

# Ambiguities in gravitational lens models: impact on time delays of the source position transformation

Olivier Wertz, Bastian Orthen, and Peter Schneider

Argelander-Institut für Astronomie, Auf dem Hügel 71, 53121 Bonn, Germany  
 e-mail: owertz@alumni.ulg.ac.be

Received 6 November 2017 / Accepted 11 December 2017

## ABSTRACT

The central ambition of the modern time delay cosmography consists in determining the Hubble constant  $H_0$  with a competitive precision. However, the tension with  $H_0$  obtained from the *Planck* satellite for a spatially flat  $\Lambda$ CDM cosmology suggests that systematic errors may have been underestimated. The most critical of these errors probably comes from the degeneracy existing between lens models that was first formalized by the well-known mass-sheet transformation (MST). In this paper, we assess to what extent the source position transformation (SPT), a more general invariance transformation which contains the MST as a special case, may affect the time delays predicted by a model. To this aim, we have used pySPT, a new open-source python package fully dedicated to the SPT that we present in a companion paper. For axisymmetric lenses, we find that the time delay ratios between a model and its SPT-modified counterpart simply scale like the corresponding source position ratios,  $\Delta\hat{i}/\Delta t \approx \hat{\beta}/\beta$ , regardless of the mass profile and the isotropic SPT. Similar behavior (almost) holds for nonaxisymmetric lenses in the double image regime and for opposite image pairs in the quadruple image regime. In the latter regime, we also confirm that the time delay ratios are not conserved. In addition to the MST effects, the SPT-modified time delays deviate in general no more than a few percent for particular image pairs, suggesting that its impact on time delay cosmography seems not be as crucial as initially suspected. We also reflected upon the relevance of the SPT validity criterion and present arguments suggesting that it should be reconsidered. Even though a new validity criterion would affect the time delays in a different way, we expect from numerical simulations that our conclusions will remain unchanged.

**Key words.** cosmological parameters – gravitational lensing: strong

## 1. Introduction

The use of the gravitational lensing phenomenon as a cosmological tool offers an independent way to probe the nature of the universe (for the early work, see [Blandford & Narayan 1992](#)). To date, numerous weak and strong lensing observations have been employed to infer the fundamental cosmological parameters with an increasingly competitive precision. In the strong lensing regime, [Refsdal \(1964\)](#) established that multiple-image systems can theoretically be used to infer the Hubble parameter  $H_0$ . The method relies upon the idea that the propagation time of light rays emitted from a background source (typically an active galactic nucleus, AGN) towards the observer differs from one lensed image to another. The corresponding difference in arrival times, known as the time delay, is inversely proportional to  $H_0$ . This idea lays the basis of the modern time delay cosmography, which has been extensively addressed in literature; see, for example, the recent review [Treu & Marshall \(2016\)](#) and references therein.

For sake of clarity, we recall few key results of the well-known theory of time delays. Relative to an unperturbed ray emitted by a source located at  $\beta$ , the extra light travel time  $T(\theta)$  at an image position  $\theta$  is formally defined by

$$T(\theta) = \frac{D_{\Delta t}}{c} \left\{ \frac{1}{2} [\theta - \beta(\theta)]^2 - \psi(\theta) \right\} =: \frac{D_{\Delta t}}{c} \tau(\theta), \quad (1)$$

where  $\psi(\theta)$  is the deflection potential produced by a dimensionless surface mass density  $\kappa(\theta) = \nabla^2 \psi(\theta)/2$ ,  $\tau(\theta)$  is known as the Fermat potential, and  $D_{\Delta t}$  is referred to as the time delay distance

$$D_{\Delta t} = (1 + z_d) \frac{D_d D_s}{D_{ds}} \propto H_0^{-1}, \quad (2)$$

where  $z_d$  is the redshift of the deflector and  $D$  is the angular diameter distances between the observer and deflector (d), observer and source (s), and deflector and source (ds). In Eq. (1), the first term in brackets describes the geometrical deviation of the light ray due to the lens, whereas the second describes the time delay that a ray experiences as it crosses the deflection potential. The relative time delay  $\Delta t_{ij}$  between a pair of lensed images  $\theta_i$  and  $\theta_j$  is obtained by differencing the corresponding extra light travel time:

$$\Delta t_{ij} = T(\theta_i) - T(\theta_j) = \frac{D_{\Delta t}}{c} [\tau(\theta_i) - \tau(\theta_j)] =: \frac{D_{\Delta t}}{c} \Delta \tau_{ij}. \quad (3)$$

From Eq. (3),  $H_0$  inference can be conceptually performed by constraining the time delay distance  $D_{\Delta t}$ , provided that both accurate time delay measurements and a mass model which predicts  $\Delta \tau_{ij}$  can be obtained. At present, a few percent precision time delays have been measured for several multiple-image systems based on different light curve analysis methods (see e.g., [Vuissoz et al. 2008](#); [Paraficz & Hjorth 2010](#); [Courbin et al. 2011](#); [Fohlmeister et al. 2013](#); [Eulaers et al. 2013](#); [Tewes et al. 2013](#); [Rathna Kumar et al. 2013](#); [Bonvin et al. 2017](#); [Akhunov et al. 2017](#)). In the foreseeable future, we can expect thousands of lensed quasars to be discovered by the next generation of instruments ([Jean et al. 2001](#); [Coe & Moustakas 2009](#); [Oguri & Marshall 2010](#); [Finet et al. 2012, 2015](#); [Liao et al. 2015](#); [Finet & Surdej 2016](#)). Among them, numerous suitable candidates for

robust time delay measurements should lead the time delay cosmography to the next level. However, constraining the lens mass distribution turns out to be as decisive as measuring time delays with high precision. Given a measured time delay between two lensed images, more concentrated mass distributions lead to shorter time delay distance estimations, hence to larger values of  $H_0$  (Kochanek 2002). The Fermat potential difference  $\Delta\tau_{ij}$  is primarily sensitive to the strong lensing effects produced by the main lens. However, a realistic time delay cosmography should also consider the lensing effects of any external mass structures located in the vicinity of the main lens, as well as along the line of sight (LOS; e.g., Seljak 1994; Bar-Kana 1996). If the LOS mass effects are sufficiently small, they can be approximated by an external shear and an external convergence, usually denoted as  $\kappa_{\text{ext}}$ , which need to be characterized<sup>1</sup> (see e.g., Keeton 2003; Fassnacht et al. 2006; Suyu et al. 2010, 2013; Wong et al. 2011, 2017). Otherwise, these external mass structures need to be explicitly included in the mass model, for instance by considering the full multi-plane lensing formalism (Schneider 2014b; McCully et al. 2014, 2016).

