

## Momentum-Based Learning of Nash Equilibria for LISA Pointing Acquisition

Gomez, Aitor Ramirez; Ahdab, Mohamad Al

*Published in:*  
IFAC-PapersOnLine

*DOI (link to publication from Publisher):*  
[10.1016/j.ifacol.2023.10.646](https://doi.org/10.1016/j.ifacol.2023.10.646)

*Creative Commons License*  
CC BY-NC-ND 4.0

*Publication date:*  
2023

*Document Version*  
Publisher's PDF, also known as Version of record

[Link to publication from Aalborg University](#)

*Citation for published version (APA):*  
Gomez, A. R., & Ahdab, M. A. (2023). Momentum-Based Learning of Nash Equilibria for LISA Pointing Acquisition. *IFAC-PapersOnLine*, 56(2), 6012-6017. <https://doi.org/10.1016/j.ifacol.2023.10.646>

### General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal -

### Take down policy

If you believe that this document breaches copyright please contact us at [vbn@aub.aau.dk](mailto:vbn@aub.aau.dk) providing details, and we will remove access to the work immediately and investigate your claim.

# Momentum-Based Learning of Nash Equilibria for LISA Pointing Acquisition

Aitor R. Gomez\* Mohamad Al Ahadab\*

\* Aalborg University, Aalborg, Denmark

(e-mail: {arg,maah}@es.aau.dk).

**Abstract:** This paper addresses the pointing acquisition phase of the Laser Interferometer Space Antenna (LISA) mission as a guidance problem. It is formulated in a cooperative game setup, which solution is a sequence of corrections that can be used as a tracking reference to align all the spacecraft' laser beams simultaneously within the tolerances required for gravitational wave detection. We propose a model-free learning algorithm based on residual-feedback and momentum, for accelerated convergence to stable solutions, i.e. Nash Equilibria. Each spacecraft has 4 degrees of freedom, and the only measured output considered are laser misalignments with the local interferometer sensors. Simulation results demonstrate that the proposed strategy manages to achieve absolute misalignment errors  $< 1\mu\text{rad}$  in a timely manner.

Copyright © 2023 The Authors. This is an open access article under the CC BY-NC-ND license (<https://creativecommons.org/licenses/by-nc-nd/4.0/>)

**Keywords:** Space exploration, Decision making and autonomy, High accuracy pointing, Extremum seeking and model free adaptive control, Game theory, LISA, Satellite constellation.

## 1. INTRODUCTION

The Laser Interferometer Space Antenna (LISA) is a highly complex space mission from NASA and ESA, aiming at detecting and characterizing gravitational waves (GWs). It is the result of the coordinated efforts of three spacecraft maintaining frequency-stabilized laser links between each other, resulting in an equilateral triangle formation as illustrated in Fig. 1. Detection of GWs is enabled by the monitoring of frequency fluctuations of the lasers, produced by space-time ripples altering the distance between proof masses –the interferometer arm length– located inside each spacecraft. The barycentre of the formation follows an heliocentric orbit lying in the ecliptic plane, with radius of 1AU and 20 degrees behind the Earth, and the interferometer arm length is  $5 \cdot 10^6 \text{ km}$ . This design has been meticulously chosen to allow for observations of gravitational radiation within a spectrum not reachable by its ground-based predecessor, the Laser Interferometer Gravitational-wave Observatory (LIGO). While LIGO is capable of detecting frequencies that are limited between 10 Hz and 10 kHz, LISA has the potential of complementing those observations by detecting GWs with frequencies ranging between  $10^{-4}$  Hz and 1 Hz.

The detection mode of GWs –or *science mode*– is the last of several intricate phases of the mission, most of which present very demanding requirements. The one studied in this paper is referred to as *Pointing Acquisition and Tracking* (PAT) mode, namely, the task of aligning the local interferometer detectors with the lasers received from the other two spacecraft companions, while satisfying the remarkably low pointing tolerances required to engage science mode. To that end, every spacecraft is equipped with two identical optical assemblies (OAs) as payload, each of which contain, among others, the proof mass, a telescope, a laser source and laser interferometer detectors. The OAs have the telescopes pointing in a particular configuration

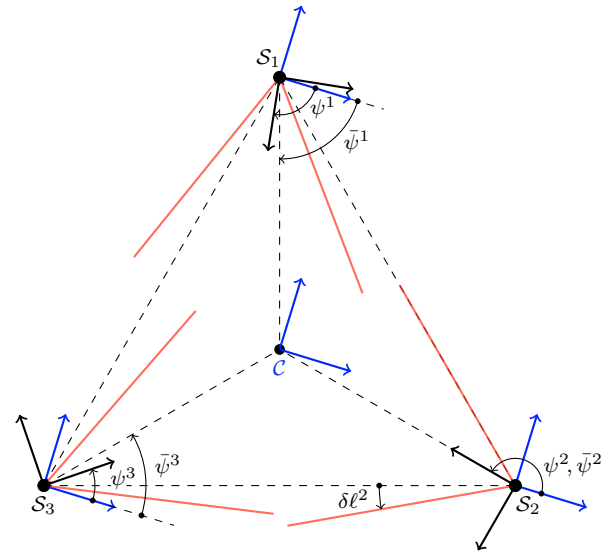


Fig. 1. 2-D geometry of the problem with overemphasized angles. It illustrates the constellation –inertial– frame (blue), and the spacecraft' moving frames (black). The red lines represent the directions of the laser beams,  $\psi^i$  the global state (yaw) and  $\bar{\psi}^i$  the target global state (target yaw) of spacecraft  $S_i$ .

