Received: Added at production | Revised: Added at production | Accepted: Added at production

DOI: xxx/xxxx

ARTICLE TYPE

Optimal control of a solar sail

Alesia Herasimenka*[1] | Lamberto Dell'Elce[2] | Jean-Baptiste Caillau[1] | Jean-Baptiste Pomet[2]

[1] Université Côte d'Azur, CNRS, Inria, LJAD, Nice, France
[2]Inria, Sophia Antipolis, France

Correspondence
*Email:
alesia.herasimenka@univ-cotedazur.fr

Present Address
This is sample for present address text this is sample for present address text

Summary

Solar sails belong to controlled systems with a positivity constrained on the control, because they can not generated force toward the Sun. This constraint is even more restricted, as the resulting force is contained in some convex cone with the origin in its vertex. Consider a scenario of a solar sail orbiting around a planet or asteroid, which is wisely used for different mission like deorbiting or observation. We propose a methodology for solving optimal control problem for orbital maneuvers of the sails. Convex optimisation allows to find the admissible controls that are used as the initial guess for an optimal control problem. Pontryagin's maximum principle gives necessary conditions for optimality. thorough analysis of the dynamics makes possible to find a switch function allowing to detect the structure of the solution for a given costate. Finally, indirect technique of multiple shooting combined with homotopy is used to solve the problem.

KEYWORDS:
solar sailing, conical constraints, sum of squares relaxation, multiple shooting, differential continuation

1 | INTRODUCTION

1.1 | Solar radiation pressure

Solar sails are satellittes that use solar radiation pressure (SRP) as propulsion for orbital maneuvers. Caused by interaction between photons and the surface of the sail, SRP has magnitude which depends on the distance between the Sun and the sail, denoted by r. Let us denote the fixed solar flux at the Earth's distance $r_\oplus = 1$ AU by $\Phi_{SR} = 1367$ W m^{-2}. With c the speed of light, a simple model for the SRP is given by [1, Chap. 3]:

$$P_{SR} = \frac{\Phi_{SR}}{c}\left(\frac{r_\oplus}{r}\right)^2. \tag{1}$$

In this paper, similarly to [2], we consider a flat sail with surface A and mass m. Different optical and geometrical properties have impact on the resulting SRP force, which has components of the incoming, reflected, and thermal radiations, namely f_a, f_r, and f_e. Moreover, the reflected radiation has specular and diffuse contributions, f_{rs} and f_{ru}, respectively. Each force component has different magnitude and direction, that can be identified through the Sun-sail direction, denoted as $\hat{s}$ and the unit vector normal to the sail having a positive component along $\hat{s}$, $\hat{n}$. That is we assume that both sides of the sail have the same optical properties, so only the (non-oriented) direction of the normal to the plane representing the plane will describe its attitude. We also assume that it is possible to control the attitude, and the actual control will be the force generated by this attitude (see (8)). In this model, $\hat{n}$ belongs the projective plane $\mathbf{R}P^2$ that one can describe as the union of one open hemisphere (whose axis is $\hat{s}$) with a circle whose antipodal points are identified. Fixing some reference vector $\hat{t}$ in $\{\hat{s}\}^\perp$, one defines coordinates $(\beta, \delta) \in (-\pi/2, \pi/2) \times \mathbf{R}$

for $\hat{n}$ in the open hemisphere part setting as usual (see Figure 1b)

$$\hat{n} = \sin\beta(\cos\delta\hat{t} + \sin\delta\hat{o}) + \cos\beta\hat{s},$$

where $\hat{o} := \hat{s} \times \hat{t}$ (in order that $(\hat{t}, \hat{o}, \hat{s})$ defines a direct orthogonal frame). Note that this is not a chart[1] as no δ, even restricted to $\mathbf{R}/\pi\mathbf{Z}$, can be uniquely associated with the direction $\hat{s}$. (See also Remark **??**.) The angle β is the so-called solar-sail *pitch angle*. As shown in Fig. 1a, let us introduce the direction of specular reflection given by $\hat{\xi}$, and the tangent unit vector $\hat{t}$ lying in the plane generated by $\hat{s}$ and $\hat{n}$. These vectors are defined as:

$$\hat{t} := \frac{\hat{n} \times \hat{s}}{\|\hat{n} \times \hat{s}\|} \times \hat{n} = \frac{\hat{s} - \cos\beta\,\hat{n}}{\sin\beta}. \tag{2}$$

As shown in the sketch of Figure 1a, the force due to the incoming radiation, f_a, points along $\hat{s}$. The force provided by the specularly reflected radiation, f_{rs}, points along $\hat{\xi}$ and is caused by photons that are reflected symmetrically with respect to the normal of the sail, thus yielding an exchange of momentum. Diffuse reflection stems from the sail surface roughness, which causes photons to be uniformly reflected in all directions, yielding a component of the force toward the direction normal to the sail, $\hat{n}$. Finally, as the absorbed photons are re-radiated in all directions, the force f_e is generated, which is orthogonal to the sail surface and points again along $\hat{n}$.

We follow[3, Chap. 2] and express the unit vectors $\hat{s}$ and $\hat{\xi}$ in terms of $\hat{n}$ and $\hat{t}$:

$$\begin{aligned} \hat{s} &= \cos\beta\,\hat{n} + \sin\beta\,\hat{t} \\ \hat{\xi} &= \cos\beta\,\hat{n} - \sin\beta\,\hat{t}, \end{aligned} \tag{3}$$

so that the above-presented forces can be expressed as[4]:

$$\begin{aligned} f_a &= \varepsilon\cos\beta\,\hat{s} = \varepsilon\,\cos\beta(\cos\beta\,\hat{n} + \sin\beta\,\hat{t}) \\ f_{rs} &= \varepsilon\rho s\cos\beta\,\hat{\xi} = \varepsilon\,\rho s\,\cos\beta(\cos\beta\,\hat{n} - \sin\beta\,\hat{t}) \\ f_{ru} &= \varepsilon\,B_f\,\rho(1-s)\cos\beta\,\hat{n} \\ f_e &= \varepsilon\,(1-\rho)\frac{\varepsilon_f B_f - \varepsilon_b B_b}{\varepsilon_b + \varepsilon_f}\cos\beta\,\hat{n} \end{aligned} \tag{4}$$

AL / L: rewrite theses expressions in terms of b_1, b_2, b_3, and factor out $\cos\beta$ In (4), ε is equal to $AP_{SR}\,m^{-1}$, which combines optical and physical parameters of the sail, has small magnitude, $\rho \in [0,1]$ is the fraction of reflected radiation to total amount of radiation illuminating the sail, $s \in [0,1]$ the fraction of specularly reflected radiation to total reflected radiation, ε_b and ε_f are the back and front surface emissivity coefficients, respectively, and B_b and B_f are back and front non-Lambertian coefficients, respectively. The SRP force is found as:

$$f_{SRP} = f_a + f_{rs} + f_{ru} + f_e. \tag{5}$$

To fully describe the orientation of the solar sail in the 3D space we introduce second angle, δ, equal to an angle of the projection of $\hat{n}$ on the plane perpendicular to the sun line $\hat{s}$ and the orbital plane of the solar sail. Let us denote $\mathcal{S}$ the frame attached to the solar sail and whose vectors are formed by $\hat{s}$, and two orthogonal vectors, one lying in the orbital plane, and the other one orthogonal to it. Projection of $\hat{n}$ and $\hat{t}$ in $\mathcal{S}$ are given by:

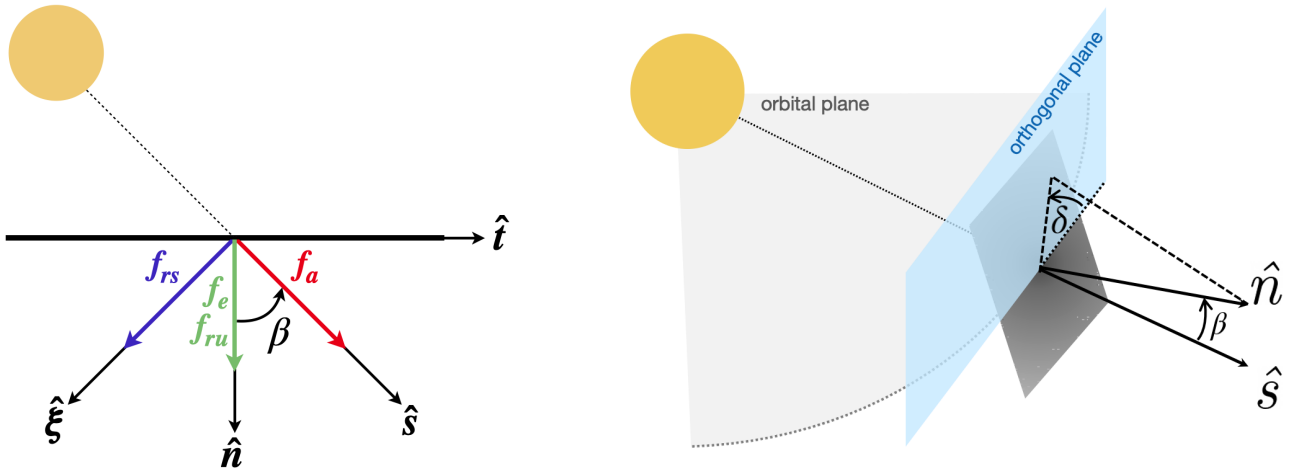
$$\hat{n} = \begin{pmatrix} \cos\beta \\ \sin\beta\sin\delta \\ \sin\beta\cos\delta \end{pmatrix}, \quad \hat{t} = \begin{pmatrix} \sin\beta \\ -\cos\beta\sin\delta \\ -\cos\beta\cos\delta \end{pmatrix}. \tag{6}$$

Finally, the SRP force projected on $\mathcal{S}$ is (note that because of the radial symmetry, its norm is independent of δ):

$$f_{SRP} = \cos\beta \begin{pmatrix} \cos^2\beta(1+\rho s) + B_f\rho(1-s)\cos\beta + (1-\rho)\frac{\varepsilon_f B_f - \varepsilon_b B_b}{\varepsilon_f + \varepsilon_b}\cos\beta + (1-\rho s)\sin^2\beta \\ 2\rho s\sin\beta\cos\beta\sin\delta + B_f\rho(1-s)\sin\beta\sin\delta + (1-\rho)\frac{\varepsilon_f B_f - \varepsilon_b B_b}{\varepsilon_f + \varepsilon_b}\sin\beta\sin\delta \\ 2\rho s\sin\beta\cos\beta\cos\delta + B_f\rho(1-s)\sin\beta\cos\delta + (1-\rho)\frac{\varepsilon_f B_f - \varepsilon_b B_b}{\varepsilon_f + \varepsilon_b}\sin\beta\cos\delta \end{pmatrix}. \tag{7}$$

Remark 1. In our modeling, the magnitude of the SRP is continuous, going to zero when the Sun direction is contained into the sail plane (orthogonality of $\hat{s}$ and $\hat{n}$), but its direction is not: when going through $\hat{s} \perp \hat{n}$, the illuminated side of the sail (a thickless 2D object embedded into 3D space) is changed and the orientation of $\hat{n}$ is changed to opposite ($\beta = \pm\pi/2$ being changed to $-\beta$,

[1] One actually retrieves the universal cover of the pointed open hemisphere by restricting to (β, δ) in $(0, \pi/2) \times \mathbf{R}$.



(a) Schematic representation.

(b) Orientation angles β and δ of the solar sail.

Figure 1 Components of the SRP force and orientation of th.

still defining the same direction—a perpendicular to $\hat{s}$—in the projective plane). The resulting force, going to zero in such cases, is continuous but not smooth. This singularity is inherent to the modeling and would be removed in a more realistic approach describing the sail as a genuine 3D object. This lack of smoothness is nonetheless not crucial here since, as will be clear from the optimality analysis in Section 2, an optimal force will have discontinuities, being either zero or with $\beta \in (-\beta^*, \beta^*)$ and $0 < \beta^* < \pi/2$ (if we exclude the ideal case for which $\beta^* = \pi/2$). So flips of illuminated side will not be encountered.

AL: define b_1, b_2, b_3 as in Mengali'2005 and rewrite theses expressions in terms of b_1, b_2, b_3

1.2 | Parametrisation of the control set

Controlling the sail attitude, *i.e.* the normal vector $\hat{n}$, allows to change the direction and magnitude of the resulting SRP. A reliable inference of optical coefficients is indeed mandatory to accurately estimate the mapping between $\hat{n}$ and f_{SRP}. To carry out our analysis, solar sail dynamics is conveniently modeled as a nonlinear control-affine system (see Section 1.3), where the control variable is homogeneous to the force vector: $u := f_{SRP}/\varepsilon$. The control set $U \subset \mathbf{R}^3$ is then given by:

$$U = \left\{ u = \frac{f_{SRP}(\hat{n})}{\varepsilon},\ \hat{n} \in \mathbf{R}P^2 \right\}. \tag{8}$$

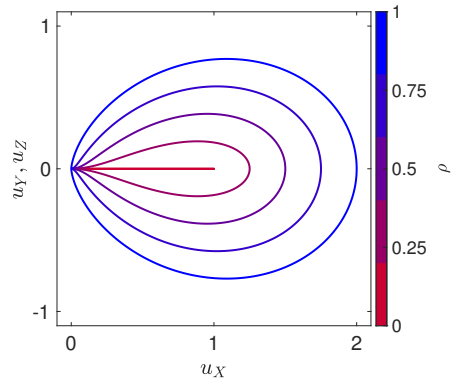
As for the normal vector $\hat{n}$, we use spherical coordinates (β, δ) as a set of local coordinates to parametrise the control set. Figure **??** shows the intersection of U on the plane generated by $\hat{n}$ and $\hat{s}$ for various optical properties. The set is a surface of revolution with axis $\hat{s}$, and it is not convex unless $\rho = s = 1$. Note that the interior of the surface is not part of U. When re-emitted radiation is neglected, which is most often a reasonable assumption for control purposes, U contains the origin but mapping between $\hat{n}$ and u is non-smooth at this point. Two extreme cases can be identified: ideal sails are constituted by perfectly-reflective surfaces ($\rho = s = 1$), whereas perfectly absorptive surfaces are the worst-case scenario ($\rho = 0$, f_e neglected) because SRP is systematically parallel to $\hat{s}$. Although sails are designed to be as close to ideal as possible, partial absorption of the energy is unavoidable in real-life applications and, in addition, optical properties exhibit degradation with time. Hence, the fraction of reflected radiation decreases with lifetime of the satellite, as discussed in[5,6].

AL: Comments here + figures about convexity/non-convexity of the control set: horizontal or vertical tangent

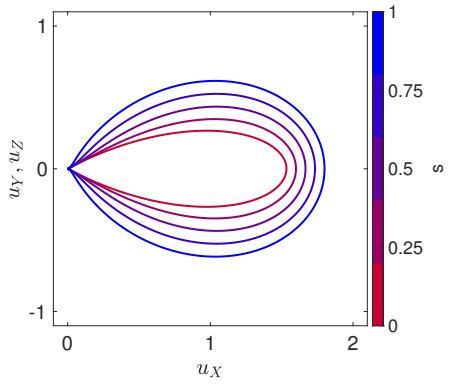
1.3 | Equations of motion

The following assumptions are introduced:

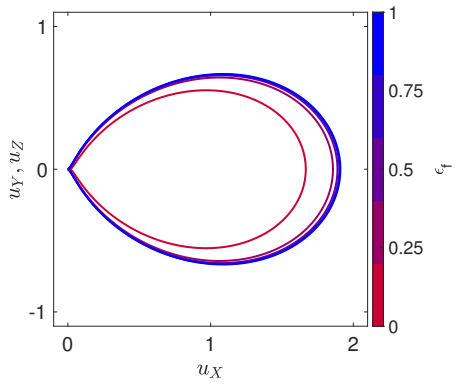
1. Orbital period of the sail is much smaller than the one of the heliocentric orbit of the attractor, so that variations of the Sun direction $\hat{s}$ over a single orbit of the sail are neglected.