As first shown in Falco et al. (1985), the dimensionless surface mass density  $\kappa(\theta)$  and the class of mass models  $\kappa_l(\theta)$  defined as

$$\kappa_l(\theta) = \lambda \kappa(\theta) + (1 - \lambda), \quad (4)$$

along with the corresponding unobservable source rescaling  $\beta \rightarrow \lambda\beta$ , lead to identical lensing observables, except for the time delays between pairs of lensed images which are transformed such that  $\Delta t \rightarrow \lambda\Delta t$ . If not broken, this degeneracy, referred to as the mass-sheet transformation (MST), may significantly affect cosmographic inferences, including  $H_0$  (see e.g., Gorenstein et al. 1988; Saha 2000; Wucknitz 2002; Koopmans et al. 2003; Liesenborgs & De Rijcke 2012; Schneider & Sluse 2013; Schneider 2014a,b; Xu et al. 2016). It is worth mentioning that the external convergence  $\kappa_{\text{ext}}$  is based on physical effects whereas the MST (Eq. (4)) stems from a pure mathematical degeneracy (Schneider & Sluse 2013; hereafter SS13). Different solutions have been proposed to reduce the degeneracy induced by the MST in time delay cosmography (see e.g., Birrer et al. 2016; Treu & Marshall 2016, and references therein). A commonly used method consists in assuming a specific lens model, typically a power-law, and independently estimating the lens mass with the measurement of its velocity dispersion. However, SS13 have shown experimentally that two different classes of galaxy models with compatible velocity dispersions were able to reproduce equally well a set of image positions, but predicted significantly different time delays. Furthermore, because the time delay ratios were not constant, they suggested that the transformation between these two models was not exactly an MST but a more general one. This has naturally raised some concerns about the reliability of the  $H_0$  determination from time delay cosmography.

Schneider & Sluse (2014; hereafter SS14) laid the theoretical basis for an approximate invariance transformation, the so-called source-position transformation (SPT), of which the MST is a special case. Unruh et al. (2017; hereafter USS17) explored further its properties, such as defining a criterion to determine whether an SPT is valid or not and exploring the density profile of SPT-modified mass distributions. They also pointed out that the degeneracy found experimentally in SS13 between the two

<sup>1</sup> We note that the time delay distance  $D_{\Delta t}^{\text{model}}$  inferred from a model that neglects the impact of  $\kappa_{\text{ext}}$  is related to the true time delay distance by  $D_{\Delta t}^{\text{model}} = (1 - \kappa_{\text{ext}}) D_{\Delta t}$ .

models mimics an SPT, which thereby confirmed that it occurs in real lens modeling. To date, it is not clear whether the conclusions drawn in SS13 and SS14 about time delays and  $H_0$  could be generalized to other mass distributions modified under exact SPTs or only reflect the behavior of a very special case. In this paper, we address this question by studying how time delays are sensitive to the effects of the SPT.

This paper is organized as follows. For readers who are not familiar with the SPT, we outline its basic principles in Sect. 2. In particular, we recall the importance of identifying a validity criterion. Owing to the valuable insight it offers for more general cases, we consider the SPT-modified time delays for axisymmetric lenses in Sect. 3. For nonaxisymmetric lenses, we discuss in Sect. 4 the relevance of the validity criterion defined in USS17 and analyze the SPT-modified time delays in detail. We summarize our findings and conclude in Sect. 5.

## 2. The principle of the source position transformation

This section focuses on the principle of the SPT and the most recent theoretical results. For a detailed discussion, the reader is referred to SS14 and USS17. All the analytical results presented in this paper have been implemented into a user friendly python package called pySPT. All the numerical results and figures have also been obtained from pySPT without using any extra software. For an overall description of the package, we refer the reader to the companion paper Wertz & Orthen (2018).

The basic idea underlying the SPT can be simply summarized as follows. A given general mass distribution  $\kappa(\theta)$  defines a deflection law  $\alpha(\theta)$  which describes how the light paths are affected in the vicinity of the deflector. The  $n$  lensed image angular positions  $\theta_i$  of a point-like source at unobservable position  $\beta$  are those which satisfy the lens equation  $\beta = \theta_i - \alpha(\theta_i)$ . Then, from astrometric observations, we can infer the constraints

$$\theta_i - \alpha(\theta_i) = \theta_j - \alpha(\theta_j), \quad (5)$$

for all  $1 \leq i < j \leq n$ , leading to the mapping  $\theta_i(\theta_1)$  defined by the relative image positions of the same source. The SPT addresses the following question: can we define an alternative deflection law, denoted as  $\hat{\alpha}(\theta)$ , which preserves the mapping  $\theta_i(\theta_1)$  for a unique source? If such a deflection law exists, it will necessarily be associated with the alternative source position  $\hat{\beta} = \theta_i - \hat{\alpha}(\theta_i)$ , defining a new lens mapping, in such a way that

$$\theta = \beta + \alpha(\theta) = \hat{\beta} + \hat{\alpha}(\theta). \quad (6)$$

An SPT consists in a global transformation of the source plane formally defined by a mapping  $\hat{\beta}(\beta)$  which gives rise to the transformed deflection law:

$$\hat{\alpha}(\theta) = \alpha(\theta) + \beta - \hat{\beta}(\beta) = \alpha(\theta) + \beta - \hat{\beta}(\theta - \alpha(\theta)), \quad (7)$$

where in the first step we used Eq. (6) and in the last step we inserted the original lens equation. The mapping  $\hat{\beta}(\beta)$  is chosen so that it satisfies  $\det(\partial\hat{\beta}/\partial\beta) \neq 0$  for all  $\beta$  in the region of interest, hence  $\hat{\beta}(\beta)$  is one-to-one. This property of the source mapping guarantees the pairing of images to be conserved. With  $\hat{\alpha}$  defined this way, Eq. (6) guarantees that all images of a given source  $\beta$  under the original deflection law  $\alpha(\theta)$  are also images of the source  $\hat{\beta}$  under the modified deflection law  $\hat{\alpha}(\theta)$ . Therefore, the mapping  $\theta_i(\theta_1)$  is preserved for all source positions.

