consider the solution x to the initial-value problem (2) and, respectively, the solution y to the initial-value problem (10). The significance of the adjoint variables x and y as the cumulative one-sided gains of the objective value implies several global optimality conditions, cumulating in an exact characterization of solutions to (P).

- (i) A *necessary* optimality condition for *any* solution t^* of the global optimization problem (P) is that $x(T - t^*) = 0$ (resp., $y(t^*) = 0$).
- (ii) The fact that $x(T - \hat{t}) = 0$ for a given point $\hat{t} \in [0, T]$ is a *sufficient* condition for the existence of a solution to (P) in $[0, \hat{t}]$ (resp., if $y(\hat{t}) = 0$, then (P) has a solution on $[\hat{t}, T]$).
- (iii) For local maxima which are not solutions of (P), the condition $x(T - \hat{t}) = 0$ holds *if and only if* \hat{t} globally maximizes F on $[\hat{t}, T]$ (resp., $y(\hat{t}) = 0$ *if and only if* \hat{t} globally maximizes F on $[0, \hat{t}]$).
- (iv) By Theorem 3.1 (resp., Corollary 3.1), the smallest (resp., largest) solution to (P) is $t^* = T - \sup\{s \in [0, T] : x(s) = 0\}$ (resp., $t^{**} = \sup\{t \in [0, T] : y(t) = 0\}$). Additional solutions can be found using Corollary 3.3, as well as Remark 3.3.
- (v) By Theorem 3.2, a point \hat{t} solves (P) *if and only if* $\lambda(\hat{t}) = 0$, using the “combined” adjoint variable $\lambda(t) \equiv \max\{x(T - t), y(t)\}$. This condition, which can be checked pointwise, effectively supersedes the local necessary optimality condition (1) by Fermat. Furthermore, by Corollary 3.4 one obtains $\lambda(t) \equiv F^* - F(t)$. Applied to the interval boundaries, this invariance property implies that the distance to the optimal value is attained by the appropriate one-sided adjoint variable at each endpoint; see Corollary 3.5 for details.

Statements (i)–(v) also apply to points and solutions at the boundaries of the interval $[0, T]$, i.e., they are not limited to interior points, unlike standard (local) first-order optimality conditions such as (1). In particular, statement (v) provides a crisp representation of the solution set: $\mathcal{P} = \{t \in [0, T] : \lambda(t) = 0\}$.

Remark 5.1 As noted after Theorem 2.1, in practice the adjoint variable x representing the right-sided gain can be efficiently computed by repeatedly applying the operator \mathbf{P} in Eq. (6) a (usually small) number of times to ϕ , where $\phi(s) = \int_0^s f(T - \zeta) \, d\zeta = G(0) - G(s)$ for all $s \in [0, T]$, as illustrated in Fig. 1. That is, $x = \lim_{k \rightarrow \infty} \mathbf{P}^k \phi$.⁷ Similarly, the adjoint variable y representing the left-sided gain can be obtained using the operator $\hat{\mathbf{P}}$ in Eq. (11), so $\lim_{k \rightarrow \infty} \hat{\mathbf{P}}^k \hat{\phi} = y$, where $\hat{\phi}(t) = -\int_0^t f(\theta) \, d\theta = F(0) - F(t)$, for all $t \in [0, T]$.⁸

5.2 Dynamic-Systems Interpretation

The equivalence of global optimization on an interval and optimal stopping (see Remark 4.1) suggests a dynamic-systems interpretation of the solution method proposed in Sect. 3. By introducing the state variable $\xi(t)$ and the adjoint variable (“co-state”) $\psi(t) \equiv x(T - t)$, the solution of (P), given in Theorem 3.1, satisfies the following two-point boundary-value problem for $t \in [0, T]$:

$$\dot{\xi}(t) = \mu(\psi(t)), \quad \xi(0) = 0, \tag{14}$$

$$\dot{\psi}(t) = -\Phi(f(t), \psi(t)), \quad \psi(T) = 0, \tag{15}$$

where the function $\mu : \mathbb{R} \rightarrow \mathbb{R}$ in Eq. (14) implements the (optimal) stopping policy using a co-state feedback: $\mu(\hat{\psi}) := \mathbf{1}_{\{\hat{\psi} > 0\}}$, for all $\hat{\psi} \in \mathbb{R}$. The state $\xi(t)$ partitions the domain $[0, T]$ into a *continuation region* $[0, t^*]$ (where $\xi(t) = 0$) and a *stopping region* $(t^*, T]$ (where $\xi(t) > 0$). The co-state $\psi(t)$, independently determined by Eq. (15), is nonnegative and provides global information about possible improvements by continuing a search for the optimum to the right of the current t . Given the solution $(\xi, \psi)(t)$ of Eqs. (14)–(15) for $t \in [0, T]$, the current value $v(t)$ solves the initial-value problem

$$\dot{v}(t) = \mathbf{1}_{\{\xi(t) \leq 0\}} f(t), \quad v(0) = F(0),$$

for $t \in [0, T]$, so that

$$v(t) = \begin{cases} F(t), & \text{if } t \leq t^*, \\ F^*, & \text{otherwise,} \end{cases}$$

⁷ While our proofs make use of the fact that the objective function F in (P) is continuously differentiable, the iteration method for x and the optimality conditions work numerically if F is merely continuous (and f is approximated by means of differences), as long as the discretization steps are fine enough.

⁸ The starting functions ϕ and $\hat{\phi}$ are lower bounds for the respective adjoint variables. The first iterates ($\mathbf{P}\phi$ and $\hat{\mathbf{P}}\hat{\phi}$) are upper bounds; see Lemma A.1 for details.

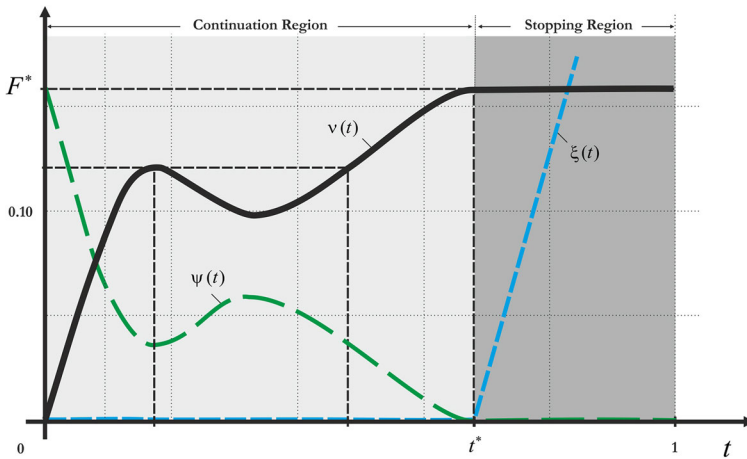


Fig. 3 State $\xi(t)$, co-state $\psi(t)$, and current value $v(t)$ in Example 4.2

where $F^* = v(T)$ is the optimal value of (P) and t^* is the (smallest) solution of (P); see Fig. 3 for an illustration using the primitives of Example 4.2. This formalizes the heuristic that it is globally optimal to walk the ‘mountain range’ defined by $F(t)$, starting at $t = 0$, toward the right, until the view toward the right becomes unimpeded. The global information about the function values not yet experienced during the walk is contributed by the co-state variable ψ . Alternately, it is possible to start walking on the interval at $t = T$ toward the left, leading to an analogous solution, as formulated in Corollary 3.1. While the results by themselves do not offer a ‘magic potion’ for finding a solution to a global optimization problem without checking the entire interval, they shed light on the importance of global information, unlike the local optimality conditions, such as (1), usually employed to identify candidates for interior local optima. The two-point boundary problem (14)–(15) is reminiscent of the Hamiltonian system which leads to a similar two-point boundary-value problem as part of the Pontryagin maximum principle [19]; see also [20].⁹ As Bellman’s principle of optimality ([21], Ch. III.3) would suggest, the adjoint variable provides in fact a solution to an entire family of nested optimization problems. It thus gives a “complete contingent plan,” in the sense that if for some reason a global optimum t^* was missed when walking from left to right, then for any $t \in]t^*, T[$ the adjoint variable still provides an optimal stopping rule on the interval $[t, T]$.