that makes both of the lasers emitted by a spacecraft form an angle of 60 degrees by default. To account for the small drifting of the spacecraft throughout the heliocentric orbit that slightly deforms the equilateral formation, one of the telescopes is assumed to have an extra degree of freedom (see  $\delta l^i$  in Fig. 2). The laser interferometer detectors consist of a charge-coupled device (CCD) and a quadrant photodiode (QPD); the former used for signal acquisition with a field-of-view of  $155\mu\text{rad}$ , and the latter for fine pointing with a field-of-view of  $1\mu\text{rad}$ . Precise measurements of the misalignments between the local and

the received laser beams can be determined based on these sensor readings by employing the well-known method of differential wave-front sensing (Heinzel et al. (2004)).

A preliminary pointing phase precedes PAT mode, in which Star Trackers (STRs) are first utilized to point the spacecraft. Due to multiple sources of uncertainty, the final pointing error of this phase is estimated in the LISA (2000) technical report to be within a  $9\mu\text{rad}$  half-cone. Since the nominal laser beam is spanned by a smaller cone, two main strategies are usually considered (Diehl, 2009) to deal with this uncertainty cone (UC) in the PAT phase: 1) defocusing of the laser beam to cover the UC and 2) scanning of the UC. Maghami et al. (2005) developed an attitude control policy based on beam-defocusing, achieving the desired pointing accuracy, however, with no specification of time duration. Cirillo and Gath (2009), on the other hand, studied the scanning method and developed a control strategy to establish the laser links within a 70 min interval, approximately, for a  $15\mu\text{rad}$  UC. Albeit, both strategies perform one link acquisition at a time.

In this paper, we propose an algorithm based on beam-defocusing for a simultaneous laser acquisition and fine-pointing, showing that fast and accurate convergence can potentially be achieved. As stated in the internal report LISA (2009), defocus of the laser beam is, additionally, considered to be more robust than scanning. Note that simultaneity of these operations imply that all spacecraft must use their lasers and detectors at the same time. While the detection of the received laser is deemed nonviable if the local lasers are always on, this inconvenient could be avoided by emitting pulses, rather than a continuous laser wave. The repetition rate must be chosen to ensure receiving a pulse when the local laser is not emitting, and part of the travel time that each pulse takes to reach the opposite spacecraft –16 seconds for a  $5 \cdot 10^6 \text{km}$  interferometer arm length– can be used to build up energy between pulses. At every detection of a pulse, the strategy described in this paper can be adopted to guide each spacecraft in unison. Alternatively, this methodology has the advantage of being combinable, by performing an initial scanning step to further reduce the UC and posteriorly defocusing the laser, or just as a high-precision pointing algorithm.

### 1.1 Contribution

The main contribution is Algorithm 1, designed to provide a tracking reference to the spacecraft' controllers for a fast, simultaneous and accurate alignment of the laser beams into the interferometer detectors. It is based on the work by Tatarenko and Kamgarpour (2019) with the addition of two extra terms: the so-called momentum (Qian, 1999) and residual-feedback (Zhang et al., 2022). We propose to include these in the context of game theory in order to improve the learning rate of equilibria.

We show the performance of Algorithm 1 for the laser acquisition and pointing problem in the LISA mission introduced in the previous section. The problem has been formulated as a cooperative game enabled by the suggested pulsed laser strategy, and successfully solved for 1000 realizations with different initial conditions.

### 1.2 Notation

Let  $\|\mathbf{x}\| := \sqrt{\mathbf{x}^\top \mathbf{x}}$  be the Euclidean norm of a vector  $\mathbf{x}$ , and define the closed ball of dimension  $d$  and radius  $r$  as  $B_r^d := \{\mathbf{x} \in \mathbb{R}^d : \|\mathbf{x}\| \leq r\}$ , with  $\mathcal{U}(B_d)$  the uniform distribution over  $B_r^d$ . Similarly, let  $S_r^{d-1} := \{\mathbf{x} \in \mathbb{R}^d : \|\mathbf{x}\| = r\}$  denote the sphere of radius  $r$ , and  $\mathcal{U}(S_r^{d-1})$  the uniform distribution over  $S_r^{d-1}$ . For an arbitrary compact set  $\Theta$ , we define a projection operator  $\Pi_\Theta(\mathbf{x}) := \operatorname{argmin}_{\theta \in \Theta} \|\theta - \mathbf{x}\|$ . For a vector  $\mathbf{x} \in \mathbb{R}^n$  we write  $\mathbf{x}^{-i} := [x^1, \dots, x^{i-1}, x^{i+1}, \dots, x^n]^\top$ . Consider an inertial frame  $\mathcal{C} := \{c, \hat{i}_X, \hat{i}_Y, \hat{i}_Z\}$  fixed in the barycentre of the constellation, and three moving spacecraft frames  $\mathcal{S}_i := \{s_i, \hat{i}_x, \hat{i}_y, \hat{i}_z\}$ , where  $i \in I_s = \{1, 2, 3\}$ , fixed in each spacecraft. We chose  $\hat{i}_x$  pointing in the bisector of the two local lasers in their default configuration,  $\hat{i}_z$  to be orthogonal to the plane spanned by the axes of the local lasers, and  $\hat{i}_y$  completing the right-hand-rule.

