

# S2 Coursework Report

Keying Song

4th April 2025

**Word Count:** 2994 words

## Abstract

This coursework explores a statistical analysis of the Antikythera mechanism's calendar ring using MCMC methods to infer the real parameters of this ancient artifact. The analysis compares two error models: an isotropic Gaussian model with uniform error in all directions, and a radial-tangential model with direction-dependent errors, which is considered more aligned with data and reality. Analytical gradients are derived for both models and verified using finite difference methods. The point estimates obtained via MLE informed the prior range used in the subsequent Bayesian inference, and the results of both approaches were mutually supportive. NUTS sampling generates posterior distributions that reveal significantly larger errors in the tangential direction compared to the radial direction. Thermodynamic integration confirms the superiority of the anisotropic model with an odds ratio of 0.0515.

## Contents

<b>1</b>	<b>Introduction</b>	ii
<b>2</b>	<b>Data Exploration</b>	ii
<b>3</b>	<b>Hole Position Modelling</b>	ii
<b>4</b>	<b>Two Error Modelling Methods</b>	iv
4.1	Error Decomposition . . . . .	iv
4.2	Isotropic Gaussian Model . . . . .	v
4.3	Radial-Tangential Gaussian Model . . . . .	v
<b>5</b>	<b>Gradient Calculation</b>	v
5.1	Gradient for Isotropic Model . . . . .	v
5.2	Gradient for Anisotropic Model . . . . .	vi
5.3	Correctness check . . . . .	vii
<b>6</b>	<b>Maximum Likelihood Predictions</b>	viii
6.1	Methodology . . . . .	viii
6.2	Results . . . . .	ix
<b>7</b>	<b>Posterior Sampling with NUST</b>	x
7.1	Methodology . . . . .	x
7.2	Results . . . . .	x
<b>8</b>	<b>Model Comparison</b>	xii
8.1	Explanation . . . . .	xii
8.2	Methodology . . . . .	xiii
8.3	Result . . . . .	xiv
<b>References</b>		xv

## 1) Introduction

This report presents a MCMC-based statistical analysis of the Antikythera mechanism's calendar ring, focusing on inferring the original number of holes with the measured data of the fragments from [1], thereby ascertaining whether it functioned as a lunar or solar calendar.

In section 2, I will explore the measured data, visualising the positions of the 81 surviving holes across eight fragmented sections of the ring. In section 3, the mathematical model for the hole positions from [2] will be described, accounting for the displacement and rotation of the fragmented sections. Then, section 4 will introduce two Gaussian-like models, an isotropic model with a single variance parameter and a radial-tangential model with direction-dependent variances, to describe the total errors except for the displacements of fragments. In section 5, I will derive the analytical gradients of the log-likelihood functions for both models and check their correctness with the finite difference method. This is followed by section 6, which will present the maximum likelihood parameter estimates. Subsequently, section 7 will employ NUTS method to sample from the posterior distributions, yielding credible intervals for the model parameters and visualising the predictive distribution for a selected hole. Finally, section 8 will deeply explain the two kind of covariance matrices and conduct a formal comparison between the isotropic and anisotropic models using the thermodynamic integration method.

## 2) Data Exploration

### Problem Statement

- (a) Download the data from Ref [1]. Make a plot showing the measured hole locations  $\mathbf{d}_i \in \mathbb{R}^2$  in the  $x - y$  plane, where  $i$  labels the holes, indicating clearly which holes are in each fractured section of the ring. [4]

By visualising the csv file containing measured coordinates of 81 holes distributed across 8 contiguous arc sections, figure 1 shows these measured hole locations in the  $x - y$  plane, with distinct colours indicating the different fragmented sections.

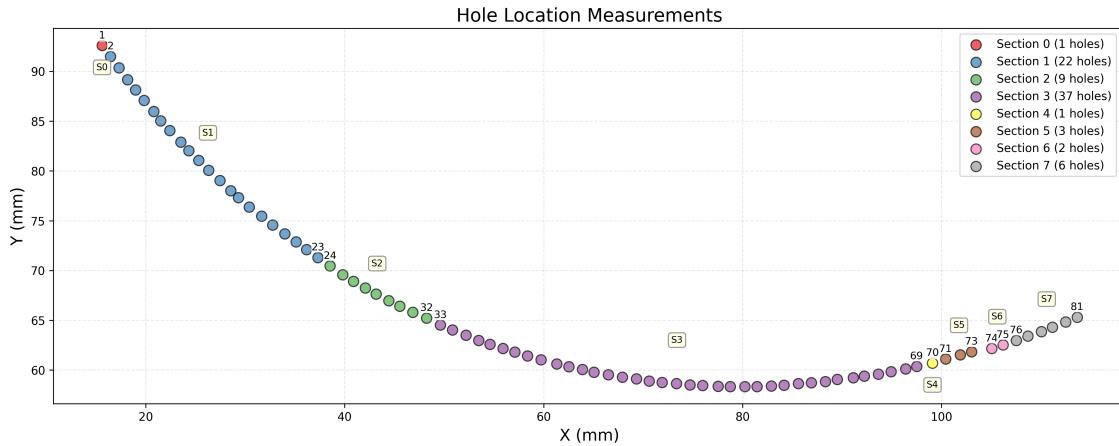


Figure 1: The measured hole position data from Ref [1].

## 3) Hole Position Modelling

### Problem Statement

- (b) Use the model for the hole locations described in section 2 of Ref [2], assume that all parts of the disk lie in the  $x - y$  plane. [5]

In section 2 of Ref [2], the model assumes that the original  $N$  holes sited around a circle of radius  $r$ . Currently, after this circle being fragmented and partial, there are only 81 holes on 8 contiguous arc sections left, with slight displacement and rotation between each section. The notation conventions are summarised in table 1 and the model sketch is shown in figure 2.