6 Conclusions

Keeping track of one-sided improvements on an interval $[0, T]$ in the form of adjoint variables $x(T - t)$ and $y(t)$, for all $t \in [0, T]$, allows for a characterization of all solutions to the global optimization problem (P). The two-sided adjoint variable $\lambda(t) = \max\{x(T - t), y(t)\}$, as the upper envelope of both one-sided adjoint

⁹ Endpoint transversality, $\psi(T) = 0$, also holds at a global optimum: $\psi(t^*) = 0$.

variables, vanishes at a point \hat{t} of the interval if and only if that point is a solution of (P), so $\hat{t} \in \mathcal{P}$. The adjoint variables are uniquely determined as solutions to the initial-value problems (2) and (10), and they can be obtained using a Picard iteration that usually terminates in a finite number of steps. Conceptually, the adjoint variables incorporate not only all the global information needed for solving (P) but also for solving subproblems of (P): A one-sided adjoint variable, say $y(t)$, describes a (‘stopping’) policy for optimizing on a subinterval $[0, t]$ from the current point t to the corresponding endpoint of the interval (0 for the left-sided adjoint variable y); $y(t) = 0$ if and only if t is a global maximum on $[0, t]$. Finally, an analytical description of all solutions to the global optimization problem (P) may be used to check solution properties, such as the monotonicity in problem parameters, that may or may not be satisfied at points implied by imprecise optimality conditions such as Eq. (1).

Acknowledgements The author would like to thank participants of the 14th EUROPT Workshop on Advances in Continuous Optimization in Warsaw, Poland, and the MODU 2016 Workshop in Melbourne, Australia, as well as several anonymous referees for helpful comments and suggestions.

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.

Appendix

Proof of Theorem 2.1 We first show existence and then uniqueness of a solution to the initial-value problem (2).

(i) *Existence:* $\mathcal{R} \neq \emptyset$. Consider a sequence of absolutely continuous functions, $\sigma := (x_k)_{k=0}^\infty \subset \mathcal{W}^{1,1}([0, T])$, defined by the recursion¹⁰

$$x_0(s) := \phi(s), \quad x_{k+1}(s) := (\mathbf{P}x_k)(s), \quad s \in [0, T],$$

for all $k \geq 0$, where $\phi(s) = \int_0^s \varphi(\zeta) \, d\zeta = F(T) - F(T - s) = z(s)$ is the difference between the boundary value and the current value of the objective function.¹¹ Consider now the sequence of the largest possible horizons s_k such that the consecutive elements of this sequence coincide, $x_k(s) = x_{k-1}(s)$, for all $s \in [0, s_k]$:

$$s_k := \sup\{s \in [0, T] : x_k(\zeta) = x_{k-1}(\zeta), \zeta \in [0, s]\}, \quad k \geq 1, \tag{16}$$

¹⁰ The idea of Picard iterations of this type dates back to Picard [22] and Lindelöf [23]. It is commonly employed for proving the existence of solutions to ordinary differential equations (see, e.g., Coddington and Levinson [24]). In this case however, the Banach fixed-point theorem cannot be used, as no Lipschitz constant for the usual contraction mapping is available—because of the discontinuous system function Φ in Eq. (2).