## 2. PROBLEM FORMULATION

An essential aspect of this problem is that the misalignments between laser beams are in the order of  $\mu\text{rad}$ – $\text{mrad}$ . This fact can lead to numerical complications if the problem is not formulated properly. Employing the CCDs and QPDs as sensing devices is not only motivated by the high-precision requirements, but also by the fact that it enables the derivation of a *relative* formulation that, in turn, will allow us to center our states around zero, reducing the scale of the units to  $\mu\text{rad}$  and avoiding numerical issues.

### 2.1 System Model

We start by considering three spacecraft with local moving frames  $\mathcal{S}_i$ , where  $i \in I_s = \{1, 2, 3\}$ , in triangular formation as represented in Fig. 1. Each spacecraft has a state vector  $\mathbf{x}^i := \mathbf{x}^i(t) = [\alpha^i(t) \ \varphi^i(t) \ \psi^i(t) \ \ell^i(t)]^\top \in \mathbb{R}^4$  the first three elements denoting Euler angles that describe the inertial orientation of  $\mathcal{S}_i$ , and an element  $\ell^i(t)$  describing the angle between both local lasers. For convenience, we consider the constellation frame  $\mathcal{C}$  as inertial. It follows naturally to define the initial conditions of the spacecraft  $\mathbf{x}_0^i := \mathbf{x}^i(0) = [\alpha_0^i \ \varphi_0^i \ \psi_0^i \ \ell_0^i]^\top \in \mathbb{R}^4$ , at the beginning of the laser acquisition phase. In order to adhere to a relative state formulation, the following relation becomes practical.

$$\mathbf{x}^i = \mathbf{x}_0^i + \delta \mathbf{x}^i, \quad i \in I_s. \quad (1)$$

where  $\delta \mathbf{x}^i := \delta \mathbf{x}^i(t) = [\delta \alpha^i(t) \ \delta \varphi^i(t) \ \delta \psi^i(t) \ \delta \ell^i(t)]^\top \in \mathbb{R}^4$  are small deviations from the initial condition. Let us group the three relative state vectors into a full state vector  $\delta \mathbf{x}(t) = [\delta \mathbf{x}^{1\top} \ \delta \mathbf{x}^{2\top} \ \delta \mathbf{x}^{3\top}]^\top$ , or just  $\delta \mathbf{x}$ , for compactness. Determining a full model for  $\delta \dot{\mathbf{x}}$  lies out of the scope of this paper. We will assume, nonetheless, that each spacecraft has its own internal controller with actuation resolution in the order of nanoradians. The spacecraft' states will then evolve according to their closed-loop dynamics denoted by the function  $f$ ,

$$\delta \dot{\mathbf{x}} = f(\delta \mathbf{x}, \mathbf{u}_{k-1}), \quad t \in [t_{k-1}, t_k]. \quad (2)$$

which has piecewise-constant reference vectors  $\mathbf{u}_k := \mathbf{u}(t_k) = [\mathbf{u}_k^{1\top} \ \mathbf{u}_k^{2\top} \ \mathbf{u}_k^{3\top}]^\top$ , with  $\mathbf{u}_k^i := \mathbf{u}^i(t_k) \in U^i \subseteq \mathbb{R}^4$ , as inputs. Using a constraint box for the references with  $b \in \mathbb{R}$ , we have that

$$U^i = [-b, b]^4, \quad \forall i \in I_s, \quad (3)$$

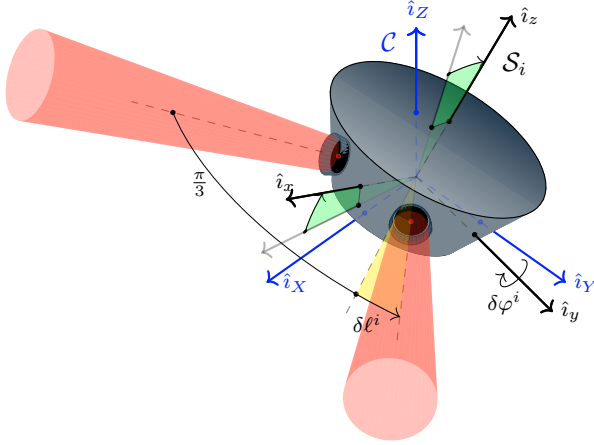


Fig. 2. Constellation frame  $\mathcal{C}$  (blue) and the spacecraft frame  $\mathcal{S}_i$  (black). The green angle illustrates  $\delta\varphi^i$ , i.e. an attitude change wrt its initial attitude (gray), and the yellow angle represents  $\delta\ell^i$ , a change in the telescope position wrt its default 60deg configuration.

Considering the system behaviour under the closed-loop dynamics presented above, we make the following fair assumption about the internal controller.

**Assumption 2.1.** Taking the system (2), we assume that the states  $\delta\mathbf{x}$  at time  $t \in [t_{k-1}, t_k]$  will converge to the reference  $\mathbf{u}_{k-1}$  with a constant time  $\tau_c \in \mathbb{R}_{\geq 0}$  such that,

$$\lim_{t \rightarrow t_k} \|\delta\mathbf{x}(t) - \mathbf{u}_{k-1}\| \leq \exp\left(-\frac{(t - t_k)}{\tau_c}\right). \quad (4)$$

Additionally, we build the guidance system so that it provides references at a slower pace than the closed-loop convergence rate (4), which supports the next assumption.