Table 1: Notation conventions for the hole position model, the first five are model parameters

Notation	Description
$N$	Number of holes in the full circle
$r$	Radius of the circle model
$\mathbf{r}_{0j} = (x_{0j}, y_{0j})$	The circle-centre of the $j^{th}$ section arcs $j \in \{0, 1, \dots, 7\}$
$\alpha_j$	Representing the degree of rotation of the $j^{th}$ fragment
$\mathbf{C}$	The covariance matrix of measurement error
$S_j$	The $j^{th}$ fragment
$\mathbf{d}_i = (x_i, y_i)$	The measured data of the $i^{th}$ hole position
$\mathbf{m}_{ij} = (x_{ij}, y_{ij})$	The model prediction for the $i^{th}$ hole position set in the $j^{th}$ section

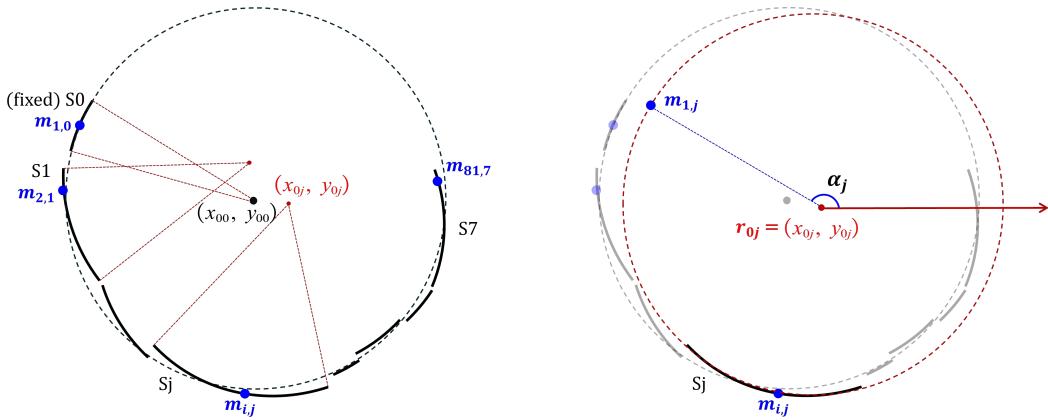


Figure 2: The hole position model in the absolute coordinate system (left) and in a relative coordinate system with respect to  $\mathbf{r}_{0j}$  (right).

For clarity, in the left graph of figure 2, I fix the first section  $S_0$  as the original location to convert the relative displacement and rotation of each fragment into absolute radial displacement and rotation with respect to the original position. Then, two parameters  $(x_{0j}, y_{0j})$  are set to describe the centre of the circle extended from the  $j^{th}$  fragment. However, the total displacement vector  $\mathbf{r}_{0j} - \mathbf{r}_{01}$  is result from the combined effect of radial translation and rotation, and it is not possible to determine the specific rotation angle and radial translation distance separately (one less degree of freedom). Hence, another parameter  $\alpha_j$  is introduced to describe the degree of rotation, which is defined as the angular position of the first hole in the full circle patched from the  $j^{th}$  fragment (the right side of figure 2). Therefore, the exact angle of rotation of  $S_j$  is  $\alpha_j - \alpha_0$ . As a result, the apparent angular position of the  $i^{th}$  hole in the  $j^{th}$  section with respect to its arc-centre is:

$$\phi_{ij} = 2\pi \frac{i-1}{N} + \alpha_j \quad (3.1)$$

So that the predicted location of the  $i^{th}$  hole in the  $j^{th}$  section under Cartesian coordinates is:

$$\begin{aligned} \mathbf{m}_{ij} &= \mathbf{r}_{0j} + (r \cos \phi_{ij}, r \sin \phi_{ij}) \\ &= (x_{0j} + r \cos \phi_{ij}, y_{0j} + r \sin \phi_{ij}) \end{aligned} \quad (3.2)$$

## 4) Two Error Modelling Methods

### Problem Statement

- (c) Use a Gaussian likelihood function for each hole location and assume that the error in the placement of different holes is independent. Consider two options of the covariance matrix. [5]

Besides the translation and rotation of the fragments, there are also random measurement errors in the data provided. The measurement error of the  $i^{th}$  hole represented in the Cartesian coordinate system is:

$$\begin{aligned} \mathbf{e}_{ij} &= \mathbf{d}_i - \mathbf{m}_{ij} \\ &= (x_i - x_{0j} - r\cos\phi_{ij}, y_i - y_{0j} - r\sin\phi_{ij}) \end{aligned} \quad (4.1)$$

### 4.1 Error Decomposition

Considering the actual sources and nature of errors, it is better to decompose the errors into radial and tangential directions and convey their representation to a relative coordinate system. As shown in figure 3, a pair of orthogonal unit vectors  $\hat{\mathbf{r}}_i, \hat{\mathbf{t}}_i$  was chosen as the bases of the relative coordinate system:

$$\begin{aligned} \hat{\mathbf{r}}_i &= (\cos\phi_{ij}, \sin\phi_{ij}) \\ \hat{\mathbf{t}}_i &= (\sin\phi_{ij}, -\cos\phi_{ij}) \end{aligned} \quad (4.2)$$

Therefore, the error coordinates in the relative system can be written as:

$$\mathbf{e}'_{ij} = (e_r, e_t)^T \quad (4.3)$$

Where the radial and the tangential components are:

$$\begin{aligned} e_r &= \mathbf{e}_{ij} \cdot \hat{\mathbf{r}}_i = (x_i - x_{0j})\cos\phi_{ij} + (y_i - y_{0j})\sin\phi_{ij} - r \\ e_t &= \mathbf{e}_{ij} \cdot \hat{\mathbf{t}}_i = (x_i - x_{0j})\sin\phi_{ij} - (y_i - y_{0j})\cos\phi_{ij} \end{aligned} \quad (4.4)$$

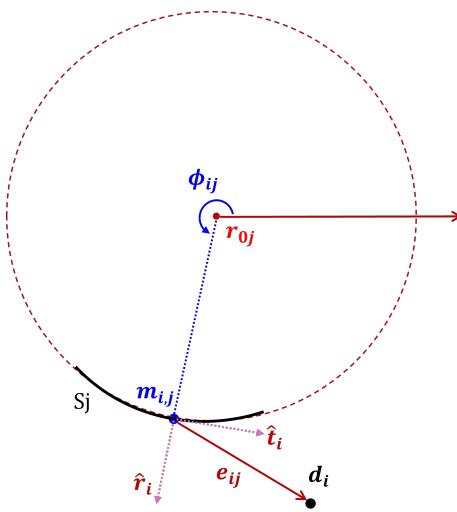


Figure 3: Measurement error in relative coordinate system

Finally, a 2-D Gaussian model are built to describe the measurement errors:

$$PDF_{ij} = \frac{\exp(-\frac{1}{2}\mathbf{e}'_{ij}^T \mathbf{C}^{-1} \mathbf{e}'_{ij})}{2\pi\sqrt{|\mathbf{C}|}} \quad (4.5)$$