¹¹ A better seed for the Picard iteration is $\phi = z_+ := \max\{0, z\}$, corresponding to the lower bound for x in Lemma 2.1.

with the additional definition $s_0 := 0$. We now show the following statement:

$$\mathcal{A}(k) : s_k < T \Rightarrow s_k < s_{k+1} \leq T,$$

for all $k \geq 1$. For this, note first that $x_1 = \mathbf{P}x_0 = \mathbf{P}\phi$, with

$$x_1(s) = \phi(s) + \int_0^s (-\varphi_-(s)) \mathbf{1}_{\{\phi(s) \leq 0\}} d\zeta \geq \phi(s) = x_0(s), \quad s \in [0, T], \quad (17)$$

so $0 \leq s_1 = \inf\{s \in [0, T] : \phi(s) \leq 0\}$. Since by definition $\phi(0) = 0$, the preceding infimum is nonnegative, and by Eq. (17) it describes $s_1 \in [0, T]$ as introduced in Eq. (16). By a contradiction argument, it is straightforward to see that $s_1 > 0$. Indeed, if $s_1 = 0$, then $\phi(s) > 0$ for all $s \in (0, T]$. Thus, by the continuity of φ there exists an $\varepsilon_0 \in (0, T]$ such that $\varphi(s) > 0$ for all $s \in (0, \varepsilon_0)$. This implies $\varphi_-(s) = 0$ and by Eq. (17) therefore $x_1(s) = x_0(s)$ on $[0, \varepsilon_0]$, whence by Eq. (16): $s_1 \geq \varepsilon_0 > 0$, as claimed. If $s_1 = T$, then $\mathcal{A}(1)$ holds automatically. Consider now the interesting case where $0 < s_1 < T$. By the definition of s_1 , there exists an $\varepsilon_1 \in (0, T - s_1)$ such that for all $s \in (s_1, s_1 + \varepsilon_1)$: $\phi(s) < 0 = x_1(s)$, whence $\mathbf{1}_{\{\phi(s) \leq 0 < x_1(s)\}} = 0$. With this, the inequality in (17) yields

$$\begin{aligned} x_2(s) &= x_1(s) + \int_0^s (-\varphi_-(\zeta)) (\mathbf{1}_{\{x_1(\zeta) \leq 0\}} - \mathbf{1}_{\{\phi(\zeta) \leq 0\}}) d\zeta \\ &= x_1(s) - \int_0^s (-\varphi_-(\zeta)) \mathbf{1}_{\{\phi(\zeta) \leq 0 < x_1(\zeta)\}} d\zeta \\ &= x_1(s) - \int_{s_1 + \varepsilon_1}^{\max\{s, s_1 + \varepsilon_1\}} (-\varphi_-(\zeta)) \mathbf{1}_{\{\phi(\zeta) \leq 0 < x_1(\zeta)\}} d\zeta, \end{aligned} \quad (18)$$

for all $s \in [0, T]$. This means that $x_1(s) = x_2(s)$ for all $s \in [0, s_1 + \varepsilon_1]$, so necessarily

$$s_2 \geq s_1 + \varepsilon_1 > s_1.$$

Thus, the statement $\mathcal{A}(1)$ is true. The following auxiliary result establishes an important monotonicity property for the sequence σ , useful in the sequel of the proof.

Lemma A.1 *The even and odd subsequences $(x_{2j})_{j=0}^\infty$ and $(x_{2j+1})_{j=0}^\infty$ of σ are both monotonic, and its elements are such that $x_{2j} \leq x_{2j+2} \leq x_{2j+3} \leq x_{2j+1}$, for all $j \geq 0$.*

Proof All claims are implied by the validity of the statement

$$\mathcal{B}(j) : x_{2j} \leq x_{2j+2} \leq x_{2j+3} \leq x_{2j+1},$$

for $j \geq 0$. To show that the statement $\mathcal{B}(j)$ holds for any nonnegative integer j , we use mathematical induction (see, e.g., [25]). The inequality in (17) is equivalent to $x_0 \leq x_1$, while Eq. (18) immediately yields $x_2 \leq x_1$. Using the telescopic sum