**Assumption 2.2.** Let the guidance system provide reference samples every time period of  $\tau_g := t_k - t_{k-1} \in \mathbb{R}_{\geq 0}$  with  $\tau_g \gg \tau_c$ . Then, if Assumption 2.1 holds, it follows that

$$\|\delta\mathbf{x}(t_k) - \mathbf{u}_{k-1}\| \leq \exp\left(-\frac{\tau_g}{\tau_c}\right), \quad (5)$$

which indicates that  $\delta\mathbf{x}_k \approx \mathbf{u}_{k-1}$ , where  $\delta\mathbf{x}_k := \delta\mathbf{x}(t_k)$ .

## 2.2 Objective Functions

Let us introduce the objective functions  $h^i: \mathcal{U} \rightarrow \mathbb{R}_{\geq 0}$  of  $\mathbf{u}$  that each spacecraft  $\mathcal{S}_i$  aims to minimize, where  $\mathcal{U} = \mathcal{U}^1 \times \mathcal{U}^2 \times \mathcal{U}^3$  denotes the set of joint references. Each function  $h^i$  penalizes the total misalignment measured by both detectors on each spacecraft. Provided that all spacecraft receive two beam pulses, then  $h^i$  is designed to be zero only when the misalignment between both local sensors and their received beams is also zero. We define another function  $y_{ij}: \mathcal{U}^i \times \mathcal{U}^j \rightarrow \mathbb{R}_{\geq 0}$  as a model for the individual misalignment measured by each individual sensor. Given the relation between  $\delta\mathbf{x}_k$  and  $\mathbf{u}_{k-1}$  from Assumption 2.2, we propose the objective function to be

$$h^i(\mathbf{u}_{k-1}) = \sum_{j \in I_s^{ij}} w y_{ij}^2(\mathbf{u}_{k-1}^i, \mathbf{u}_{k-1}^j), \quad \forall i \in I_s. \quad (6)$$

We normalize the objectives with a scalar weight  $w \in \mathbb{R}$  using the worst achievable value of  $h^i$  to avoid numerical issues. Herein, the worst misalignment corresponds to the size of the UC, i.e.  $9\mu\text{rad}$ .

Note that this setup is clearly a cooperative game that will be denoted as  $\Gamma(3, \{h^i\}, \{U^i\})$ , given that each spacecraft can be deemed an agent trying to minimize a common objective, i.e. the sum of all the individual objectives  $h^i$ . However, any arbitrary spacecraft  $\mathcal{S}_i$  can only observe its local objective  $h^i$  by means of its own sensors. Moreover, any reference  $\mathbf{u}^i \in U^i$  selected by this spacecraft will have an impact on the objectives of the other two spacecraft.

We proceed to determine an expression for  $y_{ij}$  that rely first on the relative states  $\delta\mathbf{x}$ , and later on  $\mathbf{u}$ . To that end, let  $\bar{\mathbf{x}}^i \in \mathbb{R}^4$  be some desired inertial state to be attained by each spacecraft and its maneuverable telescope, for which all the lasers are perfectly aligned.

**Remark 2.1.** Note that the target attitude vectors  $\bar{\mathbf{x}}$  depend on the actual position and orientation of all the spacecrafts. Errors in these quantities are unknown to each spacecraft and yield to non-equilateral triangle formations.

It follows to define the initial error,  $\delta\mathbf{x}_0^i \in \mathbb{R}^4$ , of spacecraft  $\mathcal{S}_i$  from its target  $\bar{\mathbf{x}}^i$ .

$$\delta\mathbf{x}_0^i = \bar{\mathbf{x}}^i - \mathbf{x}_0^i. \quad (7)$$

It is evident, then, that both quantities,  $\bar{\mathbf{x}}^i$  and  $\delta\mathbf{x}_0^i$ , are unknowns. Additionally, notice that in order to achieve such a triangular formation, the difference between the relative states of any pair of spacecraft  $\mathcal{S}_i$  and  $\mathcal{S}_j$ , with  $(i, j) \in \{i, j \in I_s : i \neq j\}$ , has to result in a constant, but unknown, angle vector  $\bar{\mathbf{x}}_{ij} \in \mathbb{R}^4$ .

$$\bar{\mathbf{x}}_{ij} = \bar{\mathbf{x}}^i - \bar{\mathbf{x}}^j, \quad (8)$$

which value depends on the spacecraft pair involved.

**Remark 2.2.** Since the triangular formation might not be equilateral, see Remark 2.1, the relative vector  $\bar{\mathbf{x}}_{ij}$  is unknown.

Based on the previous equations, we propose the following function that models the misalignment of the beam emitted by  $\mathcal{S}_j$  and measured by  $\mathcal{S}_i$  at discrete time  $t_k$ .

$$y_{ij}(\mathbf{x}_k^i, \mathbf{x}_k^j) = \|(\mathbf{x}_k^i - \mathbf{x}_k^j) - \bar{\mathbf{x}}_{ij}\|, \quad (9)$$