From the Jacobi matrix  $\hat{\mathcal{A}}(\boldsymbol{\theta}) = \partial\hat{\boldsymbol{\beta}}/\partial\boldsymbol{\theta} = (\partial\hat{\boldsymbol{\beta}}/\partial\boldsymbol{\beta})(\partial\boldsymbol{\beta}/\partial\boldsymbol{\theta})$  of the modified lens mapping  $\hat{\boldsymbol{\beta}} = \boldsymbol{\theta} - \hat{\boldsymbol{\alpha}}(\boldsymbol{\theta})$ , SS14 have shown that both the magnification ratios of image pairs and their relative shapes remain unchanged under an SPT. In general, the Jacobi matrix  $\hat{\mathcal{A}}(\boldsymbol{\theta})$  is not symmetric, which indicates that the modified deflection law  $\hat{\boldsymbol{\alpha}}$  is not a curl-free field:

$$|\nabla \times \hat{\boldsymbol{\alpha}}(\boldsymbol{\theta})| = |\hat{\mathcal{A}}_{12}(\boldsymbol{\theta}) - \hat{\mathcal{A}}_{21}(\boldsymbol{\theta})| \neq 0, \quad (8)$$

where the subscript indices refer to the matrix entries. Therefore,  $\hat{\boldsymbol{\alpha}}$  cannot be in general expressed as the gradient of a deflection potential  $\hat{\psi}$  and does not correspond to the deflection produced by a gravitational lens. Thus, there exists no physical mass distribution  $\hat{\kappa}$  leading to the modified deflection law  $\hat{\boldsymbol{\alpha}}$ . The only cases for which  $\hat{\mathcal{A}}(\boldsymbol{\theta})$  is globally symmetric occur either when the SPT simply reduces to an MST  $\hat{\boldsymbol{\beta}}(\boldsymbol{\beta}) = \lambda\boldsymbol{\beta}$ , or when axisymmetric lenses are transformed under SPTs corresponding to a general radial stretching of the form

$$\hat{\boldsymbol{\beta}}(\boldsymbol{\beta}) = [1 + f(|\boldsymbol{\beta}|)]\boldsymbol{\beta}, \quad (9)$$

where  $f$  is called the deformation function. For such cases, we can always define  $\hat{\kappa}$  so that  $2\hat{\kappa} = \nabla \cdot \hat{\boldsymbol{\alpha}} = \nabla^2 \hat{\psi}$ . However, even in this case there is still no guarantee that  $\hat{\kappa}$  corresponds to a physical mass distribution. Depending on the SPT, the modified mass profile may become nonmonotonic or even non-positive definite in particular regions of the lens plane.

Provided the curl component of  $\hat{\boldsymbol{\alpha}}$  is sufficiently small, it was shown in USS17 that one can define a curl-free deflection law  $\tilde{\boldsymbol{\alpha}}$  which is very similar to  $\hat{\boldsymbol{\alpha}}$  in the sense that their difference is smaller than the astrometric accuracy  $\varepsilon_{\text{acc}}$  of current observations:

$$|\tilde{\boldsymbol{\alpha}}(\boldsymbol{\theta}) - \hat{\boldsymbol{\alpha}}(\boldsymbol{\theta})| = |\Delta\boldsymbol{\alpha}(\boldsymbol{\theta})| < \varepsilon_{\text{acc}}, \quad (10)$$

in a finite region  $\mathcal{U}$  where multiple images occur. Therefore,  $\tilde{\boldsymbol{\alpha}}$  can be derived as the gradient of a deflection potential  $\tilde{\psi}$ , which is caused by a mass distribution  $\tilde{\kappa}$  corresponding to a gravitational lens. The central question of the validity of an SPT was addressed in USS17. Whereas  $\hat{\boldsymbol{\alpha}}$  yields exactly the same lensed image positions as the original lens,  $\tilde{\boldsymbol{\alpha}}$  does not. Because of observational uncertainties and additional physical reasons such as substructures in the mass distribution, we cannot reproduce observed positions to better than a few milliarcseconds (mas) with a smooth mass model (for a detailed discussion, see SS14). A given SPT  $\hat{\boldsymbol{\beta}}(\boldsymbol{\beta})$  should be flagged as being valid as long as a corresponding curl-free  $\tilde{\boldsymbol{\alpha}}$  leads to lensed image shifts smaller than a few mas. In this sense, the SPT is only an approximate invariance transformation. The condition (10) was chosen in USS17 as the criterion to assess whether an SPT is valid or not. The relevance of this choice is reconsidered in detail in Sect. 4.1.

Because it will be of practical interest for deriving SPT-modified time delays in the nonaxisymmetric case (see Sect. 4), we recall here the explicit expressions for  $\tilde{\psi}$  and  $\tilde{\boldsymbol{\alpha}} = \nabla\tilde{\psi}$ . These can essentially be obtained by formulating the ‘‘action’’

$$S = \int_{\mathcal{U}} |\nabla\tilde{\psi} - \hat{\boldsymbol{\alpha}}|^2 d^2\boldsymbol{\theta} \quad (11)$$

for which finding a minimum leads to the Neumann problem:

$$\begin{cases} \nabla^2 \tilde{\psi} = \nabla \cdot \hat{\boldsymbol{\alpha}} =: 2\hat{\kappa}, & \text{for all } \boldsymbol{\theta} \in \mathcal{U}, \\ \nabla\tilde{\psi} \cdot \mathbf{n} = \hat{\boldsymbol{\alpha}} \cdot \mathbf{n}, & \text{for all } \boldsymbol{\theta} \in \partial\mathcal{U}, \end{cases} \quad (12)$$

where  $\partial\mathcal{U}$  represents the boundary curve of  $\mathcal{U}$  and  $\mathbf{n}$  the outward directed normal vector. From Eq. (11), we see that the condition (10) plays a central role in defining a curl-free counterpart  $\tilde{\boldsymbol{\alpha}}$  of the SPT-modified deflection law  $\hat{\boldsymbol{\alpha}}$ . We also note that the first relation in Eq. (12) implies  $\tilde{\kappa} = \hat{\kappa}$  for all  $\boldsymbol{\theta} \in \mathcal{U}$ . The Neumann problem can be solved by means of a Green’s function for which an analytical solution is known when  $\mathcal{U}$  is a disk of radius  $R$ . Thus, the deflection potential  $\tilde{\psi}$  evaluated at the position  $\boldsymbol{\vartheta}$  in the lens plane explicitly reads (Unruh et al. 2017; Wertz & Orthen 2018)

$$\tilde{\psi}(\boldsymbol{\vartheta}) = \langle \tilde{\psi} \rangle + 2 \int_{\mathcal{U}} H_1(\boldsymbol{\vartheta}; \boldsymbol{\theta}) \hat{\kappa}(\boldsymbol{\theta}) d^2\boldsymbol{\theta} - \int_{\partial\mathcal{U}} H_2(\boldsymbol{\vartheta}; \boldsymbol{\theta}) \hat{\boldsymbol{\alpha}} \cdot \mathbf{n} ds, \quad (13)$$

where  $\langle \tilde{\psi} \rangle$  is the average of  $\tilde{\psi}$  on  $\mathcal{U}$ ,  $ds$  the line element of the boundary curve  $\partial\mathcal{U}$ ,

$$H_1(\boldsymbol{\vartheta}; \boldsymbol{\theta}) = \frac{1}{4\pi} \left[ \ln \left( \frac{|\boldsymbol{\vartheta} - \boldsymbol{\theta}|^2}{R^2} \right) + \ln \left( 1 - \frac{2\boldsymbol{\vartheta} \cdot \boldsymbol{\theta}}{R^2} + \frac{|\boldsymbol{\vartheta}|^2 |\boldsymbol{\theta}|^2}{R^4} \right) - \frac{|\boldsymbol{\theta}|^2}{R^2} \right], \quad (14)$$

and

$$H_2(\boldsymbol{\vartheta}; \boldsymbol{\theta}) = \frac{1}{4\pi} \left[ 2 \ln \left( \frac{|\boldsymbol{\vartheta} - \boldsymbol{\theta}|^2}{R^2} \right) - 1 \right]. \quad (15)$$