## 4.2 Isotropic Gaussian Model

Considering a 2-D isotropic Gaussian distribution with standard deviation  $\sigma$ , the covariance matrix is:

$$\mathbf{C} = \begin{pmatrix} \sigma^2 & 0 \\ 0 & \sigma^2 \end{pmatrix} \quad (4.6)$$

Exploiting the equations 4.3, 4.5 and 4.6, the likelihood function is:

$$\mathcal{L}(D | \theta) = \prod_{j=0}^7 \prod_i^{i \text{ in } j} PDF_{ij} = (2\pi\sigma^2)^{-n} \prod_{j=0}^7 \prod_i^{i \text{ in } j} \exp \left[ -\frac{e_r^2 + e_t^2}{2\sigma^2} \right] \quad (4.7)$$

Where the number of measurements  $n=81$ , the measurement data  $D = \{\mathbf{d}_i \mid i = 1, 2, \dots, n\}$  and the parameter vector  $\theta = (N, r, \{(x_{0j}, y_{0j})\}, \{\alpha_j\}, \mathbf{C})$ . Therefore, the log likelihood is:

$$\log \mathcal{L}(D | \theta) = -n \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{j=0}^7 \sum_i^{i \text{ in } j} (e_r^2 + e_t^2) \quad (4.8)$$

## 4.3 Radial-Tangential Gaussian Model

Considering a 2-D Gaussian distribution with principal axes aligned with the radial and tangential directions for each hole, the standard deviations in each direction are  $\sigma_r$  and  $\sigma_t$ , so that the covariance matrix is:

$$\mathbf{C} = \begin{pmatrix} \sigma_r^2 & 0 \\ 0 & \sigma_t^2 \end{pmatrix} \quad (4.9)$$

Again, the likelihood function is:

$$\mathcal{L}(D | \theta) = (2\pi\sigma_r\sigma_t)^{-n} \prod_{j=0}^7 \prod_i^{i \text{ in } j} \exp \left[ -\frac{e_r^2}{2\sigma_r^2} - \frac{e_t^2}{2\sigma_t^2} \right] \quad (4.10)$$

Therefore, the log likelihood is:

$$\log \mathcal{L}(D | \theta) = -n \log(2\pi\sigma_r\sigma_t) - \frac{1}{2} \sum_{j=0}^7 \sum_i^{i \text{ in } j} \left( \frac{e_r^2}{\sigma_r^2} + \frac{e_t^2}{\sigma_t^2} \right) \quad (4.11)$$

## 5) Gradient Calculation

### Problem Statement

- (d) Make use of the gradient of the log-likelihood. Include the details of how the derivatives were calculated in the report and brief details of any code tests performed. [10]

For efficient parameter estimation in the following, I derive analytical gradients of the log-likelihood with respect to each parameter for both models.

### 5.1 Gradient for Isotropic Model

According to equation 4.8, there are two generic form when calculating the derivatives of the log likelihood with respect to each component  $\theta_\mu$  of the parameter vector  $\theta$ :

1. For  $\theta_\mu = \sigma$ :

$$\frac{\partial \log \mathcal{L}}{\partial \sigma} = -\frac{2n}{\sigma} + \frac{1}{\sigma^3} \sum_{j=0}^7 \sum_i^{i \text{ in } j} (e_r^2 + e_t^2) \quad (5.1)$$

Which can be calculated using equation 4.4.

2. For  $\theta_\mu \neq \sigma$ :

$$\frac{\partial \log \mathcal{L}}{\partial \theta_\mu} = -\frac{1}{\sigma^2} \sum_{j=0}^7 \sum_i^{i \text{ in } j} (e_r \cdot \frac{\partial e_r}{\partial \theta_\mu} + e_t \cdot \frac{\partial e_t}{\partial \theta_\mu}) \quad (5.2)$$

So that the partial derivatives of the log-likelihood can be calculated respectively for each parameter:

- For N:

$$\begin{aligned} \frac{\partial \log \mathcal{L}}{\partial N} &= -\frac{1}{\sigma^2} \sum_{j=0}^7 \sum_i^{i \text{ in } j} (e_r \cdot \frac{\partial e_r}{\partial N} + e_t \cdot \frac{\partial e_t}{\partial N}) \\ &= \frac{r}{\sigma^2} \frac{2\pi}{N^2} \sum_{j=0}^7 \sum_i^{i \text{ in } j} (i-1)[(x_i - x_{0j}) \sin \phi_{ij} - (y_i - y_{0j}) \cos \phi_{ij}] \end{aligned} \quad (5.3)$$

- For r:

$$\begin{aligned} \frac{\partial \log \mathcal{L}}{\partial r} &= -\frac{1}{\sigma^2} \sum_{j=0}^7 \sum_i^{i \text{ in } j} (e_r \cdot \frac{\partial e_r}{\partial r} + e_t \cdot \frac{\partial e_t}{\partial r}) \\ &= \frac{1}{\sigma^2} \sum_{j=0}^7 \sum_i^{i \text{ in } j} [(x_i - x_{0j}) \cos \phi_{ij} + (y_i - y_{0j}) \sin \phi_{ij} - r] \end{aligned} \quad (5.4)$$

- For  $x_{0j}$ :

$$\begin{aligned} \frac{\partial \log \mathcal{L}}{\partial x_{0j}} &= -\frac{1}{\sigma^2} \sum_{j=0}^7 \sum_i^{i \text{ in } j} (e_r \cdot \frac{\partial e_r}{\partial x_{0j}} + e_t \cdot \frac{\partial e_t}{\partial x_{0j}}) \\ &= \frac{1}{\sigma^2} \sum_{j=0}^7 \sum_i^{i \text{ in } j} (x_i - x_{0j} - r \cos \phi_{ij}) \end{aligned} \quad (5.5)$$

- For  $y_{0j}$ :

$$\begin{aligned} \frac{\partial \log \mathcal{L}}{\partial y_{0j}} &= -\frac{1}{\sigma^2} \sum_{j=0}^7 \sum_i^{i \text{ in } j} (e_r \cdot \frac{\partial e_r}{\partial y_{0j}} + e_t \cdot \frac{\partial e_t}{\partial y_{0j}}) \\ &= \frac{1}{\sigma^2} \sum_{j=0}^7 \sum_i^{i \text{ in } j} (y_i - y_{0j} - r \sin \phi_{ij}) \end{aligned} \quad (5.6)$$