As expected, the measured misalignment  $y_{ij}$  is zero only when the difference between the inertial states of the spacecraft pair is equal to the target vector  $\bar{\mathbf{x}}_{ij}$ , and positive otherwise. Saturation on  $y_{ij}$  when it exceeds  $9\mu\text{rad}$  is disregarded, since it would be required to switch to another guidance scheme governed by other sensors if the states were to overshoot outside of the UC. We continue to modify (9) due to its dependency on inertial variables. Replacing (1), (7) and (8) to express  $y_{ij}$  in terms of relative states, leads to the final measurement equation,

$$y_{ij}(\delta\mathbf{x}_k^i, \delta\mathbf{x}_k^j) = \|(\delta\mathbf{x}_k^i - \delta\mathbf{x}_k^j) - (\delta\mathbf{x}_0^i - \delta\mathbf{x}_0^j)\|,$$

which according to Assumption 2.2,

$$\begin{aligned} y_{ij}(\delta\mathbf{x}_k^i, \delta\mathbf{x}_k^j) &\approx y_{ij}(\mathbf{u}_{k-1}^i, \mathbf{u}_{k-1}^j) \\ &= \|(\mathbf{u}_{k-1}^i - \mathbf{u}_{k-1}^j) - (\delta\mathbf{x}_0^i - \delta\mathbf{x}_0^j)\|, \end{aligned} \quad (10)$$

Observe that, in practice, we only obtain values for the measurement  $y_{ij}$ , and the model (10) is just intended for simulation purposes.

This concludes the description of the game  $\Gamma(3, \{h^i\}, \{U^i\})$ .

### 3. EXTREMUM SEEKING

Throughout this section, we will use the words spacecraft and players indistinctly. First, we will show that the game  $\Gamma(3, \{h^i\}, \{U^i\})$  has at least one *Nash Equilibrium* (NE), i.e. a stable solution  $\mathbf{u}^*$  which satisfies

$$h^i(\mathbf{u}^{i*}, \mathbf{u}^{-i*}) \leq h^i(\mathbf{u}^i, \mathbf{u}^{-i*}), \quad \mathbf{u}^i \in U^i, \quad \forall i \in I_s. \quad (11)$$

Afterwards, we will present a payoff-based strategy in which each spacecraft  $\mathcal{S}_i$  at time  $t_k$  computes an update  $\mathbf{u}_k^i$  based on their most recent reference,  $\mathbf{u}_{k-1}^i$ , and the value of their objective,  $h^i(\mathbf{u}_{k-1})$ .

#### 3.1 Existence of Nash Equilibria

To show the existence of a NE in the game  $\Gamma(3, \{h^i\}, \{U^i\})$ , we first put forward the following theorem.

**Theorem 1.** (Proposition 2.2.9 in Facchinei and Pang (2003)) *Consider a game  $\Gamma(N, \{q^i\}, \{A^i\})$  with  $N \in \mathbb{Z}_{>0}$ . Let  $\mathbf{a}^i \in A^i$ , with  $A^i \subset \mathbb{R}^{n_i}$  be compact and convex, and  $q^i$  be continuously differentiable. If each function  $q^i(\mathbf{a}^i, \mathbf{a}^{-i})$  is convex in  $\mathbf{a}^i$  for every fixed  $\mathbf{a}^{-i}$ , then the set of Nash equilibria is nonempty and compact.*

In the game  $\Gamma(3, \{h^i\}, \{U^i\})$  studied in this paper, we have  $U^i \subset \mathbb{R}^4$ , defined as (3), being convex and compact. Furthermore, all functions  $h^i$  are continuously differentiable, and  $\forall i \in I_s$  it holds that  $h^i(\mathbf{u}^i, \mathbf{u}^{-i})$  is convex in  $\mathbf{u}^i$  for a fixed  $\mathbf{u}^{-i}$ , with  $y_{ij}$  modeled as in (10). Therefore, Theorem 1 states that the set of Nash equilibria for the game  $\Gamma(3, \{h^i\}, \{U^i\})$  is nonempty and compact.

#### 3.2 Learning of Nash Equilibria

In order to seek a NE, we propose a strategy inspired by the work of Tatarenko and Kamgarpour (2019). In their work, a gradient estimate of the objectives of each player is obtained using a randomization technique. We assume that at time  $t_k$  each spacecraft randomly selects –or *mixes*– their strategy according to the following

$$\mathbf{u}_k^i = \mu_k^i + r_k \zeta_k^i, \quad \zeta_k^i \sim \mathcal{U}(S_1^3), \quad (12)$$

where  $\mu_k^i \in \mathbb{R}^4$  is the mean values of the reference  $\mathbf{u}_k^i$ , with an update equation that will be discussed later in this section, and  $r_k \in \mathbb{R}$  is a time varying exploration parameter. Afterwards, each spacecraft obtains a gradient estimate  $\mathbf{g}_k^i \in \mathbb{R}^4$  based on the residual-feedback method introduced first in Zhang et al. (2022) as follows

$$\mathbf{g}_k^i = \frac{4}{r_k} (h^i(\mathbf{u}_k) - h^i(\mathbf{u}_{k-1})) \zeta_k^i. \quad (13)$$

One of the key properties of  $\mathbf{g}_k^i$  is the fact that

$$\mathbb{E}_{\zeta_k^i}[\mathbf{g}_k^i] = \frac{\partial \tilde{h}^i(\mu_k)}{\partial \mu_k^i},$$

where  $\tilde{h}^i$  is a smooth version of  $h^i$  obtained by averaging the function  $h^i$  with a uniform random variable  $\xi_k$  over the unit ball. That is

$$\tilde{h}^i(\mu_k) := \mathbb{E}_{\xi_k}[\tilde{h}^i(\mu_k + r_k \xi_k)], \quad \xi_k \sim \mathcal{U}(B_1^{12}),$$

which is a standard result in zeroth-order optimization (see Chen et al. (2022); Flaxman et al. (2005)). In other words, the value  $\mathbf{g}_k^i$  is an unbiased estimator for the gradient of the smoothed version  $\tilde{h}^i$ . Using two function evaluations for

the estimator  $\mathbf{g}_k^i$  in (13) was shown by Zhang et al. (2022) to reduce the variance of the estimates, which provides faster convergence than the case of using one function evaluation as done by Tatarenko and Kamgarpour (2019).