The corresponding deflection angle  $\tilde{\boldsymbol{\alpha}}$  can be derived by obtaining the gradient of  $H_1$  and  $H_2$  with respect to  $\boldsymbol{\vartheta}$ , which reads

$$\begin{aligned} \tilde{\boldsymbol{\alpha}}(\boldsymbol{\vartheta}) = & \frac{1}{\pi} \int_{\mathcal{U}} \left( \frac{\boldsymbol{\vartheta} - \boldsymbol{\theta}}{|\boldsymbol{\vartheta} - \boldsymbol{\theta}|^2} + \frac{|\boldsymbol{\theta}|^2 \boldsymbol{\vartheta} - R^2 \boldsymbol{\theta}}{R^4 - 2R^2 \boldsymbol{\vartheta} \cdot \boldsymbol{\theta} + |\boldsymbol{\vartheta}|^2 |\boldsymbol{\theta}|^2} \right) \hat{\kappa}(\boldsymbol{\theta}) d\boldsymbol{\theta} \\ & - \frac{1}{\pi} \int_{\partial\mathcal{U}} \frac{\boldsymbol{\vartheta} - \boldsymbol{\theta}}{|\boldsymbol{\vartheta} - \boldsymbol{\theta}|^2} \hat{\boldsymbol{\alpha}} \cdot \mathbf{n} ds. \end{aligned} \quad (16)$$

Of course,  $\tilde{\psi}$  and  $\tilde{\boldsymbol{\alpha}}$  depend on the radius  $R$  of the circular region  $\mathcal{U}$  and must satisfy the two conditions  $R > |\boldsymbol{\vartheta}|$  and  $R$  not too large to ensure the criterion (10) to be satisfied.

To quantitatively study the impact of the SPT on time delays, it will be necessary to explicitly define a mapping  $\hat{\boldsymbol{\beta}}(\boldsymbol{\beta})$ . We will focus most of this work on an isotropic SPT described by the radial stretching of the form (9). In particular, we will consider the special case where the deformation function  $f(|\boldsymbol{\beta}|)$  is the lowest-order expansion of more general functions:

$$f(|\boldsymbol{\beta}|) = f_0 + \frac{f_2}{2\theta_E^2} |\boldsymbol{\beta}|^2, \quad (17)$$

where  $f_0 := f(0)$ ,  $f_2 := \theta_E^2 f''(0)$ , and  $\theta_E$  is the Einstein angular radius. When  $f_2 = 0$ , Eq. (17) reduces to  $f_0$  and the radial stretching (9) simplifies to a pure MST with  $\lambda = 1 + f_0$ . Such as in SS14 and USS17, we only consider SPT parameters which yield to physically meaningful modified mass profiles.

### 3. Time delays: the axisymmetric case

Owing to its simplicity, the study of how an SPT affects time delays between lensed images produced by an axisymmetric lens provides a valuable insight on the general nonaxisymmetric case. Since  $\boldsymbol{\alpha}$  and  $\boldsymbol{\theta}$  are collinear, the original lens mapping becomes one-dimensional and reads  $\boldsymbol{\beta} = \boldsymbol{\theta} - \alpha(\boldsymbol{\theta})$ . We set  $\beta > 0$  and only

consider the two outer<sup>2</sup> lensed images  $\theta_A$  and  $\theta_B$  located on opposite sides of the lens center, that is  $\theta_B < 0 < |\theta_B| < \theta_A$ . From Eq. (1), we readily deduce the one-dimensional form of the original time delay  $\Delta t_{AB}$  between the image pair  $(\theta_A, \theta_B)$ :

$$\Delta t_{AB} = \frac{D_{\Delta t}}{c} [\tau(\theta_A) - \tau(\theta_B)] =: \frac{D_{\Delta t}}{c} \Delta \tau_{AB}. \quad (18)$$

The one-dimensional radial stretching (9) simply reads

$$\hat{\beta}(\beta) = [1 + f(\beta)] \beta, \quad (19)$$

where  $f(-\beta) = f(\beta)$  to preserve the symmetry. With no loss of generality,  $1 + f(\beta) + \beta df(\beta)/d\beta > 0$  ensures the SPT to be one-to-one. For the axisymmetric case, the SPT is an exact invariance transformation. Thus, the deflection law  $\hat{\alpha}$  is a curl-free field,  $\hat{\alpha} = \hat{\alpha}$ , and there exists a deflection potential  $\hat{\psi}$  such as

$$\frac{d\hat{\psi}(\theta)}{d\theta} = \hat{\alpha}(\theta) = \alpha(\theta) - f(\beta(\theta)) \beta(\theta), \quad (20)$$

where in the second step we used the one-dimensional form of Eq. (7). From Eq. (1), we deduce that the SPT-modified extra light travel time  $\hat{T}$  reads

$$\hat{T}(\theta) = \frac{D_{\Delta t}}{c} \left[ \frac{1}{2} (\theta - \hat{\beta}[\beta(\theta)])^2 - \hat{\psi}(\theta) \right] =: \frac{D_{\Delta t}}{c} \hat{\tau}(\theta). \quad (21)$$

From Eqs. (3) and (21), the SPT-modified time delay between image pair  $(\theta_A, \theta_B)$  of the same source thus becomes

$$\Delta \hat{t}_{AB} = \hat{T}(\theta_A) - \hat{T}(\theta_B) = \frac{D_{\Delta t}}{c} (\hat{\tau}(\theta_A) - \hat{\tau}(\theta_B)) =: \frac{D_{\Delta t}}{c} \Delta \hat{\tau}_{AB}. \quad (22)$$

With Eqs. (18) and (22), we show in Sect. 3.1 that the time delay ratios  $\Delta \hat{t}_{AB}/\Delta t_{AB} \equiv \Delta \hat{\tau}_{AB}/\Delta \tau_{AB}$  can be highly simplified, revealing an elegant expression in terms of  $\beta$  and  $\hat{\beta}(\beta)$ , and valid for any axisymmetric lens and deformation function  $f(\beta)$ . We also propose an equivalent form of this relation in terms of the original and SPT-modified mean surface mass densities. We illustrate the analytical results with some examples in Sect. 3.2.