- For  $\alpha_j$ :

$$\begin{aligned} \frac{\partial \log \mathcal{L}}{\partial \alpha_j} &= -\frac{1}{\sigma^2} \sum_{j=0}^7 \sum_i^{i \text{ in } j} (e_r \cdot \frac{\partial e_r}{\partial \alpha_j} + e_t \cdot \frac{\partial e_t}{\partial \alpha_j}) \\ &= -\frac{1}{\sigma^2} \sum_{j=0}^7 \sum_i^{i \text{ in } j} [(x_i - x_{0j}) r \sin \phi_{ij} - (y_i - y_{0j}) r \cos \phi_{ij}] \end{aligned} \quad (5.7)$$

## 5.2 Gradient for Anisotropic Model

According to equation 4.11, there are two generic form when calculating the derivatives of the log likelihood with respect to each component  $\theta_\mu$  of the parameter vector  $\theta$ :

1. For  $\theta_\mu = \sigma$ :

$$\begin{aligned} \frac{\partial \log \mathcal{L}}{\partial \sigma_r} &= -\frac{n}{\sigma_r} + \frac{1}{\sigma_r^3} \cdot \sum_{j=0}^7 \sum_i^{i \text{ in } j} e_r^2 \\ \frac{\partial \log \mathcal{L}}{\partial \sigma_t} &= -\frac{n}{\sigma_t} + \frac{1}{\sigma_t^3} \cdot \sum_{j=0}^7 \sum_i^{i \text{ in } j} e_t^2 \end{aligned} \quad (5.8)$$

Which can be calculated using equation 4.4.

2. For  $\theta_\mu \neq \sigma$ :

$$\frac{\partial \log \mathcal{L}}{\partial \theta_\mu} = - \sum_{j=0}^7 \sum_i^{i \text{ in } j} \left( \frac{e_r}{\sigma_r^2} \cdot \frac{\partial e_r}{\partial \theta_\mu} + \frac{e_t}{\sigma_t^2} \cdot \frac{\partial e_t}{\partial \theta_\mu} \right) \quad (5.9)$$

So that the partial derivatives of the log-likelihood can be calculated respectively for each parameter:

- For N:

$$\begin{aligned} \frac{\partial \log \mathcal{L}}{\partial N} &= - \frac{2\pi}{N^2} \sum_{j=0}^7 \sum_i^{i \text{ in } j} (i-1) \\ &\quad \left\{ \left( \frac{1}{\sigma_r^2} - \frac{1}{\sigma_t^2} \right) [(x_i - x_{0j})^2 \sin \phi_{ij} \cos \phi_{ij} - (x_i - x_{0j})(y_i - y_{0j}) \cos 2\phi_{ij} - (y_i - y_{0j})^2 \sin \phi_{ij} \cos \phi_{ij}] \right. \\ &\quad \left. - \frac{r}{\sigma_r^2} [(x_i - x_{0j}) \sin \phi_{ij} - (y_i - y_{0j}) \cos \phi_{ij}] \right\} \end{aligned} \quad (5.10)$$

- For r:

$$\frac{\partial \log \mathcal{L}}{\partial r} = \frac{1}{\sigma_r^2} \sum_{j=0}^7 \sum_i^{i \text{ in } j} [(x_i - x_{0j}) \cos \phi_{ij} + (y_i - y_{0j}) \sin \phi_{ij} - r] \quad (5.11)$$

- For  $x_{0j}$ :

$$\begin{aligned} \frac{\partial \log \mathcal{L}}{\partial x_{0j}} &= \sum_{j=0}^7 \sum_i^{i \text{ in } j} \left\{ \frac{1}{\sigma_r^2} [(x_i - x_{0j}) \cos^2 \phi_{ij} + (y_i - y_{0j}) \sin \phi_{ij} \cos \phi_{ij} - r \cos \phi_{ij}] \right. \\ &\quad \left. + \frac{1}{\sigma_t^2} [(x_i - x_{0j}) \sin^2 \phi_{ij} - (y_i - y_{0j}) \sin \phi_{ij} \cos \phi_{ij}] \right\} \end{aligned} \quad (5.12)$$

- For  $y_{0j}$ :

$$\begin{aligned} \frac{\partial \log \mathcal{L}}{\partial y_{0j}} &= \sum_{j=0}^7 \sum_i^{i \text{ in } j} \left\{ \frac{1}{\sigma_r^2} [(x_i - x_{0j}) \sin \phi_{ij} \cos \phi_{ij} + (y_i - y_{0j}) \sin^2 \phi_{ij} - r \sin \phi_{ij}] \right. \\ &\quad \left. + \frac{1}{\sigma_t^2} [(y_i - y_{0j}) \cos^2 \phi_{ij} - (x_i - x_{0j}) \sin \phi_{ij} \cos \phi_{ij}] \right\} \end{aligned} \quad (5.13)$$

- For  $\alpha_j$ :

$$\begin{aligned} \frac{\partial \log \mathcal{L}}{\partial \alpha_j} &= - \sum_{j=0}^7 \sum_i^{i \text{ in } j} \\ &\quad \left\{ \left( \frac{1}{\sigma_r^2} - \frac{1}{\sigma_t^2} \right) [(x_i - x_{0j})(y_i - y_{0j}) \cos(2\phi_{ij}) + (y_i - y_{0j})^2 \sin \phi_{ij} \cos \phi_{ij} - (x_i - x_{0j})^2 \sin \phi_{ij} \cos \phi_{ij}] \right. \\ &\quad \left. + \frac{1}{\sigma_r^2} [(x_i - x_{0j}) r \sin \phi_{ij} - (y_i - y_{0j}) r \cos \phi_{ij}] \right\} \end{aligned} \quad (5.14)$$

### 5.3 Correctness check

Using the results above, all gradient calculations were implemented in a Python script `models.py`, which encapsulates a class named `AntikytheraModel` containing not only the computation of log likelihoods and gradients of both models, but also their correctness check. In the function `check_grad()` inside of `AntikytheraModel`, I compared the analytical results with numerical approximations using finite difference method, taking an initial set of parameter values as shown in table 2:

Table 2: Parameter values for gradient check

Parameters	$N$	$r/\text{mm}$	$\sigma, \sigma_{r,t}/\text{mm}$	$(x_{0j}, y_{0j})/\text{mm}$	$\alpha_j/\text{radians}$
Values	354	100	1	$(0, 0)$	0

The results were printed and presented as a table in the note book `run.ipynb`, showing that for both models, the gradients computed analytically and numerically matched to within a relative error of  $10^{-6}$ , confirming the correctness of the analytical derivations.