After introducing the gradient estimate  $\mathbf{g}_k^i$ , the reference  $\mathbf{u}_k^i$  for each spacecraft is calculated using (12) in which  $\mu_k^i$  is determined according to

$$\mu_k^i = \Pi_{\tilde{U}_k^i}(\mu_{k-1}^i - \eta_{k-1}^i \mathbf{g}_{k-1}^i + \rho(\mu_{k-1}^i - \mu_{k-2}^i)), \quad (14)$$

where  $\eta_k \in \mathbb{R}_{>0}$  is a time-varying step size and the term  $\rho(\mu_{k-1}^i - \mu_{k-2}^i)$  represents a momentum term with parameter  $\rho \in [0, 1)$ . Our introduction of the momentum term in (14) leads to a faster convergence for the seeking scheme to a NE. See Qian (1999) for additional details regarding momentum in gradient descent algorithms. The set  $\tilde{U}_k^i = [-b_k^i, b_k^i]^4$  in (14) is time-dependent, with bounds

$$b_k^i = \beta^k b_0^i + (1 - \beta^k) b, \quad 0 \leq \beta < 1,$$

where  $b_0 \in \mathbb{R}$  is the initial size chosen for the box constraint and  $b$  the final, coinciding with the original constraint in (3). This time dependency on the box constraints is introduced to help avoiding violent changes in the references at the beginning of the phase, which can cause overshooting outside the UC. Note that we also propose a uniform distribution for  $\zeta_k^i$ , instead of a normal distribution as done by Tatarenko and Kamgarpour (2019), since it has a finite support as opposed to the normal distribution. This ensures the boundedness of  $\mathbf{u}$  which is more suitable for the restriction (3) to hold. The summary of the seeking strategy used by each spacecraft  $\mathcal{S}_i$  is provided in Algorithm 1.

**Remark 3.1.** *It is easy to check that  $\Gamma(3, \{h^i\}, \{U^i\})$  is a convex and monotone game. The structure present by  $n$ -player convex and monotone games rules out the existence of non-Nash equilibria (Rosen (1965)) and thus, a gradient-play strategy will converge to a NE (Mazumdar et al. (2020)). Moreover, gradient-bandit strategies have been proposed and shown to converge to a NE in monotone games, see Tatarenko and Kamgarpour (2019) and Gao and Pavel (2022). In this paper, we demonstrate empirical convergence of our gradient-bandit algorithm to a NE. Yet, a more comprehensive theoretical analysis and proof are necessary for a complete understanding of the algorithm's behavior.*

---

#### Algorithm 1: NE seeking for spacecraft $\mathcal{S}_i, i \in I_s$

---

**Parameters:**  $\gamma_r > 0$ ,  $a_r > 0$  such that  $r_k = \gamma_r/(k)^{a_r}$ ,  
 $\gamma_\eta > 0$ ,  $a_\eta > 0$  such that  $\eta_k = \gamma_\eta/(k)^{a_\eta}$ ,  
 $0 \leq \rho < 1$ ,  $0 \leq \beta < 1$ , and  $\varepsilon > 0$ .

**Input:** initial  $\mu_0^i = \mu_{-1}^i = \mathbf{0}$ .

**Output:**  $\mathbf{u}_k^i$

```

1  $k = 0$ 
2 while  $\|h^i(\mathbf{u}_k) - h^i(\mathbf{u}_{k-1})\| \geq \varepsilon \mathbf{do}$ 
3    $k \leftarrow k + 1$ ,
4   Obtain the measurements  $y_{ij}$ ,  $\forall j \in I_s^{ij}$  and compute
      $h^i(\mathbf{u}_{k-1})$  from (6).
5   Compute  $\mathbf{g}_{k-1}^i$  using (13).
6   Compute  $\mu_k^i$  using (14).
7   Compute  $\mathbf{u}_k^i$  using (12).
8 Set  $\mathbf{u}_k^i = \mu_k^i$ .
```