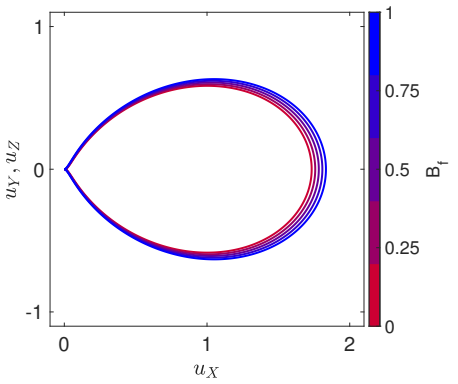
(a) Control sets for different reflectivity coefficients and $s = 1$.



(b) Control sets for different specular reflectivity coefficients and $\rho = 0.8$



(c) Control sets for different front emissivity coeffucuents.



(d) Control sets for different front non-Lambertian coeffucuentsl.

2. Solar eclipses are neglected. Introducing solar eclipses can be a way of improving the proposed algorithm.
3. Re-emitted radiation is neglected. In fact, this component of SRP can be reasonably regarded as a disturbance for control purposes.

Equations of motion are written in a set of Keplerian-like orbital elements, which leverages on the axial symmetry of the problem with respect to the Sun's direction. Namely, consider a reference frame $\mathcal{S}$ with origin at the center of the planet, $\hat{X}$ axis towards $\hat{s}$, $\hat{Y}$ lies in the plane of the planet's orbit around the Sun and is orthogonal to $\hat{X}$, and $\hat{Z}$ is chosen to form a right-hand frame. Because this study focuses on short-time controllability (characteristic time is of the order of one orbital period), motion of this frame is neglected by virtue of the first assumption above. Figure 3 represents the vectors h, e and $\hat{N}$, which denote the angular momentum, eccentricity and ascending node vectors, respectively. Let $\gamma_1, \gamma_2, \gamma_3$ be Euler angles orienting the eccentricity vector according to a X-Y-X rotation as depicted in Fig. 3, so that γ_2 is the angle between the angular momentum of the orbit and the Sun direction, and a, e, and f be semi-major axis, eccentricity and true anomaly, respectively. The motion of slow elements, $I = \left(\gamma_1, \gamma_2, \gamma_3, a, e\right)^T \in \mathcal{M}$, where $\mathcal{M}$ is the configuration manifold, and fast angle f is governed by

$$\begin{aligned} \frac{\mathrm{d}\, I}{\mathrm{d}\, t} &= \varepsilon \sqrt{\frac{a\left(1-e^2\right)}{\mu}} G_0(I,f)\, R(I,f)\, u, \\ \frac{\mathrm{d}\, f}{\mathrm{d}\, t} &= \omega(I,f) + \varepsilon\, F(I,f)\, R(I,f)\, u, \end{aligned} \tag{9}$$

where components of u are in the reference frame $\mathcal{S}$, $R(I,f) = R_X(\gamma_3 + f)R_Y(\gamma_2)R_X(\gamma_1)$ is the rotation matrix from reference to local-vertical local-horizontal frames[2],

$$\omega(I,f) = \sqrt{\frac{\mu}{a\,(1-e^2)^3}}(1 + e\cos f)^2. \tag{10}$$

[2]Here, $R_A(f)$ denotes the rotation matrix of angle f about the axis $\hat{A}$.

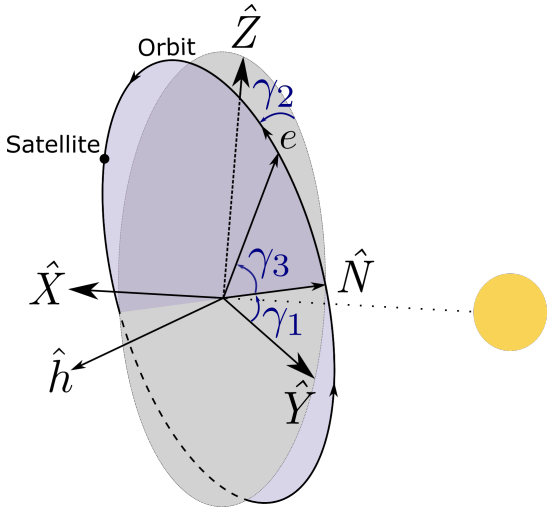


Figure 3 Euler angles γ_i orienting the orbit according to a $\hat{X}$-$\hat{Y}$-$\hat{X}$ rotation with respect to the reference frame $\mathcal{S}$. Here, h and e denote the angular momentum and eccentricity vectors.

Both $F(I, f)$ and $G_0(I, f)$ can be deduced from Gauss variational equations (GVE) of classical elements, where

$$G_0(I,f) = \begin{pmatrix} 0 & 0 & \frac{\sin\left(\gamma_3+f\right)}{\sin\gamma_2(1+e\cos f)} \\ 0 & 0 & \frac{\cos\left(\gamma_3+f\right)}{1+e\cos f} \\ -\frac{\cos f}{e} & \frac{2+e\cos f}{1+e\cos f}\frac{\sin f}{e} & \frac{\cos\left(\gamma_3+f\right)}{1+e\cos f} \\ \frac{2ae}{1-e^2}\sin f & \frac{2ae}{1-e^2}(1+e\cos f) & 0 \\ \sin f & \frac{e\cos^2 f+2\cos f+e}{1+e\cos f} & 0 \end{pmatrix}. \tag{11}$$

This peculiar choice of Euler angles follows from the symmetry of System (2), namely axial symmetry with respect to the axis $\hat{\boldsymbol{X}}$, and it has the main consequence that controllability results in Section V are independent of γ_1, which is a rotation about this axis.

2 | CONTROL OVER ONE ORBITAL PERIOD

2.1 | Optimal control formulation

We are interested in moving solar sail in the desired direction after one orbital period. Therefore, it is interesting to rewrite System (9) in terms of displacement of the slow state elements, denoted δI. Therefore, its dynamics is given by

$$\delta I' = \varepsilon G(I, f)u \tag{12}$$

where $' := \mathrm{d}/\mathrm{d}f$ and with

$$G(I,f) := \frac{a\left(1-e^2\right)^2}{\mu(1+e\cos f)^3} G_0(I,f)R(I,f).$$

As mentioned earlier, SRP has a very small magnitude, this is why it is usually considered as a perturbation. Thus, changes on slow variables are very small over one orbital period so that I will be assumed *constant* in the rest of the paper. The goal is to maximise the size of the displacement in a given direction fixed by a unit vector, d_I, so that the final value of δI is parallel to d_I. This problem can be written in Mayer form as follows (note the simple form of the dynamics, given by an explicit integral,

as the right-hand side does not depend on δI in our approximation):

$$\max_{u(f)\in U} (\delta I(2\pi)|d_I) \quad \text{subject to} \quad \delta I' = \varepsilon \sum_{i=1}^{3} u_i G_i(I,f), \quad \delta I(0) = 0, \quad \delta I(2\pi) \text{ parallel to } d_I. \tag{13}$$

Building upon results in [11?], we have access to an effective test (related to the convex SDP approximation discussed in Section 3) to check that it is indeed possible to move in the direction d_I after one revolution. So we assume in the sequel that the problem is controllable.

2.2 | Existence and necessary conditions for optimality

We first consider the relaxation of (13) obtained by replacing the control set U by its convex hull: $u(f) \in \text{conv}(U)$ (see Figure **??**). As the control set is now compact and convex, and since we have assumed controllability using controls valued in $U \subset \text{conv}(U)$, Filippov theorem entails that

Proposition 1. The relaxed problem has a solution.

To formulate the necessary optimality conditions for the problem on $\text{conv}(U)$ we introduce the costate $p_{\delta I}$ of δI, a covector of dimension 5. The Hamiltonian associated with the dynamics is

$$H(I,f,p_{\delta I},u) = \varepsilon p_{\delta I} G(I,f) u. \tag{14}$$

Remember that I is a constant, and note that the Hamiltonian does not depend on the state δI because of the very simple form of the dynamics. (The ODE defines a mere quadrature, here.) Clearly, $p_{\delta I}$ is constant and transversality conditions write

$$(p_{\delta I}|d_I) = -p^0 \|d_I\|^2 = -p^0 \tag{15}$$

where p^0 is the nonpositive multiplier associated with the cost. In particular, $p_{\delta I}$ is not zero, since otherwise both p^0 and $p_{\delta I}$ would vanish. By homogeneity in $(p^0, p_{\delta I})$ there are two cases: (i) the abnormal case ($p^0 = 0$) when $(p_{\delta I}|d_I) = 0$ and where one can normalise setting $\|p_{\delta I}\| = 1$; (ii) the normal case ($p^0 < 0$) when $(p_{\delta I}|d_I) > 0$ and where one can normalise setting $(p_{\delta I}|d_I) = 1$. Let us set $\psi := p_{\delta I} G(I,f)$.