Anisotropic model gradient check:  
Max relative difference: 1.099508e-06

Param	Analytical	Numerical	Rel Diff
N	509.029084	509.029109	2.431333e-08
r	-2224.360999	-2224.361058	1.322940e-08
sigma_r	103645.695286	103645.539784	7.501622e-07
sigma_t	283130.281859	283129.857038	7.502215e-07
x0_0	-84.400690	-84.400672	1.080141e-07
y0_0	92.596530	92.596543	7.157034e-08
alpha_0	-1563.220455	-1563.220460	1.559213e-09
x0_1	1334.971312	1334.971312	7.371911e-11
...	...	...	...

(a) Example 1

Isotropic model gradient check:  
Max relative difference: 1.099508e-06

Param	Analytical	Numerical	Rel Diff
N	509.029084	509.029109	2.431333e-08
r	-2224.360999	-2224.361058	1.322940e-08
sigma	386775.977145	386775.396793	7.502432e-07
x0_0	-84.400690	-84.400672	1.080141e-07
y0_0	92.596530	92.596543	7.157034e-08
alpha_0	-1563.220455	-1563.220460	1.559213e-09
x0_1	1334.971312	1334.971312	7.371911e-11
y0_1	-408.368777	-408.368796	2.341078e-08
...	...	...	...

(b) Example 2

Figure 4: Comparison of analytically and numerically computed gradients with respect to the first few parameters.

## 6) Maximum Likelihood Predictions

### Problem Statement

- (e) Find the maximum likelihood parameters for both models. On your plot of the measured hole locations, add new markers showing predictions from both models for the hole locations. [8]

### 6.1 Methodology

To find the maximum likelihood parameters for both models, I employed the Limited-memory BFGS algorithm with bounds (L-BFGS-B) [3] implemented in SciPy's `minimize` function [4]. This quasi-Newton method is well-suited for this problem, as it efficiently handles the high dimensionality of the parameter space while making use of the analytical gradients derived in Section 5, without requiring the knowledge of the Hessian matrix, which is difficult to derive and compute. In the `mle.py` module, the maximum likelihood estimation along with the visualisation of the predicted hole positions were implemented. For the isotropic model, I optimised over 27 parameters:  $N$ ,  $r$ ,  $\sigma$ , and 3 parameters  $(x_{0j}, y_{0j}, \alpha_j)$  for each of the 8 sections. For the radial-tangential model, I optimised over 28 parameters, with  $\sigma_r$  and  $\sigma_t$  replacing the single  $\sigma$  parameter.

According to the nature of the parameters, sensible parameter boundaries were set, as shown in table 3, where  $N$  was treated as a continuous parameter for simplification.

Table 3: Parameter bounds for L-BFGS-B optimisation

Parameters	$N$	$r$ /mm	$\sigma, \sigma_{r,t}$ /mm	$x_{0j}$ /mm	$y_{0j}$ /mm	$\alpha_j$ /radians
Bounds	(300, 400)	(50, 100)	(1e-4, 1)	(70, 90)	(100, 150)	$(-\pi, -\frac{\pi}{2})$

Exploiting the same set of initial values shown in table 2, the results were printed and plotted in `run.ipynb`.

## 6.2 Results

After the optimisation, the maximum likelihood parameters for both models were obtained, including the intrinsic parameters  $N, r, \sigma/\sigma_{r,t}$ , which are listed in Tables 4 and 5.

Table 4: MLE for Isotropic Model

Parameter	ML Estimation
N	351.41
r (mm)	76.51
$\sigma$ (mm)	0.0880

Table 5: MLE for Anisotropic Model

Parameter	ML Estimation
N	355.25
r (mm)	77.34
$\sigma_r, \sigma_t$ (mm)	0.0255, 0.1211

It can be seen that both models yield estimates of N closer to 354 (lunar calendar hypothesis) than to 365 (solar calendar hypothesis), supporting the recent suggestion that the Antikythera mechanism's calendar ring was designed as a lunar calendar [5].

Comparing the two models, the anisotropic model estimate  $N \approx 355.25$ , which is remarkably close to the 354 holes expected for a lunar calendar, while the isotropic model's estimate of N is slightly lower. The difference in the radius estimates is relatively small, suggesting both model predictions agree well on the overall scale of the mechanism.

However, the most notable difference between the models appears in the error parameters. The isotropic model's single error parameter  $\sigma$  represents an overall uncertainty of errors, while the anisotropic model reveals a significant difference between radial and tangential uncertainty, with the tangential error being approximately 4.8 times larger than the radial error. This aligns with the fact that the manufacturing process had much better control over the radial positioning of holes than their tangential spacing, I will discuss it later in section 8.

As for the position parameters, new markers were then added to the plot of the measured hole locations, using the same colour map as in Figure 1, to indicate the hole positions predicted by the maximum likelihood method.

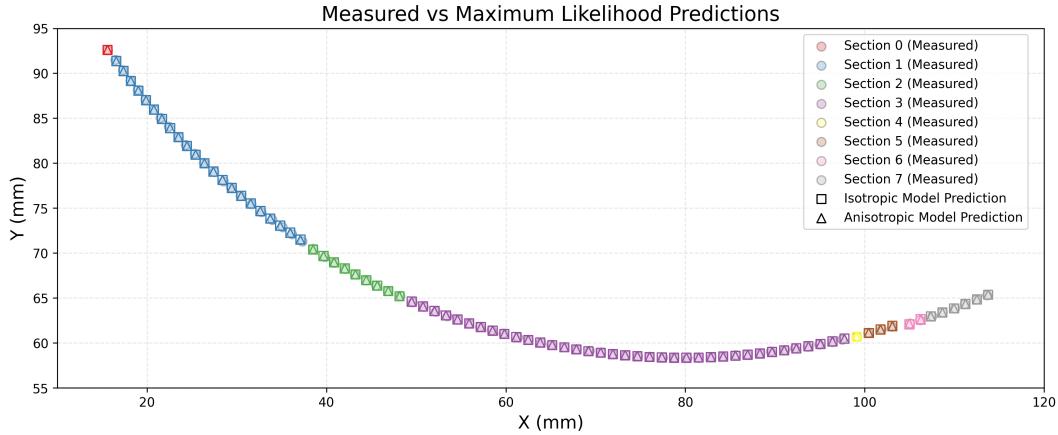


Figure 5: The maximum likelihood estimation of the hole positions with the measured positions.