### 3.1. The SPT-modified time delays

After substituting the one-dimensional form of Eqs. (6) and (20) into Eq. (21), the SPT-modified extra light travel time reads  $\hat{T}(\theta_i) = \hat{\tau}(\theta_i) D_{\Delta t}/c$  with

$$\hat{\tau}(\theta_i) = \frac{1}{2} [\alpha(\theta_i) - f(\beta(\theta_i)) \beta(\theta_i)]^2 - \psi(\theta_i) + \int_0^{\theta_i} f(\beta(\theta)) \beta(\theta) d\theta, \quad (23)$$

up to a constant independent of  $\theta$ , keeping in mind that  $\beta(\theta) = \theta - \alpha(\theta)$ . Because of  $\beta(\theta_A) = \beta(\theta_B) =: \beta_s$ , we have  $f(\beta(\theta_A)) = f(\beta(\theta_B)) = f(\beta_s)$ , and the SPT-modified time delays between the images  $\theta_A$  and  $\theta_B$  is given by  $\Delta \hat{t}_{AB} = \Delta \hat{\tau}_{AB} D_{\Delta t}/c$  with

$$\Delta \hat{\tau}_{AB} = \Delta \tau_{AB} - f(\beta_s) \beta_s (\theta_A - \theta_B) + \int_{|\theta_B|}^{\theta_A} f(\beta(\theta)) \beta(\theta) d\theta. \quad (24)$$

Due to the lens symmetry, the integral over  $[\theta_B, |\theta_B|]$  does not contribute to  $\Delta \hat{\tau}_{AB}$ . With no loss of generality, we thus integrate

<sup>2</sup> The inner lensed image is most of the time not observed.

from  $|\theta_B|$  instead of  $\theta_B$  in Eq. (24). To go a step further, the difference  $\Delta \tau_{AB}$  between the original Fermat potentials can also be written as

$$\Delta \tau_{AB} = -\beta_s (\theta_A - \theta_B) + \frac{1}{2} (\theta_A^2 - \theta_B^2) - (\psi(\theta_A) - \psi(\theta_B)), \quad (25)$$

$$= -\beta_s (\theta_A - \theta_B) + \int_{|\theta_B|}^{\theta_A} \beta(\theta) d\theta, \quad (26)$$

where in the first step we used the original lens equations  $\alpha(\theta_A) = \theta_A - \beta_s$  and  $\alpha(\theta_B) = \theta_B - \beta_s$ , and in the last step we used  $d\psi(\theta)/d\theta = \theta - \beta(\theta)$ . Combining Eqs. (24) and (26), we then obtain from Eq. (22) the SPT-modified time delay:

$$\Delta \hat{t}_{AB} = \Delta t_{AB} [1 + f(\beta_s)] + \frac{D_{\Delta t}}{c} \varepsilon_{AB}, \quad (27)$$

where we define  $\varepsilon_{AB}$  as

$$\varepsilon_{AB} = \int_{|\theta_B|}^{\theta_A} \beta(\theta) [f(\beta(\theta)) - f(\beta_s)] d\theta. \quad (28)$$

For the special case of a pure MST, the deformation function  $f$  is independent of  $\beta$ , namely  $f(\beta(\theta)) = f(\beta_s) \equiv \lambda - 1$  with  $\lambda \in \mathbb{R}$ . Therefore,  $\varepsilon_{AB} = 0$  and we find  $\Delta \hat{t}_{AB} = \lambda \Delta t_{AB}$  for all axisymmetric lenses, as expected. Considering the radial stretching (19) and a singular isothermal sphere (SIS) lens model, we show explicitly in Appendix A that  $\varepsilon_{AB} = 0$  also holds for all image pairs  $(\theta_A, \theta_B)$ , that is for  $0 \leq \beta < \theta_E$ . In fact, simple analytical arguments reveal that, in general,  $\varepsilon_{AB}$  remains very small compared to the other terms in Eq. (27) and can be neglected. The demonstration is explained in detail in Appendix B. As a result, the time delay ratios  $\Delta \hat{t}/\Delta t$  given in Eq. (27) can be simply approximated by

$$\frac{\Delta \hat{t}}{\Delta t} \approx 1 + f(\beta) \equiv \frac{\hat{\beta}(\beta)}{\beta}, \quad (29)$$

where we have dropped the subscript AB keeping in mind that the equation holds only for time delay ratios between the same pair of lensed images corresponding to the source  $\hat{\beta}$  and  $\beta$ . For a given radial stretching, Eq. (29) shows that the ratios between SPT-modified and original time delays scale basically like  $\hat{\beta}/\beta$ , implying that they depend explicitly on the deformation function  $f(\beta)$ , as it is the case for the MST.

As written, Eq. (29) misleadingly suggests that the time delay ratio is insensitive to the original lens profile  $\kappa$ . Consider two original radial mass profiles  $\kappa^{(1)}$  and  $\kappa^{(2)}$ , which are not related under an SPT, and consider a source position  $\beta$ . We locate the corresponding pairs of brighter lensed images by  $(\theta_A^{(1)}, \theta_B^{(1)})$  and  $(\theta_A^{(2)}, \theta_B^{(2)})$ . For a given deformation function  $f(\beta)$ , Eq. (29) says that  $\Delta \hat{t}_{AB}^{(1)}/\Delta t_{AB}^{(1)} \approx \Delta \hat{t}_{AB}^{(2)}/\Delta t_{AB}^{(2)}$ , but the two time delay ratios are evaluated at two different pairs of positions which depend on the lens models, that is  $\theta_A^{(1)} \neq \theta_A^{(2)}$  and  $\theta_B^{(1)} \neq \theta_B^{(2)}$ . When  $\kappa^{(2)}$  corresponds to a modified version of  $\kappa^{(1)}$  under the SPT  $\hat{\beta}(\beta) = [1 + g(\beta)] \beta$  (with  $g(\beta)$  satisfying the conditions given after Eq. (19)), we have  $\theta_A^{(1)} = \theta_A^{(2)}$  and  $\theta_B^{(1)} = \theta_B^{(2)}$ . However, this case can be reduced to an original radial mass profile  $\kappa^{(1)}$  deformed by an SPT that is defined as the composition of two other SPTs such as  $\hat{\beta}(\beta) = [1 + h(\beta)] \beta$  with  $h(\beta) = [1 + f(\beta)][1 + g(\beta)] - 1$ . Thus, this leads to  $\Delta \hat{t}/\Delta t \approx 1 + h(\beta)$ , in agreement with Eq. (29).