Lemma 1. For any I, the matrix formed by $G(I,f)$ and $\partial G(I,f)/\partial f$ has maximum rank for all $f \in [0, 2\pi]$.

Proof. This computation is actually equivalent to the rank condition that can be verified in terms of Lie brackets (and, *e.g.*, Cartesian coordinates) in [7] (check Lemma 1). □

As a result, the zeros of the dimension three covector ψ (as a function of the true anomaly f) are isolated on $[0, 2\pi]$. Indeed, the previous lemma implies that ψ and $\mathrm{d}\psi/\mathrm{d}f$ cannot vanish simultaneously as then, $p_{\delta I}$ would be orthogonal to all columns of $G(I,f)$ and of its derivative, so $p_{\delta I}$ would be zero (a contradiction). So there are only finitely many such zeros on $[0, 2\pi]$, defining a locus of codimension greater than one in the (I,f) space. For the sake of simplicity, we assume in the sequel that ψ actually never vanishes. For a detailed discussion on the associated singularities of the dynamics, see [8].

Let K_α be the convex cone generated both by U and by its convex hull, α denoting the half-angle at the cone vertex. The polar cone K_α^0 is the set of directions having a nonpositive scalar product with those in K_α. The drop-shaped curve obtained when intersecting the control set with a plane is parametrised by the angle β alone, and we denote $\beta^* \in (0, \pi/2)$ the parameter associated with the tangency point of this curve with its conical hull (see Figure **??**). In the sequel, we recall and complete the analysis from [9], providing precise bounds on the number of switchings on the control.

Lemma 2. [9] The angle β^* is solution of

$$\cos\beta^* = \frac{-b_1 b_3 - 2b_2 b_3 + \sqrt{b_1^2 b_3^2 - 4b_1 b_2 b_3^2 + 8b_1^2 b_2^2 + 4b_1 b_2^3}}{4b_1 b_2 + 2b_2^2}.$$

Proposition 2. [9] An optimal control u verifies the following: (i) when ψ belongs to the interior of K_α^0, u is zero; (ii) when ψ does not belong to K_α^0, the coordinates (β, δ) of the control verify the following relations:

$$\psi_1 \sin\beta(b_1 + 3b_2\cos^2\beta + 2b_3\cos\beta) - \sqrt{\psi_2^2 + \psi_3^2}\left(\cos^2\beta(b_2\cos\beta + b_3) - \sin^2\beta(2b_2\cos\beta + b_3)\right), \ \beta \in (-\beta^*, \beta^*), \tag{16}$$

and

$$\delta = \pi/2 - \arg(\psi_2 + i\psi_3) \bmod \pi. \tag{17}$$

Moreover, any optimal control is made of finitely many subarcs corresponding to case (i) or (ii), and has at most 8 switchings (transverse contacts with ∂K_α^0) over one period.

Proof. According to Pontrjagin maximum principle and to the expression (**??**) of the Hamiltonian, for almost all true anomaly f an optimal control must be a maximizer of the scalar product $(\psi|u)$ for u in $\mathrm{conv}(U)$. Clearly, when ψ belongs to the interior of the polar cone of K_α, this scalar product is negative for any nonzero u, so $u = 0$ is the only maximizer. Conversely, when ψ belong to the open complement of K_α^0, maximizers must annihilate the gradient of the Hamiltonian with respect to the chosen coordinates of the control,

$$\frac{\partial H}{\partial \beta} = 0, \quad \frac{\partial H}{\partial \delta} = 0,$$

which gives the expressions in alternative (ii) of the statement. Moreover, ψ belongs to the boundary of K_α^0 if and only if $\psi_1 \cos\alpha + \sqrt{\psi_2^2 + \psi_3^2}\,\sin\alpha = 0$, implying that

$$\psi_1^2 \cos^2\alpha - (\psi_2^2 + \psi_3^2)\sin^2\alpha = 0. \tag{18}$$

Every component of ψ is trigonometric in f, and this (nontrivial) equation results in a trigonometric polynomial of degree 4. As it has isolated zeros, there are finitely many zeros (at most eight, see Remark 2) defining isolated contacts with ∂K_α^0. □

Remark 2. Roots of a trigonometric polynomial can be found using companion-matrix methods [10]. Consider the degree 4 polynomial

$$\mathcal{T}(f) = \sum_{j=0}^{4} a_j \cos(jf) + \sum_{j=1}^{4} b_j \sin(jf).$$

Fourier-Frobenius companion matrix elements are

$$B_{jk} = \begin{cases} \delta_{j,k-1}, & j = 1,\dots,7,\ k = 1,\dots,8, \\ (-1)\dfrac{h_{k-1}}{a_4 - ib_4}, & j = 8,\ k = 1,\dots,8, \end{cases} \tag{19}$$

where δ_{jk} are the Kronecker functions such that $\delta_{jk} = 0$ if $j \neq k$ and $\delta_{jj} = 1$, and h_k are

$$h_k = \begin{cases} a_{4-k} + ib_{4-k}, & k = 0,\dots,3, \\ 2a_0, & k = 4, \\ a_{k-4} - ib_{k-4}, & k = 5,\dots,8. \end{cases} \tag{20}$$

The roots of $\mathcal{T}(f)$ are obtained from eigenvalues z_k of the matrix defined in Eq. (19) as

$$f_{k,m} = arg(z_k) - i\log(|z_k|) \bmod (2\pi), \quad k = 1,\dots,8.$$

Real-valued roots of $\mathcal{T}(f)$ are such that $|z_k| = 1$. Therefore, this technique allows to find roots of the switch function and, thus, find out the structure of the solution for a given costate. It is important to stress that the trigonometric polynomial is of degree 4, which means that the switching function can have at most 8 roots.

Corollary 1. The original optimal control problem (13) has a solution.

Proof. The relaxed problem has at least one solution (Proposition 1), and any control solution actually belongs to U by virtue of Proposition 2. Such controls must be optimal for the original problem, whence existence. □

3 | SOLUTION USING CONVEX OPTIMISATION AND CONTINUATION

3.1 | Convex approximation for a reliable initial guess

In order to use indirect shooting methods for solving optimal control problem, we need first a reliable initial guess for the costate $p_{\delta I}$. We propose an approximation by a convex mathematical program similar to the one used in [11] for controllability check

purposes. To this end, define the bounded cone $\hat{K}_\alpha$ obtained by truncating the K_α at its tangency points with U (check Figure 4a). This cone is bounded by a disk denoted D_α. This new control set is a subset of the convex hull of U, in order that any solution of

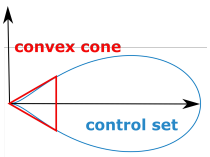
$$\max_{u(f)\in\hat{K}_\alpha} (\delta I(2\pi)|d_I) \quad \text{subject to} \quad \delta I' = \varepsilon \sum_{i=1}^{3} u_i G_i(I,f), \quad \delta I(0) = 0, \quad \delta I(2\pi) \text{ parallel to } d_I, \tag{21}$$

will define an admissible control for the convex relaxation of the original control problem. Note that existence holds for this new problem (Filippov again, as $\hat{K}_\alpha$ is convex and bounded) and that any solution will also have a bang-bang structure. A similar analysis to the one of Section 2.2 on conv(U) indeed allows to prove that

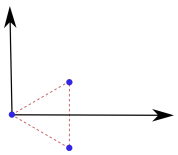
Proposition 3. An optimal control u of problem (21) on $\hat{K}_\alpha$ verifies the following: (i) when ψ belongs to the interior of K_α^0, u is zero; when ψ does not belong to K_α^0, (ii-a) the control is uniquely determined and belongs to the circle ∂D_α, unless (ii-b) ψ is colinear to the axis $\hat{s}$ of the cone K_α in which case the control still belongs to the ∂D_α but is not uniquely determined. Moreover, any optimal control is made of finitely many subarcs corresponding to case (i) or (ii-a) over one period.