$x_2 - x_0 = (x_2 - x_1) + (x_1 - x_0)$, Eqs. (17) and (18) together give that $x_0 \leq x_2$. Analogously, we obtain

$$x_3(s) - x_2(s) = \int_{s_2}^{\max\{s_2, s\}} (-\varphi_-(\zeta)) \mathbf{1}_{\{x_2(\zeta) \leq 0 < x_1(\zeta)\}} d\zeta \geq 0, \quad s \in [0, T],$$

i.e., $x_3 \geq x_2$. Using the statement $\mathcal{A}(1)$ and substituting the already computed differences into the telescopic sum $x_3 - x_1 = (x_3 - x_2) + (x_2 - x_1)$ yields $x_3 \leq x_1$. We have therefore established the validity of the induction hypothesis:

$$\mathcal{B}(0) : x_0 \leq x_2 \leq x_3 \leq x_1.$$

In the ‘induction step,’ we now show that if $\mathcal{B}(j)$ holds for some $j \geq 0$, then $\mathcal{B}(j + 1)$ must also be true. By virtue of $\mathcal{B}(j)$, the forward difference between two consecutive elements of σ , starting with x_{2j+3} , is

$$\begin{aligned} x_{2j+4}(s) - x_{2j+3}(s) &= \int_0^s (-\varphi_-(\zeta)) (\mathbf{1}_{\{x_{2j+3}(\zeta) \leq 0\}} - \mathbf{1}_{\{x_{2j+2}(\zeta) \leq 0\}}) d\zeta \\ &= - \int_0^s (-\varphi_-(\zeta)) \mathbf{1}_{\{x_{2j+2}(\zeta) \leq 0 < x_{2j+3}(\zeta)\}} d\zeta \\ &\leq 0, \end{aligned} \tag{19}$$

for all $s \in [0, T]$. Based on this, the forward difference between two consecutive elements of σ , starting with x_{2j+4} , is

$$\begin{aligned} x_{2j+5}(s) - x_{2j+4}(s) &= \int_0^s (-\varphi_-(\zeta)) (\mathbf{1}_{\{x_{2j+4}(\zeta) \leq 0\}} - \mathbf{1}_{\{x_{2j+3}(\zeta) \leq 0\}}) d\zeta \\ &= \int_0^s (-\varphi_-(\zeta)) \mathbf{1}_{\{x_{2j+2}(\zeta) \leq 0 < x_{2j+1}(\zeta)\}} d\zeta \\ &\geq 0, \end{aligned} \tag{20}$$

for all $s \in [0, T]$. The second inequality in

$$\mathcal{B}(j + 1) : x_{2j+2} \leq x_{2j+4} \leq x_{2j+5} \leq x_{2j+3},$$

corresponds to the inequality in (20). To establish the validity of $\mathcal{B}(j + 1)$, it therefore remains to be shown that $x_{2j+2} \leq x_{2j+4}$ and $x_{2j+5} \leq x_{2j+3}$. Consider the first of these two inequalities. Using the telescopic-sum idea, $x_{2j+4} - x_{2j+2} = (x_{2j+4} - x_{2j+3}) + (x_{2j+3} - x_{2j+2})$, together with Eq. (19) and $\mathcal{B}(j)$, one obtains

$$\begin{aligned} x_{2j+4}(s) - x_{2j+2}(s) &= - \int_0^s \varphi_-(s) [\mathbf{1}_{\{x_{2j+2}(\zeta) \leq 0 < x_{2j+1}(\zeta)\}} - \mathbf{1}_{\{x_{2j+2}(\zeta) \leq 0 < x_{2j+3}(\zeta)\}}] d\zeta \\ &= \int_0^s (-\varphi_-(s)) [\mathbf{1}_{\{0 < x_{2j+1}(\zeta)\}} - \mathbf{1}_{\{0 < x_{2j+3}(\zeta)\}}] \mathbf{1}_{\{x_{2j+2}(\zeta) \leq 0\}} d\zeta. \end{aligned}$$