The SPT-modified mass profile  $\hat{\kappa}$  of a radial profile  $\kappa$  is also radial (SS14). Therefore, time delays  $\Delta t_{AB}$  and  $\Delta \hat{t}_{AB}$  should

depend only on the image positions and the corresponding surface mass densities in the annulus defined between the images. In particular for  $\Delta t_{AB}$ , the major contribution comes from the mean surface mass density  $\langle \kappa \rangle_{AB}$  in the annulus  $|\theta_B| < \theta < \theta_A$  (Gorenstein et al. 1988; Kochanek 2002, 2006). We will show next that the time delay ratios (29) can be expressed only in terms of  $\langle \kappa \rangle_{AB}$  and the corresponding SPT-modified  $\langle \hat{\kappa} \rangle_{AB}$ . First, we can easily show that

$$\begin{aligned} \langle \kappa \rangle_{AB} &:= \frac{2}{\theta_A^2 - \theta_B^2} \int_{|\theta_B|}^{\theta_A} \theta \kappa(\theta) d\theta = \frac{m(\theta_A) - m(|\theta_B|)}{\theta_A^2 - \theta_B^2} \\ &= 1 - \frac{\beta_s}{\theta_A - |\theta_B|}, \end{aligned} \quad (30)$$

where in the last step we used  $m(\theta) = \theta \alpha(\theta)$ ,  $\alpha(\theta_A) = \theta_A - \beta_s$ , and  $\alpha(\theta_B) = \theta_B - \beta_s$ . Similarly, we can easily deduce that

$$\langle \hat{\kappa} \rangle_{AB} = 1 - \frac{\hat{\beta}(\beta_s)}{\theta_A - |\theta_B|}, \quad (31)$$

where we first used  $\hat{m}(\theta) = \theta \hat{\alpha}(\theta)$  and Eq. (20), then  $\hat{\alpha}(\theta_A) = \theta_A - \beta_s$  and  $\hat{\alpha}(\theta_B) = \theta_B - \beta_s$ . Combining Eqs. (29) and (31), we thus obtain for the time delay ratio in terms of mean surface mass densities:

$$\frac{\Delta \hat{t}}{\Delta t} \approx \frac{\hat{\beta}(\beta)}{\beta} = \frac{1 - \langle \hat{\kappa} \rangle}{1 - \langle \kappa \rangle}, \quad (32)$$

where we have once again dropped the AB keeping in mind that the mean surface mass densities are evaluated in the annulus defined by the inner and outer radii  $|\theta_B|$  and  $\theta_A$ , respectively. As expected, Eq. (32) shows that the ratio between SPT-modified and original time delays depends essentially on mean surface mass densities in the annulus  $|\theta_B| < \theta < \theta_A$ . Finally, we note that the second equality in Eq. (32) is exact.

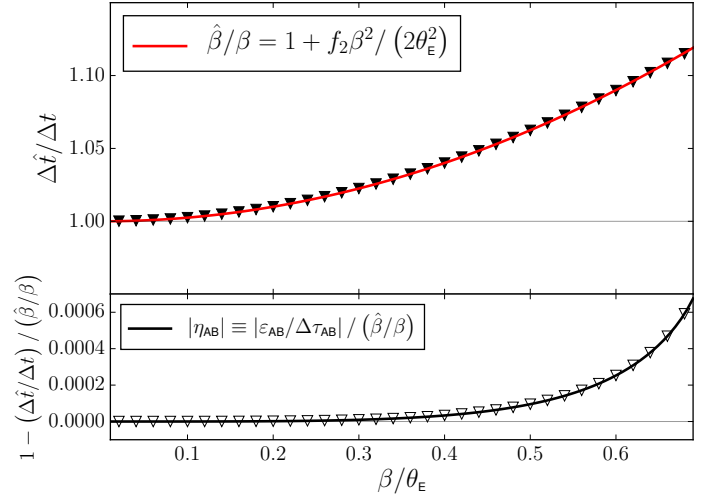
### 3.2. Some illustrative examples

To illustrate the results obtained in the previous section, we first consider the deformation function (17) with  $f_2 = 0.5$  and  $f_0 = 0$  to separate the impact of the MST from that of the SPT. As original lens model, we choose a non-singular isothermal sphere (NIS) characterized by the deflection law:

$$\alpha(\theta) = \frac{\theta_E \theta}{\sqrt{\theta_c^2 + \theta^2}}, \quad (33)$$

where the core radius  $\theta_c$  is defined such as  $\theta_c = \nu \theta_E$  with  $0 < \nu < 1$ . To derive time delays in the axisymmetric case, we only consider the three lensed image configurations where the fainter central image is omitted. Thus, we need to sample the source positions inside the radial caustic of angular radius  $\beta_r = \beta(\theta_r)$  where  $\theta_r$  represents the angular radius of the corresponding radial critical curve. For an NIS,  $\beta_r$  is simply given by  $\beta_r = \theta_E (1 - \nu^{2/3})^{3/2} \approx 0.695 \theta_E$  for  $\nu = 0.1$ . Using this simple lens model and pySPT, we create pairs of mock images for a uniform set of 34 sources covering the range  $\beta = 0.02 \theta_E$  to  $\beta = 0.68 \theta_E < \beta_r$ .

The top panel in Fig. 1 shows  $\Delta \hat{t}/\Delta t$  as a function of  $\beta$  for the corresponding pairs of lensed images. We see that the time delay ratios scale remarkably well like the function  $1 + f(\beta) \equiv \hat{\beta}(\beta)/\beta$ , as predicted by Eq. (29). According to Eq. (29), the strongest effect of the SPT on time delays arises for a source as close as possible to the radial caustic, that is for  $\beta \rightarrow \beta_r$ . Thus, in



**Fig. 1.** Impact of an SPT described by the radial stretching  $\hat{\beta}(\beta) = 1 + f_2 \beta^2 / (2 \theta_E^2)$ , with  $f_2 = 0.5$ , on time delays image pairs generated by a NIS, with core  $\theta_c = 0.1 \theta_E$ . *Top panel:* ratio between SPT-modified and original time delays (black inverted triangles) for each source position. The time delay ratios scale almost perfectly like  $\hat{\beta}/\beta$  (red curve), as predicted by Eq. (29). *Bottom panel:* numerical confirmation of the validity of the approximation performed in Eq. (29) which consists in neglecting the term  $\varepsilon_{AB}$ . The solid black curve illustrates  $|\eta_{AB}|$  as an analytical function of  $\beta$  and perfectly fits the quantity  $1 - (\Delta \hat{t}/\Delta t)/(\hat{\beta}/\beta)$  numerically evaluated for each source position.