Proof. As $\hat{K}_\alpha$ and K_α have the same polar cone, (i) is clear. Conversely, when ψ belongs to the open complement of K_α^0, the colinearity condition $\psi \wedge \hat{s} = 0$ boils down to checking a polynomial condition in f and has only isolated zeros corresponding to case (ii-b). When ψ is not colinear to $\hat{s}$, the unique maximizer of $(\psi|u)$ for u in $\hat{K}_\alpha$ indeed belongs to the circle ∂D_α, which is case (ii-a). □

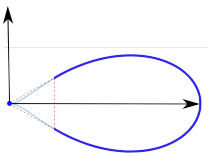
This structure being analogous to that of solutions of the original problem, one hopes to retrieve a reasonable approximation to be used to initiate a differential continuation (see Section 3.2). In particular, we note that the original problem (13) on U and problem (21) on $\hat{K}_\alpha$ share the same switching function associated with contacts with ∂K_α^0 and given by (18).



(a) Approximation of the control set by a cone bounded at the points of tangency.



(b) Optimal solution will be given by two controls on the boundary of the cone or a zero.



(c) Possible optimal solutions on the real control set.

Figure 4 Approximation of the control set by a convex cone.

Consider the following discretization of (21): the control set $\hat{K}_\alpha$ is approximated by a polyhedral cone $\hat{K}_\alpha^g \subset \hat{K}_\alpha$ generated as the convex hull of g vertices $V_1, \dots, V_g$ chosen in $\partial\hat{K}_\alpha$. (Note that the 3D cone $\hat{K}_\alpha$ is not finitely generated.) Any control in $\hat{K}_\alpha^g$ is given by a convex combination

$$u(f) = \sum_{j=1}^{g} \nu_j(f) V_j, \quad \nu_j(f) \geq 0, \quad \sum_{j=1}^{g} \nu_j(f) = 1, \quad f \in \mathbf{S}^1, \quad j = 1, \dots, g. \tag{22}$$

The functions ν_j are modeled using an N-dimensional basis of trigonometric polynomials, $\Phi(f) = \left(1, e^{if}, e^{2if}, \dots, e^{(N-1)if}\right)$:

$$\nu_j(f) = \left(\Phi(f) \mid c_j\right)_H \tag{23}$$

with $c_j \in \mathbf{C}^N$ complex-valued coordinates of ν_j in $\Phi(f)$, and $(\cdot|\cdot)_H$ stands for the Hermitian product on $\mathbf{C}^N$. To enforce the positivity constraint, we leverage on the formalism of squared functional systems outlined in [12]. It allows to recast continuous positivity constraints into linear matrix inequalities (LMI), that can be solved using convex optimisation. Actually, $\Phi(f)$ has a corresponding squared functional system given by $S^2(f) = \Phi(f)\Phi^H(f)$, with $\Phi^H(f)$ being the conjugate transpose of $\Phi(f)$. According to [12], let us define a linear operator $\Lambda_H : \mathbf{C}^N \to \mathbf{C}^{N\times N}$ that maps coefficients of a polynomial in $\Phi(f)$ to its squared base, and its adjoint operator $\Lambda_H^* : \mathbf{C}^{N\times N} \to \mathbf{C}^N$ such that

$$\left(Y|\Lambda_H(c)\right)_H = \left(\Lambda_H^*(Y)|c\right)_H, \quad Y \in \mathbf{C}^{N\times N}. \tag{24}$$

The theory of squared functional systems states that, for the trigonometric polynomial $(\Phi(f)|c)$ to be non-negative, it is sufficient that there exists a Hermitian semidefinite positive matrix $Y \succeq 0$, such that

$$\Lambda_H^*(Y) = c.$$

In the case of trigonometric polynomials, the operator Λ_H^* is defined by means of Toeplitz matrices:

$$\Lambda_H^*(Y) = \begin{bmatrix} \mathrm{tr}\left(Y|T_0\right) \\ \vdots \\ \mathrm{tr}\left(Y|T_{N-1}\right) \end{bmatrix} \tag{25}$$

with kl-coefficients of T_j such that

$$\begin{aligned} &T_0 = I, \\ &T_j^{(k,l)} = \begin{cases} 2 & \text{if } k-l=j \\ 0 & \text{otherwise} \end{cases} \quad j = 1, \dots, N-1. \end{aligned} \tag{26}$$

For an admissible control u valued in $\hat{K}_\alpha^g$, one has

$$\int_0^{2\pi} \sum_{i=1}^{3} u_i(f) G_i(I, f)\, \mathrm{d}f = \sum_{j=1}^{g} \left(L_j c_j + \bar{L}_j \bar{c}_j\right) \tag{27}$$

with $L_j(I)$ in $\mathbf{C}^{5\times N}$ defined by

$$L_j(I) = \frac{1}{2} \sum_{i=1}^{3} \int_{\mathbf{S}^1} V_{ij} G_i(I, f) \Phi^H(f)\, \mathrm{d}f, \tag{28}$$

where $V_j = (V_{ij})_{i=1,\dots,3}$. We note that the components of $L_j(I)$ are Fourier coefficients of the function $\sum_{i=1}^3 V_{ij} G_i(I, f)$. $L_j(I)$ are approximated using the discrete Fourier transform (DFT). Since vector fields G_i are smooth, truncation of the series is justified by the fast decrease of the coefficients. Finally, for a control u valued in $\hat{K}_\alpha^g$, coefficients ν_j are truncated Fourier series of order $N-1$. As a result, for a given vector d_I, the SDP approximation is

$$\begin{aligned} \max_{c_j \in \mathbf{C}^N,\ Y_j \in \mathbf{C}^{N\times N}} (\delta I | d_I) \quad \text{subject to} \quad & \delta I = \varepsilon \sum_{j=1}^{g} \left(L_j c_j + \bar{L}_j \bar{c}_j\right) \text{ parallel to } d_I \\ & Y_j \succeq 0, \quad \Lambda^*\left(Y_j\right) = c_j, \quad j = 1, \dots, g. \end{aligned} \tag{29}$$

L: add the discretization of the constraint $\sum_j \nu_j(f) = 1$ The Lagrange variable of the discretization of the equality constraint that δI is parallel to d_I from the convex program is expected to be a fair approximation of the costate $p_{\delta I}$ of (21). More importantly, it is hoped that the bang-bang control structure associated with this $p_{\delta I}$ is indeed the same as for the solution of the problem defined on $\hat{K}_\alpha$.

3.2 | Multiple shooting, differential continuation and callback

Homotopy, *aka.* continuation, allows to solve a complex problem by connecting it continuously to a simpler problem. The idea is then to follow the path (assumed to be regular enough) of solutions from the simpler problem towards the targeted one. See, *e.g.*,[13,14] for applications in optimal control. In our case, a parameter λ defined between 0 and 1 allows to connect the problem with control set the bounded convex cone $\hat{K}_\alpha$ at $\lambda = 0$, to the original problem with the non-convex drop-like control set U at $\lambda = 1$. In order to be able to solve the problem for $\lambda = 0$, we rely on the solution of the convex program on $\hat{K}_\alpha^g$ to provide an admissible solution. This solution is used not only to compute an educated guess for the initial costate but also to devise the appropriate multiple shooting function. To do so, we use the control structure corresponding to the approximation of $p_{\delta I}$ provided by the convex optimization and described at Proposition 3. This proposition tells us that, when ψ (a function of $p_{\delta I}$ and f) belongs to the open complement of the polar cone K_α^0, the control must be equal to the *dynamical feedback* described case (ii-a) (apart for some isolated points that correspond to case (ii-b) that we can neglect); we denote $u_b^0(f, p_{\delta I})$ this control. Similarly, for such values of ψ, Proposition 2 for the problem on conv(U)—and actually U, check Corollary 1—, implies that the control must be a solution of (16)-(17). (While these equations provide an explicit solution for the coordinate δ of the control,

β is only implicitly defined and we discuss its actual computation in Section 3.3.) We assume that this solution is unique and denote it $u_b^1(f, p_{\delta I})$. Then, for λ in $[0, 1]$ and ψ outside the polar cone, we define

$$u_b(f, p_{\delta I}, \lambda) := (1 - \lambda)u_b^0(f, p_{\delta I}) + \lambda u_b^1(f, p_{\delta I})$$

as the convex combination of the dynamical feedbacks for $\lambda = 0$ and $\lambda = 1$. Conversely, for any λ in $[0, 1]$ and ψ in the interior of the polar cone, the control is set to zero.