By $\mathcal{B}(j)$ it is $x_{2j+3} \leq x_{2j+1}$, so that

$$\mathbf{1}_{\{0 < x_{2j+1}(\varsigma)\}} - \mathbf{1}_{\{0 < x_{2j+3}(\varsigma)\}} = \mathbf{1}_{\{x_{2j+3}(\varsigma) \leq 0 < x_{2j+1}(\varsigma)\}} \geq 0, \quad s \in [0, T],$$

which in turn implies that $x_{2j+4} \geq x_{2j+2}$. The demonstration that $x_{2j+5} \leq x_{2j+3}$ proceeds analogously and is therefore omitted; this concludes the proof of Lemma A.1. \square

By Eqs. (17) and (18), it is $\phi = x_0 \leq x_2 \leq x_1$. By virtue of Lemma A.1, if $x_k = x_{k+1}$ (i.e., $s_{k+1} = T$), then $x_k = x_{k+n}$ (i.e., $s_{k+n} = T$) for all $n \geq 1$. In our proof of $\mathcal{A}(k)$ for $k \geq 1$ we therefore consider the nontrivial case where $s_k < T$.

As in Eq. (20), the forward difference between two consecutive elements of σ , starting with an *even* element $x_k = x_{2j+2}$, is

$$x_{2j+3}(s) - x_{2j+2}(s) = \int_0^s (-\varphi_-(\varsigma)) \mathbf{1}_{\{x_{2j+2}(\varsigma) \leq 0 < x_{2j+1}(\varsigma)\}} d\varsigma,$$

for all $s \in [0, T]$ and any integer $j \geq 0$. By the definition of s_k in Eq. (16) this yields

$$x_{k+1}(s) - x_k(s) = \int_{s_k}^{\max\{s_k, s\}} (-\varphi_-(\varsigma)) \mathbf{1}_{\{x_k(\varsigma) \leq 0 < x_{k-1}(\varsigma)\}} d\varsigma, \quad s \in [0, T].$$

Since $x_k(s_k) = x_{k-1}(s_{k-1})$, by the continuity of φ there exists an $\varepsilon_k \in]0, T - s_k]$ such that $x_k(s) > x_{k-1}(s)$ for all $s \in]s_k, s_k + \varepsilon_k[$. But then $\mathbf{1}_{\{x_k(\varsigma) \leq 0 < x_{k-1}(\varsigma)\}} = 0$ on $]s_k, s_k + \varepsilon_k[$, which (by continuity) implies that $x_{k+1}(s) = x_k(s)$ for all $s \in [s_k, s_k + \varepsilon_k]$, whence (given that $s_1 > 0$, as shown earlier):

$$s_{k+1} \geq s_k + \varepsilon_k > s_k, \quad k = 2j, \quad j \geq 0. \tag{21}$$

Similarly, as in Eq. (19), the forward difference between two consecutive elements of σ , starting with an *odd* element $x_k = x_{2j+1}$, is

$$x_{2j+2}(s) - x_{2j+1}(s) = - \int_0^s (-\varphi_-(\varsigma)) \mathbf{1}_{\{x_{2j}(\varsigma) \leq 0 < x_{2j+1}(\varsigma)\}} d\varsigma,$$

for all $s \in [0, T]$ and any integer $j \geq 0$. As a result, using again the definition of s_k :

$$x_{k+1}(s) - x_k(s) = - \int_{s_k}^{\max\{s_k, s\}} (-\varphi_-(\varsigma)) \mathbf{1}_{\{x_{k-1}(\varsigma) \leq 0 < x_k(\varsigma)\}} d\varsigma, \quad s \in [0, T].$$

The fact that $x_k(s_k) = x_{k-1}(s_k)$ implies (by continuity) that there exists an ε_k in the interval $]0, T - s_k]$ such that $x_k(s) < x_{k-1}(s)$ and therefore also $\mathbf{1}_{\{x_{k-1}(\varsigma) \leq 0 < x_k(\varsigma)\}} = 0$ for all $s \in]s_k, s_k + \varepsilon_k[$. Hence, $x_{k+1}(s) = x_k(s)$ on $[s_k, s_k + \varepsilon_k]$, resulting in

$$s_{k+1} \geq s_k + \varepsilon_k > s_k, \quad k = 2j + 1, \quad j \geq 0. \tag{22}$$