our first example, the theoretical maximum time delay ratio equals  $\Delta \hat{t}/\Delta t \approx 1.12$  for  $f_2 = 0.5$  and  $\nu = 0.1$ , leading to an impact of 12% on  $H_0$ . As long as it leads to a physical meaningful  $\hat{\kappa}$ , larger (resp. smaller) values of  $|f_2|$  lead to larger (resp. smaller) time delay ratios. To quantitatively evaluate the accuracy of Eq. (29), we compare the numerically evaluated quantity  $1 - (\Delta \hat{t}/\Delta t)/(\hat{\beta}/\beta)$  to unity, as shown in the bottom panel in Fig. 1. From Eq. (27), it follows that  $1 - (\Delta \hat{t}/\Delta t)/(\hat{\beta}/\beta) = |\eta_{AB}|$  with  $\eta_{AB} := (\varepsilon_{AB}/\Delta\tau_{AB})/(\hat{\beta}/\beta_s)$ . The quantity  $|\eta_{AB}|$  is smaller than  $10^{-4}$  for  $\beta \leq 0.5 \theta_E$ , reaching a maximum of  $|\eta_{AB}| \approx 6 \times 10^{-4} \ll 1$  for  $\beta = 0.68 \theta_E$ , confirming that  $\varepsilon_{AB}$  can be neglected in Eq. (29) in such a case. For an NIS deformed by a radial stretching characterized by Eq. (17), it is possible to derive an analytical solution for  $\varepsilon_{AB}$ , hence for  $|\eta_{AB}|$ , by solving Eq. (28). This analytical solution is represented in the bottom panel in Fig. 1 and fits perfectly the numerical evaluations of  $1 - (\Delta \hat{t}/\Delta t)/(\hat{\beta}/\beta)$  at each source position, as expected.

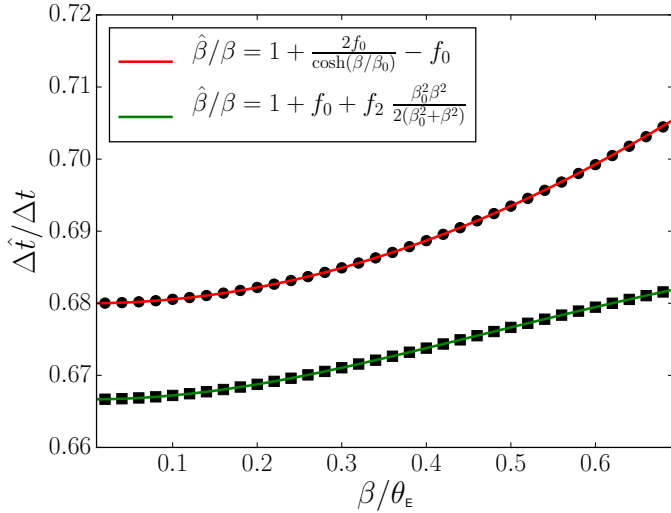
We have successfully tested the relation (29) for various axisymmetric lens profiles deformed by different deformation functions. As additional examples, we consider the two deformation functions:

$$f(\beta) = \frac{2f_0}{\cosh(\beta/\beta_0) - f_0}, \quad (34)$$

with  $\beta_0 = \theta_E \sqrt{3(1-f_0)/(1+f_0)}$  and  $f_0 = -0.32$ , and

$$f(\beta) = f_0 + \beta_0^2 f_2 \beta^2 \left[ 2(\beta_0^2 + \beta^2) \right], \quad (35)$$

with  $\beta_0 = 0.8 \theta_E$ ,  $f_0 = -1/3$ , and  $f_2 = 1/9$ . The choice for the two deformation functions (34) and (35) is justified by the fact that the resulting SPT-modified mass profiles  $\hat{\kappa}$  are approximately power laws near the tangential critical curve, that is  $\hat{\kappa}(\theta) \approx \hat{\kappa}(\theta_E)(\theta/\theta_E)^{-\nu}$  (SS14). In both cases, we adopt an NIS with  $\theta_c = 0.1 \theta_E$  as original lens model and the same source sample as



**Fig. 2.** Impact on time delays of two different SPTs defined such that the corresponding  $\hat{\kappa}$  is approximately a power law near the tangential critical curve. As predicted by Eq. (29), the time delay ratios (black dots and squares) scale almost perfectly like  $\hat{\beta}/\beta$  (red and green curves).

in the first example. Figure 2 shows the time delay ratios between the lensed images for each source. As expected,  $\Delta\hat{t}/\Delta t$  fits almost perfectly the function  $1 + f(\beta) \equiv \hat{\beta}(\beta)/\beta$ . For  $\beta = 0$ , the two deformation functions simplify to  $f(\beta) = f_0$  and the corresponding SPTs reduce to pure MSTs leading to  $\Delta\hat{t}/\Delta t = 0.68$  and  $\Delta\hat{t}/\Delta t = 2/3$ , respectively. Therefore, any changes from these values reflect the impact of the SPT. For  $\beta = 0.68 \theta_E$ , Fig. 2 shows an impact of around 3.6% and 2.2% on  $H_0$ , respectively, which is significantly smaller than what we have obtained for the first example.

Not all combinations of SPT deformation parameters and original mass profiles  $\kappa$  yield a physically meaningful SPT-modified mass profile, namely  $\hat{\kappa}$  monotonically decreasing and positive definite (Schneider & Sluse 2014). In addition, the maximum time delay ratio also depends on  $\kappa$  since the latter directly defines the size of the radial caustic  $\beta(\theta_r)$ , namely the region in the source plane that produces multiple images. In summary, the way the SPT affects the time delays is very sensitive to the choice of the deformation function  $f$ , the associated deformation parameters, the original mass profile  $\kappa$  and lensed image positions. For these reasons, we restrain ourselves to draw generalized quantitative conclusions in the axisymmetric case. However, our numerical tests suggest an effect of a few percent in general. We will show in the next section that the simple connection between the time delay ratios and the source position ratios may still be very strong in the nonaxisymmetric case.

#### 4. Time delays: the nonaxisymmetric case

In this section, we drop the axisymmetry assumption for the original lens model. The SPT-modified deflection law  $\hat{\alpha}$  is thus not a curl-free field in general and there exists no deflection potential  $\hat{\psi}$  which satisfies  $\nabla\hat{\psi} = \hat{\alpha}$ . To define a physically meaningful modified extra light travel time, we consider the deflection law  $\tilde{\alpha}$ , the closest curl-free approximation to  $\hat{\alpha}$  which satisfies the criterion (10) for all  $\theta$  over a region  $\mathcal{U}$  (see Eq. (16)), and the associated deflection potential  $\tilde{\psi}$  (see Eq. (13)). For the rest of this section, we follow USS17 and consider  $\varepsilon_{\text{acc}} \approx 5 \times 10^{-3} \theta_E$  over the circular region  $|\theta| \leq 2 \theta_E$ , where the approximation stems from the typical positional accuracy of the *Hubble* Space Telescope.