For a given λ, one has a finite sequence of arcs with either $u = u_b$ (bang arcs), or $u = 0$ (zero arcs). Contacts with ∂K_α^0 are characterized by (18) whose left-hand side defines the switching function, denoted $\varphi(f, p_{\delta I})$ (not depending on λ in our particular setting). To this finite sequence of arcs is associated a multiple shooting function in a standard fashion. Assume for instance that the structure is bang-zero-bang. Then the shooting function has three arguments: the (constant) value of the costate, $p_{\delta I}$, and the two switchings times (true anomalies) bounding the central zero arc, f_1 and f_2. (So that $(p_{\delta I}, f_1, f_2)$ belong to $\mathbf{R}^7$.) Plugging $u = u_b(f, p_{\delta I}, \lambda)$ into the dynamics of δI and integrating on $[0, f_1]$ from $\delta I(0) = 0$ allows to compute $\delta I_1 := \delta I(f_1)$. As the control is zero on $[f_1, f_2]$, δI remains constant on the coast arc and we set $\delta I_2 := \delta I_1$. The control $u = u_b(f, p_{\delta I}, \lambda)$ is eventually plugged again on $[f_2, 2\pi]$ to compute $\delta I_f := \delta I(2\pi)$, starting from δI_2. The associated value of the shooting function is obtained by concatenating the left-hand side of the four equations below, forming a vector of dimension $4+1+2 = 7$ (note that the first colinearity equation indeed has dimension $5 - 1 = 4$):

$$\delta I_f \wedge d_I = 0, \tag{30}$$

$$(p_{\delta I} | d_I) - 1 = 0, \tag{31}$$

$$\varphi(f_1, p_{\delta I}) = 0, \tag{32}$$

$$\varphi(f_2, p_{\delta I}) = 0. \tag{33}$$

This defines a shooting function $S(\xi, \lambda)$ with, for this bang-zero-bang structure, $\xi := (p_{\delta I}, f_1, f_2)$. Once the first solution for $\lambda = 0$ is obtained, the path of zeros is followed by differential continuation, typically using a parametrisation by its curvilinear abscissa:

$$s \mapsto (\lambda(s), \xi(s)) \text{ with } S(\xi(s), \lambda(s)) = 0.$$

We refer, *e.g.*, to [15] for the assumptions needed to do so. Note that, according to (32), we look for normal extremals (compare with (15)).

One important issue in practice is that it might not be possible to reach $\lambda = 1$ because, at some $\lambda(\bar{s})$ in $(0, 1)$, the structure of the solution changes; for instance because one subarcs disappears. It is crucial to be able to detect such a change during homotopy since then, the shooting function has to be redefined according to the new structure. This is achieved using a standard callback mechanism along with differential continuation. On the previous bang-zero-bang example, the continuation is monitored and, at each step of the path following procedure, a simple test is performed: if the exit time of the zero arc, f_2, becomes inferior to the entry time f_1 (this is detected by a sign change on $f_2 - f_1$, as going forward in time makes sense mathematically but is not allowed to obtain admissible trajectories), the continuation is stopped. And restarted at $\lambda(\bar{s})$ with a new shooting function (in this case, a single shooting one, as only one bang arc would be left), using $\xi(\bar{s})$ as initial guess. More elaborated tests can be constructed to detect a new arc appearing, *etc.* In our case, a callback is used to detect a structure change from 5 subarcs to 3 (see Section 4).

3.3 | Implicit treatment of the Hamiltonian maximisation

Regarding the computation of $u_b^1(f, p_{\delta I})$, we know after Proposition 2 that the control is either zero, either solution of (16-17). The first equation for the coordinate β of u has no closed form solution. There is a preliminary numerical discussion of the number of solutions in [9] (we actually look for a global maximizer of the Hamiltonian over U, which may allow to eliminate some strictly local minimizer that also verify (16)) for a particular set of values of the sail parameters. More generally, while maximization of the Hamiltonian often yields an explicit expression of the control as a dynamics feedback function of the state and the costate, it is not always the case. In such a situation, we advocate an implicit treatment of this maximization, incorporating the stationarity equation of the Hamiltonian into the shooting procedure. We sketch below a simple way to do so in a general setting.

Assume that, after applying Pontrjagin maximum principle, one has to integrate the following system (x denoting the state, p the costate):

$$\dot{x}(t) = \nabla_p H(x(t), p(t), u(t)), \quad \dot{p}(t) = -\nabla_x H(x(t), p(t), u(t)), \tag{34}$$

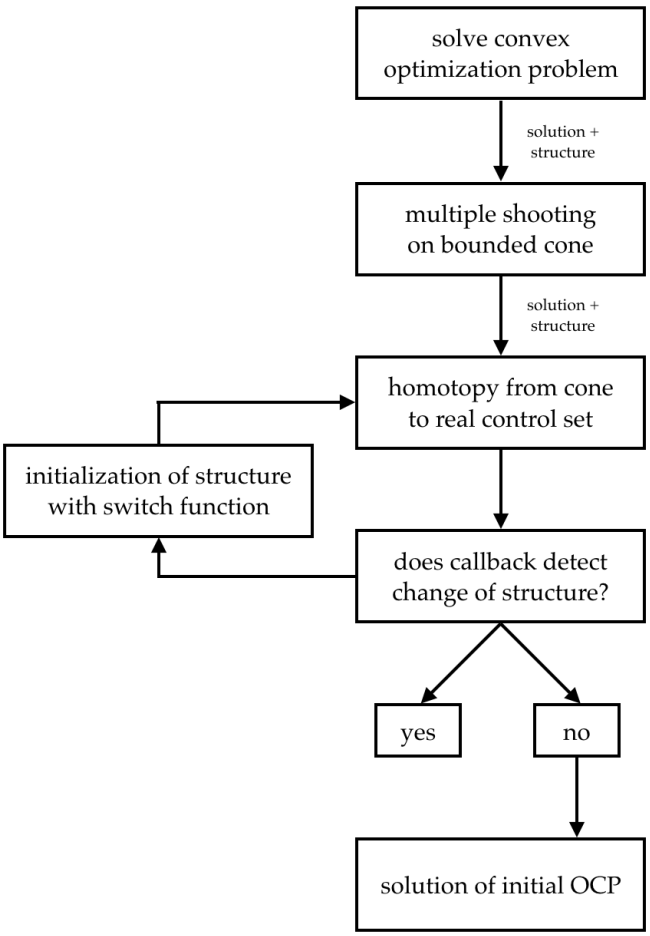


Figure 5 Algorithm for solving optimal control problem (OCP).

where, at each time t, $u(t)$ verifies

$$\nabla_u H(x(t), p(t), u(t)) = 0. \tag{35}$$

The last stationarity equation corresponds to an unconstrained situation—whereas a Lagrangian, plus an additional finite dimensional multiplier, should be considered in the presence of constraints—, and defines a semi-explicit DAE. Assume that the strong Legendre-Clebsch condition holds in an open neighborhood of the reference extremal times the open control set, $\nabla^2_{uu} H \geq cI$ for some positive constant c. Then the Hamiltonian has a unique maximizer, that satisfies $\nabla_u H = 0$, and the previous DAE is of index 1 (differentiating once (35) allows to solve for $\dot{u}$). In particular, one can extend the Hamiltonian system (34) by adding the equation

$$\dot{u} = -\nabla^2_{uu} H^{-1}(\nabla_{ux} H \cdot \nabla_p H - \nabla_{up} H \cdot \nabla_x H)(x, p, u) := g(x, p, u),$$

with initial condition $\nabla_u H(x(0), p(0), u(0)) = 0$. The new system remains Hamiltonian as is clear setting $\hat{x} := (x, u)$, $\hat{p} := (p, p_u)$ and

$$\hat{H}(x, u, p, p_u) := H(x, p, u) + (p_u | g(x, p, u))$$

with $p_u(0) = 0$. (One can obviously eliminate the trivial equation on p_u, which is an extra but identically zero costate.) In the case of a shooting approach, the value of $u(0)$ is an additional shooting variable. Keeping the system in Hamiltonian form is convenient in the algorithmic framework described in Section 4, but other approaches for DAE such as predictor-corrector ones can of course be considered. In our case, we use this approach with $x = \delta I$, $p = p_{\delta I}$ to deal with the implicit equation (16) on β (while we use (17) to solve explicitly for δ). The combination of this implicit approach with multiple shooting, homotopy and callback is described in the last section.