Overall, the predicted positions align well with the measurements, with only minor differences between predictions and measurements, suggesting that both models provide correct representations of the positional parameters. They precisely capture the rotation and translation of each fragment while effectively excluding manufacturing and measurement errors, which are encapsulated in the error parameters  $\sigma, \sigma_{r,t}$ . This is evidenced by the clear discontinuities between fragments (represented by different colours), while the hole positions within each fragment remain uniformly distributed. In this sense, the model effectively "corrects" for the measurement and manufacturing errors in the data and provides an intuitive indication of which holes contribute more significantly to the error parameters.

## 7) Posterior Sampling with NUST

### Problem Statement

(f) Sample the posterior distribution for both models using any variation of the Hamiltonian Monte Carlo sampling algorithm. Pick one of the holes, on your plot of the measured hole locations, show the posterior predictive distribution for its hole locations using both models. [8]

### 7.1 Methodology

To infer the models' parameters comprehensively, I employ the No-U-Turn Sampler (NUTS), an adaptive variant of Hamiltonian Monte Carlo to sample from the posterior distributions of both models.

Following the point estimate results from MLE and tight tolerances constrained by the geometry of the calendar ring, I take use of the uniform prior distributions with sensible bounds below:

Table 6: Bounds for uniform priors of each parameter

Parameters	$N$	$r$ /mm	$\sigma, \sigma_{r,t}$ /mm	$x_{0j}$ /mm	$y_{0j}$ /mm	$\alpha_j$ /radians
Prior Bounds	(335, 375)	(60, 90)	(0.02, 0.2)	(70, 90)	(125, 145)	(-3.0, -2.0)

In addition, the decision of the bounds of  $\sigma, \sigma_{r,t}$  prior is guided by the results of previous MLE and a few pre-experiments before the formal sampling.

The implementation of the NUTS sampling was carried out in the `sampler.py` module with PyMC library. Notably, I dropped the data of section 0 and 4, each contain only one hole, which cannot constrain 3 displacement parameters with only 2 coordinate data of each section. Thereby the 8 sections were renumbered into 0 to 5 in the code and the dimension of the parameter space was reduced to 22. Having said that, I ran 4 chains with 2000 tuning steps and 2000 sampling steps for each model, which is directly presented in `run.ipynb`.

### 7.2 Results

To visualise the predictive distribution of positions, figure 6 shows the original data with hole 16 in the second section highlighted, along with the posterior predictive distribution for that hole's location as well as the mean values of samples (marked with crosses) under both models. The sampling was sufficiently effective that the 2000 samples shown in Figure 6 are excessively clustered. For observation purpose, figure 7 provides a zoomed-in view of the same posterior distribution.

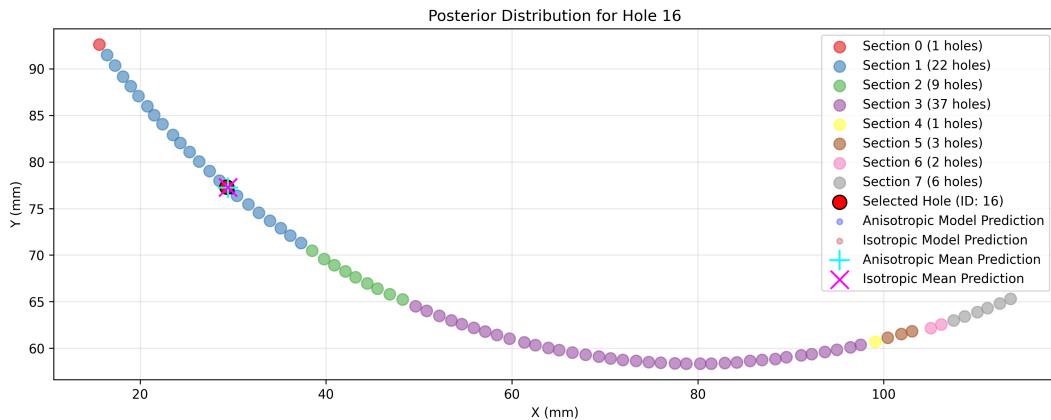


Figure 6: 2000 posterior samples from both models (blue for anisotropic, red for isotropic) for hole locations of one hole (ID: 16).

In fact, the selection of hole 16 was based on its evidently greater deviation in the tangential direction of the ring, as its position is noticeably closer to the hole in front of it (15th) and more distant from the following one (17th). It can be observed that both models lead to a predicted position for hole 16 that is more uniformly distributed along the ring (lie closer to hole 17), effectively "correcting" the positional error of its measured location.

Specifically, the anisotropic model's posterior distribution shows a notable difference in spread between the radial and tangential directions, with greater variance along the tangential direction, indicating consistency between the naked-eye observation and sampling approaches. However, the isotropic model's posterior distribution forms a more circular cloud, as expected given its equal variance assumption in all directions.

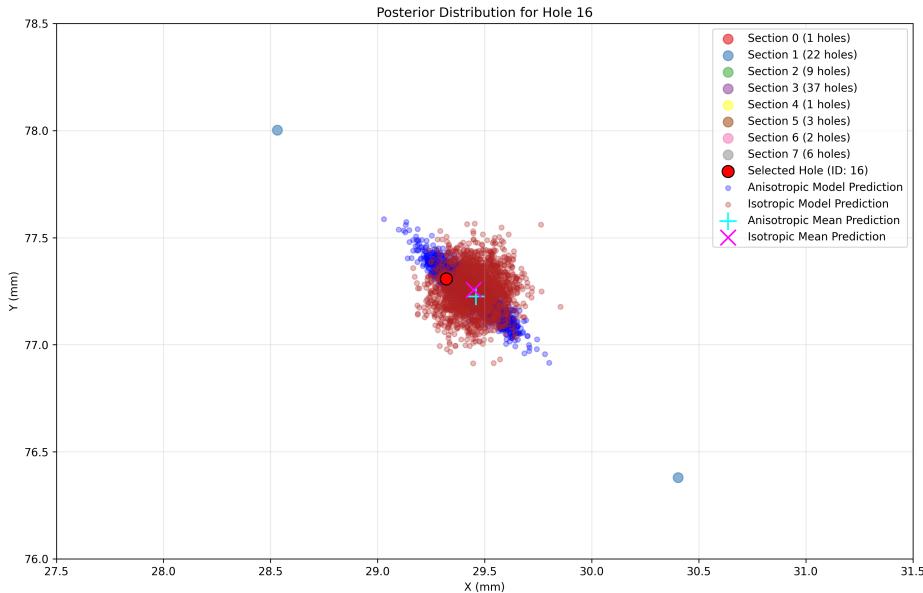


Figure 7: 2000 posterior samples from both models (blue for anisotropic, red for isotropic) for hole locations of one hole (ID: 16) (zoomed-in).