---

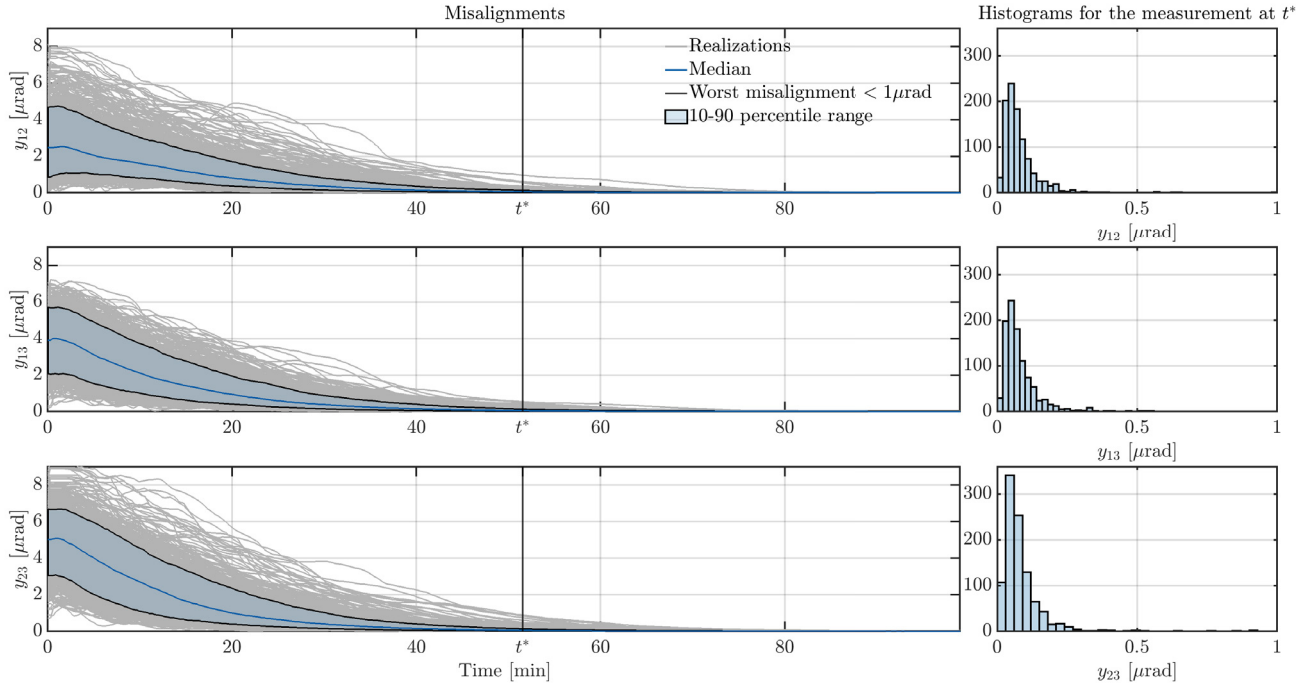


Fig. 3. Simulation of 1000 realizations with different initial conditions inside the UC. Each realization is simulated for 5000 iterations, and  $t^*$  marks the time at which the worst realization reaches the field of view of the QPD, i.e. all misalignments are below  $1\mu\text{rad}$ . The histograms illustrate the misalignments measured at time  $t^*$ .

#### 4. SIMULATION RESULTS

The following is a description of the simulation setup<sup>1</sup> to test the performance of Algorithm 1.

All three spacecraft lie inside an UC of  $9\mu\text{rad}$ , and thus we randomize the initial conditions  $\delta\mathbf{x}_0$  such that between any spacecraft pair it holds that  $\|\delta\mathbf{x}_0^i - \delta\mathbf{x}_0^j\| \leq 9\mu\text{rad}$ . In accordance with Assumption 2.1, the states have been propagated using a first-order system

$$\dot{\delta\mathbf{x}}(t) = \frac{1}{\tau_c}(\mathbf{u}_{k-1} - \delta\mathbf{x}(t)), \quad t \in [t_{k-1}, t_k]. \quad (15)$$

The convergence time of the controller is defined to be  $\tau_c = 1\text{s}$ . Given that optical measurements are spaced in time by 16 seconds—due to the interferometer arm length—then the guidance system can provide reference updates every  $\tau_g = \tau_c + 16 = 17\text{s}$ . For this time rate, it is fair to consider that  $\tau_g \gg \tau_c$  and, thus, that Assumption 2.2 also holds.

A total of 1000 realizations have been drawn for the initial conditions  $\delta\mathbf{x}_0$ . Instead of picking a value for the stopping parameter  $\varepsilon$  in line 2 of the algorithm, each of the initial states have been numerically propagated for  $K = 5000$  iterations using (15). At each iteration  $k$ , Algorithm 1 is used to update the references  $\mathbf{u}_k$ , corresponding to time  $t_k = k\tau_g$ . For convenience, a time increment  $\Delta t = \tau_g$  has been chosen, so that a new reference update is computed at each iteration. Equivalently, we have  $\Delta t = \tau_g = 1$  and  $\tau_c = 1/\Delta t = 1/17$ . The initial bound for the time-

dependent constraint  $\tilde{U}_k^i$  is calculated from the initial conditions,  $\mathbf{u}_0 = \mathbf{0}$ , as

$$b_0^i = 0.3(1 - h^i(\mathbf{0})).$$

The remaining parameters required for the algorithm are selected as follow:  $a_r = 0.2$ ,  $a_\eta = 0.5$ ,  $\gamma_r = 0.58$ ,  $\gamma_\eta = 4.5$ ,  $\rho = 0.93$ ,  $\beta = 0.01$ ,  $w = 1/9^2$  and  $b = 4.5$ .

Table 1. Average misalignments at time  $t^*$ , the time when the worst realization is below  $1\mu\text{rad}$ , using Algorithm 1 and Algorithm 1<sup>†</sup>.

Misalignment	Algorithm 1	Algorithm 1 <sup>†</sup>
$y_{12} [\mu\text{rad}]$	0.0794	4.10
$y_{13} [\mu\text{rad}]$	0.0804	4.25
$y_{23} [\mu\text{rad}]$	0.0843	4.54

The results of the Monte Carlo simulation are displayed in two fashion. In Fig. 3, we show the lasers misalignments between spacecraft,  $y_{ij}$ , for each realization. The robustness of the algorithm against different initial conditions is exhibited, by showing the convergence of each trajectory as a function of time, and their settling misalignment at time  $t^*$  in a histogram. Time  $t^*$  correspond to the time at which the worst realization enters the field of view of the QPD, and thus, all misalignments are below  $1\mu\text{rad}$ . We also show the 10-90 percentile range of the realizations in the blue shaded area, and the median in darker blue. Additionally, Table 1 provides details of the mean value of the misalignments at  $t^*$ . Moreover, these values are compared to the results obtained using a version of Algorithm 1 without momentum and residual-feedback terms. We refer to this as Algorithm 1<sup>†</sup>. Note that for such a case, the results are severely degraded in accuracy if simulated for the same period of time.