4 | NUMERICAL EXAMPLES

The OCP (**??**) is solved using *control toolbox (CT)* and *nutopy* package for python. An example of the code which executable online is available.[3] The results are presented in Sec. **??**.

Consider again System (**??**). Optical properties of the sail determining shape of U are taken from JPL Square Sail defined in[3, Table 2.1]: $\rho = 0.88$, $s = 0.94$, $\varepsilon_b = 0.55$, $\varepsilon_f = 0.05$, $B_b = 0.55$, $B_f = 0.79$.

We consider an example such that a structure change occurs during continuation. The initial conditions are: $d_I = (0, 1, 0, 0, 0)$, what translates increase of inclination γ_2, and $I = (10°, 50°, 30°, 1, 0.1)$.

Figs. 6a and 6c show solution of the convex optimisation program on a bounded cone. The adjoint vector is

$$p_I^{conv} = (-0.0837, 1, -0.0052, 0.0398, 0.0852).$$

Using it as the initial guess to solve OCP on a real control set, the solution is represented in Figs. 6b and 6d. The corresponding adjoint is

$$p_I^{OCP} = (-0.1637, 1, -0.0972, 0.0712, 1.6037).$$

The first solution has consists of 5 bangs, with 4 switches. However, the second solution has only 2 switches and contains 2 bangs and one *singular arc*. This difference is due to different co-vectors and dynamics close to the polar cone, which results in different roots quantity of the switch function.

Similarly, Fig. 6e and 6f plots trajectory resulting from integration of the system by injecting the solutions from convex programming and OCP respectively. The direction of displacement is increase of inclination, what corresponds to the requirement.

One can notice that the solutions given by convex optimisation problems have a stairs shape. It is indeed due to the discretization of the convex cone using a finite number of generators. It can be easily verified by solving the OCP using multiple shooting and the same initial guess directly on the bounded cone instead of the real control set. Fig.7 compares these two approaches using the data from the first case study. Again, solutions on the left are taken directly from the convex computation, when the right-side solutions are provided by OCP on a bounded cone. They are similar, except for the discrete form due to cone discretization as shown in Fig. **??**.

CONCLUSION

ACKNOWLEDGEMENTS

This work was partially supported by ESA (contract no. 4000134950/21/NL/GLC/my).

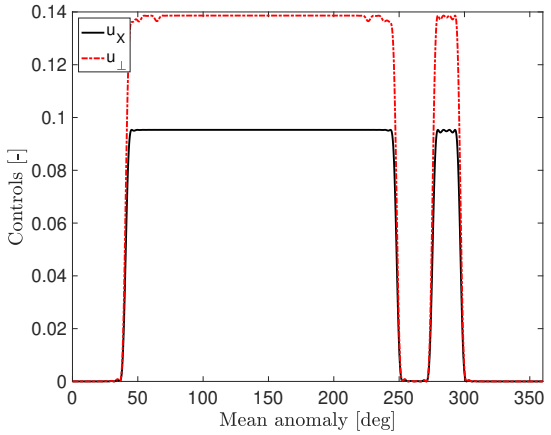
SUPPORTING INFORMATION

The following supporting information is available as part of the online article:

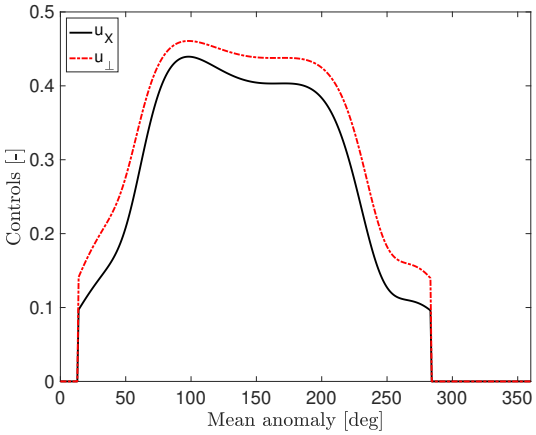
References

1. Montenbruck O, Gill E. *Satellite Orbits*. Springer Science + Business Media . 2000
2. Herasimenka A, Dell'Elce L, Caillau JB, Pomet JB. Controllability Properties of Solar Sails. *Journal of Guidance, Control, and Dynamics* 2023: 1–10. doi: 10.2514/1.g007250
3. McInnes CR. *Solar Sailing*. Springer London . 1999
4. Rios-Reyes L, Scheeres DJ. Generalized Model for Solar Sails. *Journal of Spacecraft and Rockets* 2005; 42(1): 182–185. doi: 10.2514/1.9054

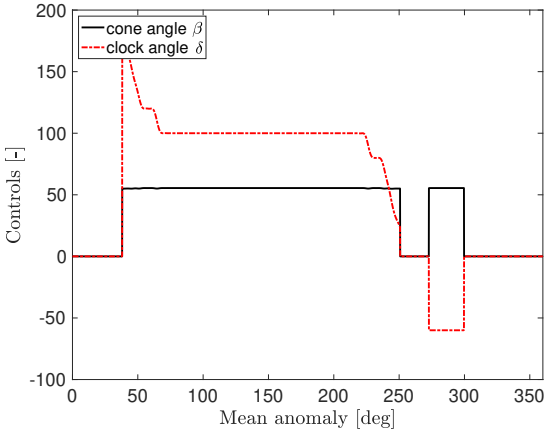
[3] https://ct.gitlabpages.inria.fr/gallery/solarsail/solarsail-simple-version.html



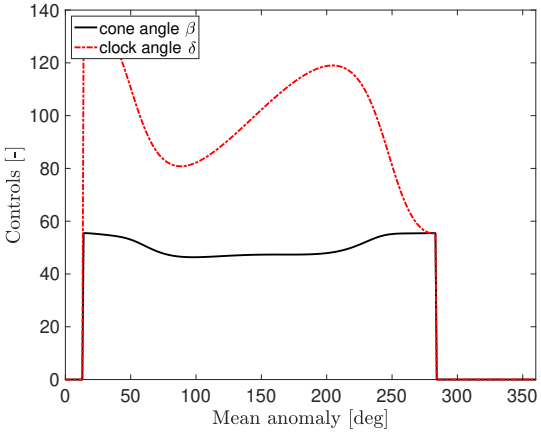
(a) Solution of the optimisation problem (29) on a bounded convex cone as projection on $\hat{s}$, $\hat{s}^\perp$.



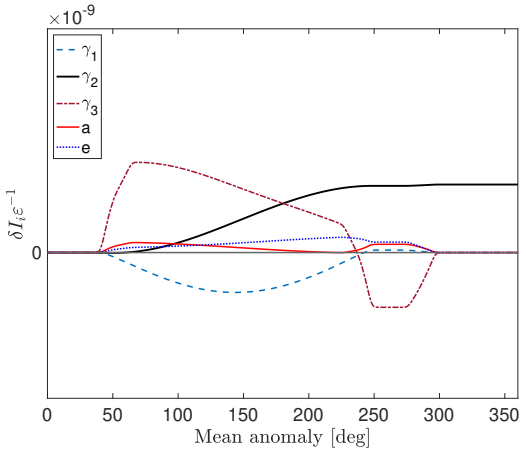
(b) Solution of the optimal control problem (**??**) on a real control set as projection on $\hat{s}$, $\hat{s}^\perp$.