As for the intrinsic parameters, the sampling results and the corner plot were shown in figure 8 and 9.

Isotropic Model Posterior Summary:										
	mean	sd	hdi_3%	hdi_97%	mcse_mean	mcse_sd	ess_bulk	ess_tail	r_hat	
N	355.374	4.341	347.697	363.841	0.135	0.071	1045.0	1930.0	1.01	
r	77.370	0.935	75.714	79.178	0.029	0.015	1043.0	1850.0	1.01	
sigma	0.094	0.006	0.084	0.105	0.000	0.000	6928.0	5259.0	1.00	
Anisotropic Model Posterior Summary:										
	mean	sd	hdi_3%	hdi_97%	mcse_mean	mcse_sd	ess_bulk	ess_tail	r_hat	
N	355.269	1.387	352.700	357.861	0.033	0.019	1771.0	3011.0	1.0	
r	77.343	0.282	76.816	77.859	0.007	0.004	1714.0	2696.0	1.0	
sigma_r	0.028	0.003	0.024	0.033	0.000	0.000	6408.0	3600.0	1.0	
sigma_t	0.129	0.011	0.110	0.150	0.000	0.000	6338.0	4348.0	1.0	

Figure 8: Sampling results and metrics from both models for the intrinsic parameters.

It can be seen that the mean sample values of intrinsic parameters for both models are very close to the previous MLE results, indicating the accuracy of sampling estimates. Comparing the models, the narrower credible intervals in the anisotropic model (more than 3 times greater standard deviations for both  $N$  and  $r$ ) suggest that accounting for direction-specific errors leads to more confident parameter estimates, reflecting the model's better alignment with the physical reality of the artifact, which will be discussed more in section 8.

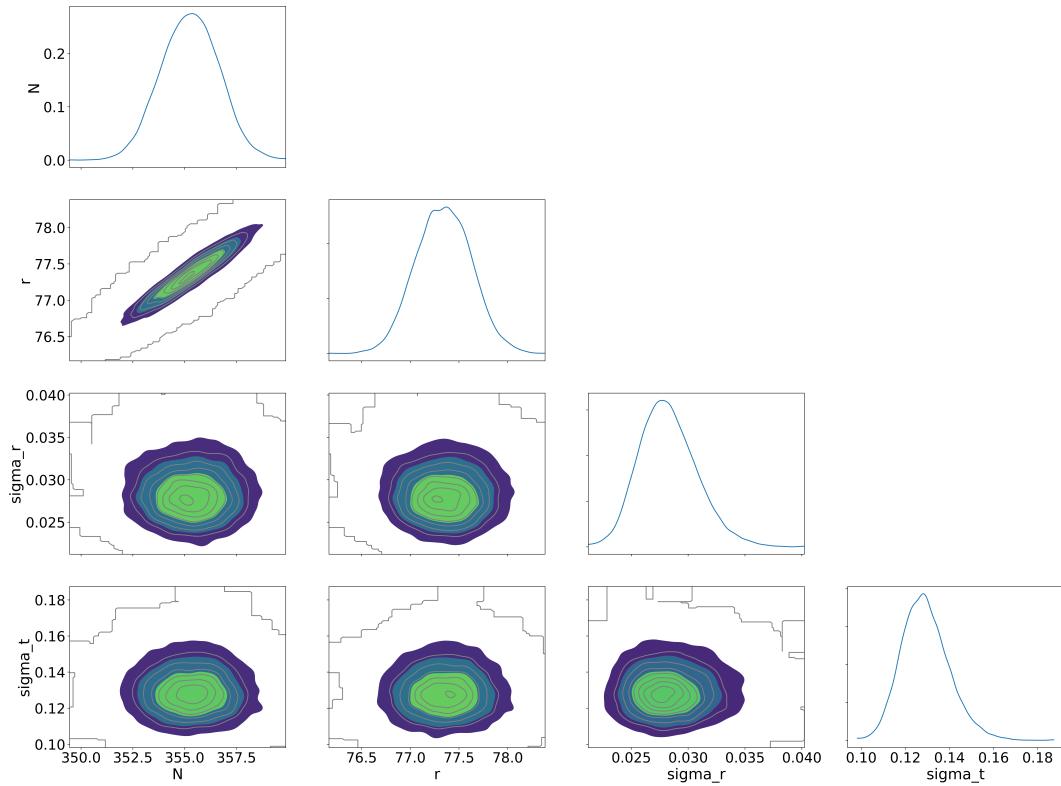


Figure 9: Posterior corner plot of anisotropic model for the intrinsic parameters.

The posterior corner plot (figure 9) reveals important parameter correlations and the shapes of the marginal distributions. The distribution between  $N$  and  $r$  shows a reasonable positive correlation while the distributions for  $\sigma_r$  and  $\sigma_t$  show clear separation with no overlap, confirming that the tangential error is indeed significantly larger than the radial error.

## 8) Model Comparison

### Problem Statement

- (g) Explain the role of the 2-dimensional covariance matrix. Explain why the radial/tangential model for the covariance matrix might a priori be considered a good model. Does the data favour this radial/tangential model over the simple isotropic model? Your answer should be justified using appropriate plots and/or quantitative measurements. [10]

### 8.1 Explanation

The covariance matrix  $\mathbf{C}$  in the models characterises the statistical properties of the measurement errors and the manufacturing errors. In the context of the measurements of Antikythera mechanism, the errors may arise from various sources, including the X-ray imaging process, the identification of hole centres in the images, and physical degradation of the artefact. Regarding manufacturing errors, these are likely to originate from the shaping of the circular calendar ring during the construction of the Antikythera mechanism (radial error), and from the process of uniformly drilling holes along a constant radius, (both radial and tangential errors).

Notably, the covariance matrix of the Gaussian error model does not account for the translational and rotational errors between fragments, as these displacement errors have already been incorporated into the model-predicted positions  $\mathbf{m}_{ij}$  through  $x_{0j}$ ,  $y_{0j}$ , and  $\alpha_j$ . The error described by the Gaussian model,  $\mathbf{e}_{ij}$ , is the result of subtracting the displacement error  $\mathbf{m}_{ij}$  from the total error  $\mathbf{d}_i$ .