Combining the monotonicity of s_k in (21) and (22), $(s_k)_{k=0}^\infty$ is an increasing sequence with upper bound T . As such it must converge ([26], p. 55), and since T is the *smallest* upper bound:¹²

$$\lim_{k \rightarrow \infty} s_k = T.$$

Employing the $\|\cdot\|_{1,1}$ -norm in Eq. (5) we can therefore conclude

$$\begin{aligned} \|x_{k+1} - x_k\|_{1,1} &\leq \int_0^T \left(\int_{s_{2k}}^T (-\varphi_-(\zeta)) d\zeta + (-\varphi_-(\zeta)) \mathbf{1}_{\{x_{2k}(s) \neq x_{2k-1}(s)\}} \right) ds \\ &\leq m(T + 1)(T - s_{2k}) \rightarrow 0, \end{aligned}$$

as $k \rightarrow \infty$, where

$$0 \leq m := \max_{t \in [0, T]} \{-\varphi_-(s)\} = \max \left\{ 0, - \min_{s \in [0, T]} \varphi(s) \right\} < \infty.$$

This in turn implies that $\sigma = (x_k)_{k=0}^\infty$ must be a Cauchy sequence. Thus, by completeness of the Banach space $\mathcal{W}^{1,1}([0, T])$, there exists an absolutely continuous function $x \in \mathcal{W}^{1,1}([0, T])$ such that $\lim_{k \rightarrow \infty} x_k = x$. The limit function x satisfies

$$0 = \|\mathbf{P}x - x\|_{1,1},$$

so

$$x(s) = \int_0^s (\varphi(\zeta) - \varphi_-(\zeta) \mathbf{1}_{\{x(\zeta) \leq 0\}}) d\zeta = \int_0^s \hat{\Phi}(\varphi(\zeta), x_k(\zeta)) d\zeta, \quad s \in [0, T],$$

which means that x solves the initial-value problem (2'), and $\mathcal{R} \neq \emptyset$.

(ii) *Uniqueness:* $x^1, x^2 \in \mathcal{R} \Rightarrow x^1 = x^2$. Indeed, for any given solutions x^1 and x^2 , consider the pointwise difference,

$$\rho(s) := x^1(s) - x^2(s), \quad s \in [0, T].$$

By the initial condition in Eq. (2') it is $\rho(0) = 0$, and

$$\dot{\rho}(s) = \dot{x}^1(s) - \dot{x}^2(s) = -\varphi_-(s) (\mathbf{1}_{\{x^1(s) \leq 0\}} - \mathbf{1}_{\{x^2(s) \leq 0\}}), \quad s \in [0, T].$$

Thus, $\dot{\rho}(s) = 0$ whenever the values $x^1(s)$ and $x^2(s)$ are *either* both positive *or* both equal to 0. On the other hand, if $x^1(s) > x^2(s) = 0$, then $\dot{\rho}(s) = \varphi_-(s) \leq 0$; and if $x^1(s) = 0 < x^2(s)$, then $\dot{\rho}(s) = -\varphi_-(s) \geq 0$. Combining these insights yields

$$\frac{d}{ds} \frac{(\rho(s))^2}{2} = (\rho(s)) \dot{\rho}(s) \leq 0, \quad s \in [0, T]. \tag{23}$$

¹² If there were another bound $\hat{t} < T$, then whenever $s_k = \hat{t}$, by virtue of $\mathcal{A}(k)$ one would obtain $s_{k+1} > \hat{t}$, i.e., a contradiction.

Together with the initial condition $\rho(0) = 0$, Eq. (23) implies

$$x^1(s) - x^2(s) = \rho(s) = 0, \quad s \in [0, T],$$

so $x^1 = x^2$, as posited at the outset of the argument.