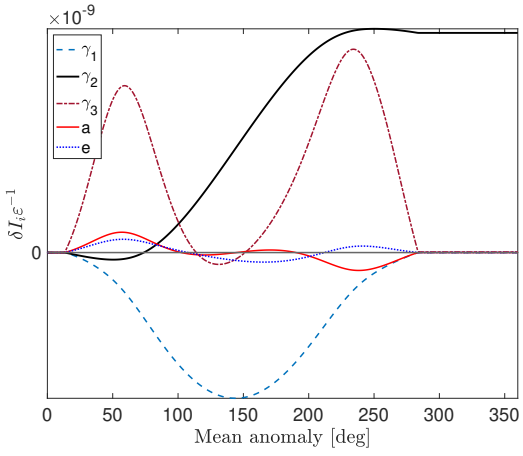
(c) Solution of the optimisation problem (29) on a bounded convex cone as orientation angles.



(d) Solution of the optimal control problem (**??**) on a real control set as orientation angles.



(e) Trajectory of the sail with controls obtained with convex programming.

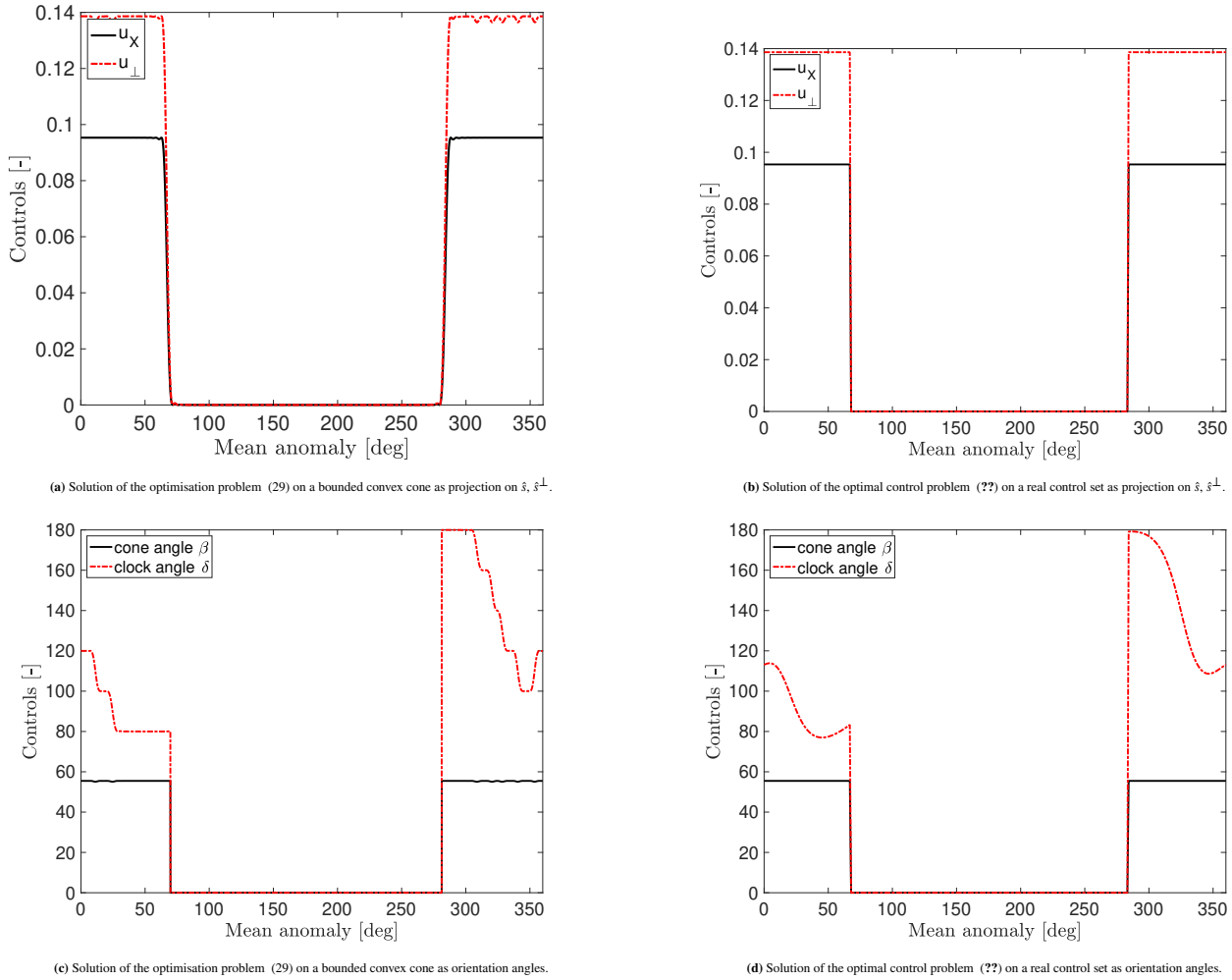


(f) Trajectory of the sail with controls obtained with OCP.

Figure 6 Solutions of two problems with $d_I = (0, 1, 0, 0, 0)$, $I = (10°, 50°, 30°, 1, 0.1)$.

5. Dachwald B, Macdonald M, McInnes CR, Mengali G, Quarta AA. Impact of Optical Degradation on Solar Sail Mission Performance. *Journal of Spacecraft and Rockets* 2007; 44(4): 740–749. doi: 10.2514/1.21432

6. Niccolai L, Quarta AA, Mengali G. Trajectory Approximation of a Solar Sail With Constant Pitch Angle and Optical Degradation. *IEEE Transactions on Aerospace and Electronic Systems* 2022; 58(4): 3643–3649. doi: 10.1109/taes.2021.3124867



(a) Solution of the optimisation problem (29) on a bounded convex cone as projection on $\hat{s}, \hat{s}^\perp$.

(b) Solution of the optimal control problem (**??**) on a real control set as projection on $\hat{s}, \hat{s}^\perp$.

(c) Solution of the optimisation problem (29) on a bounded convex cone as orientation angles.

(d) Solution of the optimal control problem (**??**) on a real control set as orientation angles.

Figure 7 Explanation of a stairs-shaped solution of convex programming.

7. Caillau JB, Daoud B. Minimum time control of the restricted three-body problem. *SIAM J. Control Optim.* 2012; 50(6): 3178-3202.

8. Caillau JB, Féjoz J, Orieux M, Roussarie R. On singularities of min time affine control systems. *SIAM J. Control Optim.* 2022; 60(2): 1143-1162.

9. Mengali G, Quarta AA. Optimal Three-Dimensional Interplanetary Rendezvous Using Non-Ideal Solar Sail. *Journal of Guidance, Control, and Dynamics* 2005; 28(1): 173-177. doi: 10.2514/1.8325

10. Boyd JP. Computing the zeros, maxima and inflection points of Chebyshev, Legendre and Fourier series: solving transcendental equations by spectral interpolation and polynomial rootfinding. *J Eng Math* 2007; 56(3): 203–219. doi: 10.1007/s10665-006-9087-5

11. Herasimenka A, Caillau JB, Dell'Elce L, Pomet JB. Controllability Test for Systems with Constrained Control. Application to Solar Sailing. In: Inria. ; 2022.

12. Nesterov Y. Squared Functional Systems and Optimization Problems. In: Pardalos PM, Hearn D, Frenk H, Roos K, Terlaky T, Zhang S., eds. *High Performance Optimization*. 33. Boston, MA: Springer US. 2000 (pp. 405–440). Series Title: Applied Optimization

13. Gergaud J, Haberkorn T. Homotopy method for minimum consumption orbit transfer problem. *ESAIM Control Optim. and Calc. Var.* 2006; 12(2): 294-310.

14. Zhu J, Trélat E, Cerf M. Geometric Optimal Control and Applications to Aerospace. *Pacific Journal of Mathematics for Industry* 2017; 9(8). doi: 10.1186/s40736-017-0033-4

15. Caillau JB, Cots O, Gergaud J. Differential pathfollowing for regular optimal control problems. *Optim. Methods Softw.* 2012; 27(2): 177-196.

AUTHOR BIOGRAPHY

empty-eps-converted-to.pdf

Author Name. This is sample author biography text this is sample author biography text.

How to cite this article: Williams K., B. Hoskins, R. Lee, G. Masato, and T. Woollings (2016), A regime analysis of Atlantic winter jet variability applied to evaluate HadGEM3-GC2, *Q.J.R. Meteorol. Soc.*, *2017;00:1–6*.