Therefore, the radial-tangential model for the covariance matrix might be considered a better model than the simple isotropic model, which assumes uniform error distribution in all directions, with a single parameter  $\sigma$  controlling the overall error magnitude. This model is concise but may not reflect the physical reality of the measurement and manufacturing processes. For example, it is much harder to distribute holes evenly around the ring during the manufacturing than to shape a circle using a circular guide. Furthermore, even minor deviations in the projection angle during X-ray imaging can lead to anisotropic error patterns in different directions.

The radial-tangential model, conversely, decomposes the error into components aligned with the ring's geometry. This model reflects the correct physical intuition that errors are allowed to differ between the radial and tangential direction, not to mention that the higher model complexity make it capable to obtain more information about the real error distribution in the two-dimensional space.

## 8.2 Methodology

In order to compare the models quantitatively with the measured data, I use thermodynamic integration method to estimate the posterior odds ratio. Denote the isotropic model as A and the anisotropic model as B, the posterior odds ratio is:

$$\mathcal{O}_{A,B} = \frac{P(A | d)}{P(B | d)} = \frac{P(d | A)}{P(d | B)} \times \frac{P(A)}{P(B)}. \quad (8.1)$$

Assume the prior ratio term  $\frac{P(A)}{P(B)} = 1$ , and allow all the model parameters to take on all possible values by marginalising over them, the equation turns to:

$$\begin{aligned} \mathcal{O}_{A,B} &= \frac{\int d\theta_A P(d, \theta_A | A)}{\int d\theta_B P(d, \theta_B | B)} \\ &= \frac{\int d\theta_A \mathcal{L}(d | \theta_A, A) \pi(\theta_A | A)}{\int d\theta_B \mathcal{L}(d | \theta_B, B) \pi(\theta_B | B)} \\ &= \frac{Z_A}{Z_B} \end{aligned} \quad (8.2)$$

Where  $Z_A$  and  $Z_B$  are the Bayesian evidences for each model. Hence, the larger evidence means a better model.

To estimate the evidences and thereby obtain the odds ratio, the posterior powered by an inverse temperature  $\beta$  were built as shown in equation 8.3, and implemented in a class named `ModelComparison`, which were encapsulated in the module `compare.py`.

$$P(\theta | d, \beta) = \frac{\mathcal{L}(d | \theta)^\beta \pi(\theta)}{Z(\beta)} \quad (8.3)$$

Therefore, the evidence as a function of  $\beta$  can be written as:

$$Z(\beta) = \int d\theta \mathcal{L}(d | \theta)^\beta \pi(\theta) \quad (8.4)$$

So that the differentiation of log-evidence with respect to  $\beta$  can be written as the expectation of log likelihood over the parameters drawn from the powered posterior distribution:

$$\begin{aligned} \frac{d}{d\beta} \log Z(\beta) &= \frac{1}{Z(\beta)} \frac{dZ}{d\beta} \\ &= \frac{1}{Z(\beta)} \int d\theta \pi(\theta) \mathcal{L}(d | \theta)^\beta \log \mathcal{L}(d | \theta) \\ &= \int d\theta P(\theta | d, \beta) \log \mathcal{L}(d | \theta) \\ &= \mathbb{E}_{\theta \sim P(\theta | d, \beta)} [\log \mathcal{L}(d | \theta)] \end{aligned} \quad (8.5)$$

Having said that, the expectation value can be approximated by sampling from the powered posterior distribution with MCMC, and computing the sample mean. Then, setting up a series of increasing values for the inverse temperatures  $0 \leq \beta_\mu \leq 1$  where  $\mu = 0, 1, \dots, M$  with  $\beta_0 = 0$  and  $\beta_M = 1$ , so that the log-evidence can be

approximated using the trapezium rule:

$$\begin{aligned} \log Z &= \int_0^1 d\beta \mathbb{E}_{\theta \sim P(\theta|d,\beta)} [\log \mathcal{L}(d|\theta)] \\ &\approx \frac{1}{2} \sum_{\mu=1}^M (\mathbb{E}_{\beta_{\mu-1}} [\log \mathcal{L}] + \mathbb{E}_{\beta_\mu} [\log \mathcal{L}]) (\beta_\mu - \beta_{\mu-1}) \end{aligned} \quad (8.6)$$

For implementation, I created 20 points evenly distributed between 0 and 1 and then raised them to the power of 5 to make more values of inverse temperatures near 0 for numerical stability when integrating, as the expected log-likelihood over the modified posterior increases sharply when  $\beta$  approaches zero. Exploiting the NUTS algorithm using PyMC library, all sampling results were saved as `model_comparison.pkl` in the `results` folder. (See `run.ipynb` and `compare.py` for more details).

### 8.3 Result

The last part of `run.ipynb` called the `ModelComparison` class and printed the comparison results after sampling from the modified posteriors of both models at each temperature. The results, listed in table 7, show that the data favour the radial-tangential model over the simple isotropic model as the anisotropic model has a larger evidence and the odds ratio  $\mathcal{O}_{A,B} = \frac{Z_A}{Z_B} < 1$ , which confirms what was suspected before the experiment.

Table 7: Comparison results using thermodynamic integration

	<b>Isotropic model</b>	<b>Anisotropic model</b>	$\mathcal{O}_{A,B}$
<b>Evidence</b>	5.77e-29	1.12e-27	0.0515

## References

- [1] A. Thoeni, C. Budiselic, and A. Ramsey, *Replication Data for: Antikythera Mechanism Shows Evidence of Lunar Calendar*, 2019.
- [2] G. Woan and J. Bayley, *An improved calendar ring hole-count for the antikythera mechanism*, 2024.
- [3] C. Zhu, R. Byrd, P. Lu, and J. Nocedal, *Algorithm 778: L-bfgs-b: Fortran subroutines for large-scale bound-constrained optimization*, *ACM Transactions on Mathematical Software* **23** (Dec., 1997) 550–560.
- [4] SciPy Community, *SciPy Optimize: minimize(method='L-BFGS-B')*, 2025.
- [5] C. Budiselic, A. T. Thoeni, M. Dubno, and A. T. Ramsey, “Antikythera mechanism shows evidence of lunar calendar.” SocArXiv, June, 2020. Revised February 2021.

## Appendix

**Declaration:** No auto-generation tools were used in this report except for generation of BibTeX references.