<sup>1</sup> Code of Algorithm 1 available at <https://github.com/aitor-rg/LISA-pointing-game>



Finally, the time taken to converge inside the field of view of the QPD,  $< 1\mu\text{rad}$ , for the worst case is 51.57min.

## 5. CONCLUSION

This work presents a guidance scheme for the LISA pointing acquisition phase, which commands three spacecraft at once to establish interspacecraft laser links for GWs detection. The guidance scheme is rooted in notions of game theory, which have been extended to deliver the main contribution of the paper, Algorithm 1, i.e. a new algorithm for learning of Nash Equilibria in cooperative games. The convergence of this algorithm to a NE is shown empirically and future work is needed to establish a rigorous proof of the algorithm's convergence properties.

The proposed methodology is based on the idea of defocusing the laser beam which, as opposed to the main scanning strategies in the literature, do not rely on ground communication during the acquisition process. It is a fully autonomous system, and it does not require changes or additional mechanisms. Each spacecraft is assumed to have micro-thrusters providing three rotational degrees of freedom, and an extra degree of freedom is considered for one of the two telescopes. Simulation results show that the algorithm is robust against initial conditions lying inside the uncertainty cone, and the worst misalignment between the laser and the interferometer detectors is  $< 1\mu\text{rad}$ . This means that the laser still is inside the field of view of the quadrant photodiode, which provide good margins of error. The time duration of the pointing acquisition –reaching below  $1\mu\text{rad}$ – for the worst simulated case is 51.57min, below 1 hour.

Moreover, a new idea has been put forward on how to manipulate the lasers during this phase. It is suggested that the lasers should emit pulses, rather than a continuous wave, in order to enable the emission and reception of optical information simultaneously between the three spacecraft. However, more technical work is needed on the laser front to ensure sufficient signal to noise ratio when defocusing the laser beam, as well as when emitting optical pulses. A certain scope for future work is the inclusion of a full dynamical model of the spacecraft with disturbances.

## REFERENCES

- Chen, X., Tang, Y., and Li, N. (2022). Improve single-point zeroth-order optimization using high-pass and low-pass filters. In *International Conference on Machine Learning*, 3603–3620. PMLR.
- Cirillo, F. and Gath, P. (2009). Control system design for the constellation acquisition phase of the lisa mission. iop publ. j. phys. conf. ser. 154.
- Diehl, R. (2009). "LISA Operations Concept Document," LISA Project internal report number LISA-OPS-RP-0001.
- Facchinei, F. and Pang, J.S. (2003). *Finite-dimensional variational inequalities and complementarity problems*. Springer.
- Flaxman, A.D., Kalai, A.T., and McMahan, H.B. (2005). Online convex optimization in the bandit setting: Gradient descent without a gradient. In *Proceedings of the Sixteenth Annual ACM-SIAM Symposium on Discrete Algorithms*, SODA '05, 385–394. Society for Industrial and Applied Mathematics, USA.
- Gao, B. and Pavel, L. (2022). Bandit learning with regularized second-order mirror descent. In *2022 IEEE 61st Conference on Decision and Control (CDC)*, 5731–5738. doi:10.1109/CDC51059.2022.9992906.
- Heinzel, G., Wand, V., García, A., Jennrich, O., Braxmaier, C., Robertson, D., Middleton, K., Hoyland, D., Rüdiger, A., Schilling, R., Johann, U., and Danzmann, K. (2004). The ltp interferometer and phasemeter. *Classical and Quantum Gravity*, 21(5), S581.
- LISA (2000). Final Technical Report of the (Phase A): Study of the Laser Interferometer Space Antenna, Project internal report number ESTEC Contract no.13631/99/NL/MS, Report No. LI-RP-DS-009.
- LISA (2009). Laser Interferometer Space Antenna (LISA) Mission Concept," LISA Project internal report number LISA-PRJ-RP-0001.
- Maghami, P., Hyde, T., and Kim, J. (2005). An acquisition control for the laser interferometer space antenna. *Quantum Gravity*, 22.
- Mazumdar, E., Ratliff, L.J., and Sastry, S.S. (2020). On gradient-based learning in continuous games. *SIAM Journal on Mathematics of Data Science*, 2(1), 103–131.
- Qian, N. (1999). On the momentum term in gradient descent learning algorithms. *Neural networks*, 12(1), 145–151.
- Rosen, J.B. (1965). Existence and uniqueness of equilibrium points for concave n-person games. *Econometrica: Journal of the Econometric Society*, 520–534.
- Tatarenko, T. and Kamgarpour, M. (2019). Learning nash equilibria in monotone games. In *2019 IEEE 58th Conference on Decision and Control (CDC)*, 3104–3109.
- Zhang, Y., Zhou, Y., Ji, K., and Zavlanos, M.M. (2022). A new one-point residual-feedback oracle for black-box learning and control. *Automatica*, 136, 110006.