The claims (i) and (ii) together imply that $|\mathcal{R}| = 1$, i.e., there exists a unique solution to the initial-value problem (2'), which by construction has the same solution set \mathcal{R} as the initial-value problem (2), thus concluding our proof. \square

References

- Spang III, H.A.: A review of minimization techniques for nonlinear functions. *SIAM Rev.* **4**(4), 343–365 (1962)
- Zangwill, W.I.: *Nonlinear Programming: A Unified Approach*. Prentice-Hall, Englewood Cliffs (1969)
- Kiefer, J.: Sequential minimax search for a maximum. *Proc. Am. Math. Soc.* **4**(3), 502–506 (1953)
- Kiefer, J.: Optimum sequential search and approximation methods under minimum regularity assumptions. *J. Soc. Ind. Appl. Math.* **5**(3), 105–136 (1957)
- Wilde, D.J.: *Optimum Seeking Methods*. Prentice-Hall, Englewood Cliffs (1964)
- Wilde, D.J., Beightler, C.S.: *Foundations of Optimization*. Prentice-Hall, Englewood Cliffs (1967)
- Boyd, S., Vandenberghe, L.: *Convex Optimization*. Cambridge University Press, Cambridge (2004)
- Shubert, B.O.: A sequential method seeking the global maximum of a function. *SIAM J. Numer. Anal.* **9**(3), 379–388 (1972)
- Sergeyev, YaD: A one-dimensional deterministic global minimization algorithm. *Comput. Math. Math. Phys.* **35**(5), 553–562 (1995)
- Breiman, L., Cutler, A.: A deterministic algorithm for global optimization. *Math. Program.* **58**(2), 179–199 (1993)
- Lera, D., Sergeyev, YaD: Acceleration of univariate global optimization algorithms working with Lipschitz functions and Lipschitz first derivatives. *SIAM J. Optim.* **23**(1), 508–529 (2013)
- Törn, A., Žilinskas, A.: *Global Optimization*. Lecture Notes in Computer Science, vol. 350. Springer, New York (1989)
- Locatelli, M.: Bayesian algorithms for one-dimensional global optimization. *J. Glob. Optim.* **10**(1), 57–76 (2007)
- Brockett, W.R.: Dynamical systems that sort lists, diagonalize matrices and solve linear equations. In: *IEEE Conference on Decision and Control*, vol. 1, pp. 799–803 (1988)
- Arnold, V.I.: *Ordinary Differential Equations*. MIT Press, Cambridge (1973)
- Bertsekas, D.P.: *Nonlinear Programming*. Athena Scientific, Belmont (1995)
- Filippov, A.F.: *Differential Equations with Discontinuous Righthand Sides*. Kluwer, Dordrecht (1988)
- Tirole, J.: *The Theory of Industrial Organization*. MIT Press, Cambridge (1988)
- Pontryagin, L.S., Boltyanskii, V.G., Gamkrelidze, R.V., Mishchenko, E.F.: *The Mathematical Theory of Optimal Processes*. Wiley Interscience, New York (1962)
- Weber, T.A.: *Optimal Control Theory with Applications in Economics*. MIT Press, Cambridge (2011)
- Bellman, R.E.: *Dynamic Programming*. Princeton University Press, Princeton (1957)
- Picard, E.: Sur l'application des méthodes d'approximations successives à l'étude de certaines équations différentielles ordinaires. *J. Math. Pure Appl.* **9**, 217–272 (1893)
- Lindelöf, E.: Sur l'application de la méthode des approximations successives aux équations différentielles ordinaires du premier ordre. *Cr. Hebd. Acad. Sci.* **116**, 454–457 (1894)
- Coddington, E.A., Levinson, N.: *Theory of Ordinary Differential Equations*. McGraw-Hill, New York (1955)
- Polya, G.: *Induction and Analogy in Mathematics*. Oxford University Press, London (1954)
- Rudin, W.: *Principles of Mathematical Analysis*, 3rd edn. McGraw-Hill, New York (1976)