Within the region  $\mathcal{U}$ , the lensed images  $\tilde{\theta}$  of the source  $\hat{\beta}$  satisfying the lens mapping  $\hat{\beta} = \tilde{\theta} - \tilde{\alpha}(\tilde{\theta})$  are expected to be sufficiently close to the corresponding original images  $\theta = \beta + \alpha(\theta) = \hat{\beta} + \hat{\alpha}(\theta)$  to not be distinguished observationally. However, we show in Sect. 4.1 that the criterion (10) defined in USS17 cannot guarantee the difference  $|\Delta\theta| := |\tilde{\theta} - \theta|$  between the SPT-modified image position  $\tilde{\theta}$  of the source  $\hat{\beta}$  and the image position  $\theta$  of the source  $\beta$  to be smaller than  $\varepsilon_{\text{acc}}$  over the whole region  $\mathcal{U}$ . However, for specific pairs of original and SPT-modified sources leading to indistinguishable image configurations, we illustrate in Sect. 4.2 the typical behavior of the time delay ratios. Finally, based solely on a numerical optimization, we slightly modify the source mapping  $\hat{\beta}(\beta)$  by relaxing the isotropic condition of the SPT. It follows that the region where  $|\Delta\theta| < \varepsilon_{\text{acc}}$  can be substantially extended. From this ad hoc source mapping,  $\tilde{\alpha}$  and  $\tilde{\psi}$ , we illustrate the corresponding alternative time delay ratios in Sect. 4.3.

##### 4.1. Criterion for the validity of an SPT

To illustrate the limit of the criterion  $|\Delta\alpha(\theta)| < \varepsilon_{\text{acc}}$ , we consider a situation similar to SS14 and USS17, namely a quadrupole lens composed of an NIS plus external shear  $\gamma_p$  (NISg) for which the deflection law is defined by

$$\alpha(\theta) = \frac{\theta_E \theta}{\sqrt{\theta_c^2 + |\theta|^2}} - \gamma_p \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \theta, \quad (36)$$

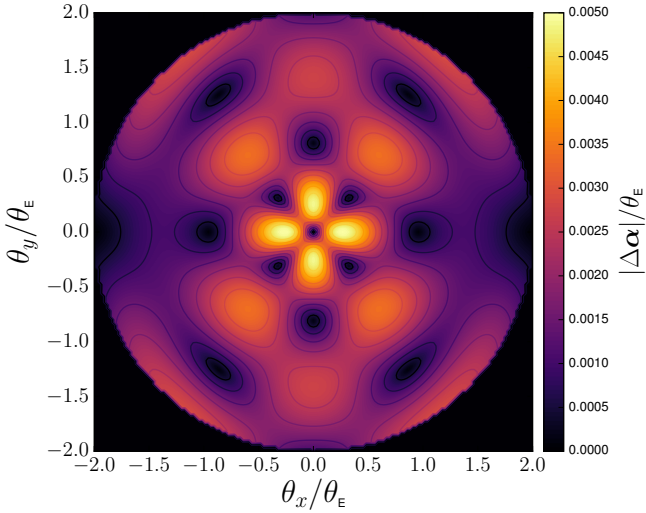
where the core radius is set to  $\theta_c = 0.1 \theta_E$ . The original source mapping is transformed by a radial stretching (9) with a deformation function  $f$  of the form (17). The adopted SPT is thus defined by

$$\hat{\beta}(\beta) = \left( 1 + f_0 + \frac{f_2}{2 \theta_E^2} |\beta|^2 \right) \beta. \quad (37)$$

For the rest of this section, we set the deformation parameter  $f_2$  to be  $f_2 = 0.4$  and the external shear magnitude to be  $\gamma_p = 0.1$ . Furthermore, we exclude the effect of a pure MST by simply choosing  $f_0 = 0$ . According to USS17 (see their Fig. 4), this specific pair  $(f_2, \gamma_p)$  constitutes an allowed pair of parameters in a sense it fulfills the criterion  $|\Delta\alpha(\theta)| < \varepsilon_{\text{acc}}$  over the circular region  $|\theta| \leq 2 \theta_E$ . Figure 3 shows the map  $|\Delta\alpha(\theta)|$  over a circular grid  $|\theta| \leq 2 \theta_E$  in the lens plane. This figure is similar to the map  $|\Delta\alpha(\theta)|$  illustrated in Fig. 7 in USS17, although they used  $f_2 = 0.55$ . It turns out that the ratio  $|\Delta\alpha(\theta_i)|/|\Delta\alpha(\theta_j)|$  remains unaffected when  $f_2$  varies, but is sensitive to variations of  $R$  or  $\gamma_p$ . Actually,  $\gamma_p$  is the only parameter that explicitly characterizes the degree of asymmetry of the original lens model.

The next step consists in determining how well the deflection law  $\tilde{\alpha}$  allows us to reproduce the original lensed images of a source. To this aim, we create a set of mock images  $\theta$  of a sample of sources  $\beta$  that cover the first quadrant of the source plane. We restrict the grid of sources to  $0 \leq |\beta| \leq 0.66 \theta_E$  where multiple images are produced (see top panel in Fig. 4). Then, we produce the images  $\tilde{\theta}$  of the corresponding SPT-modified sources  $\hat{\beta}(\beta)$ . The bottom panel in Fig. 4 shows the image positions  $\theta$  and the color-coding represents  $|\Delta\theta|$  in units of  $\theta_E$ . The same color-coding is applied to the sources (top panel) where only the largest offset, denoted by  $|\Delta\theta|_{\text{max}}$ , are shown.

An “unexpected” conclusion can be drawn from Fig. 4; even though  $|\Delta\alpha(\theta)| < \varepsilon_{\text{acc}}$  over the region  $|\theta| \leq 2 \theta_E$  (as shown in Fig. 3), many of image configurations are characterized by



**Fig. 3.** Map of  $|\Delta\alpha(\theta)|$  over the circular region  $|\theta| \leq 2\theta_E$  for  $f_2 = 0.4$ ,  $\theta_c = 0.1\theta_E$ , and  $\gamma_p = 0.1$ . This figure is similar to Fig. 7 in Unruh et al. (2017) with  $f_2 = 0.55$ , even though it is based on a different approach (see the text for more details).

$|\Delta\theta| \gg \varepsilon_{\text{acc}}$  for at least one lensed image. This implies that these image configurations can be distinguished from the original ones and the corresponding SPT can no longer be flagged as valid. Furthermore, the largest offsets  $|\Delta\theta|$  occur near the tangential critical curve. It comes with no surprise that the corresponding regions in the source plane are thus located near the tangential caustic curve. To address this behavior, we first consider what the quantity  $|\Delta\alpha(\theta)|$  really represents. As defined in Eq. (10), both  $\hat{\alpha}$  and  $\tilde{\alpha}$  are evaluated at the same position  $\theta$  in the lens plane. Therefore, we have  $\hat{\alpha}(\theta) = \theta - \hat{\beta}$  and  $\tilde{\alpha}(\theta) = \theta - \tilde{\beta}$  where  $\tilde{\beta}$  is the source position of the image  $\theta$  under the deflection law  $\tilde{\alpha}$ . Combining the two latter equations leads to

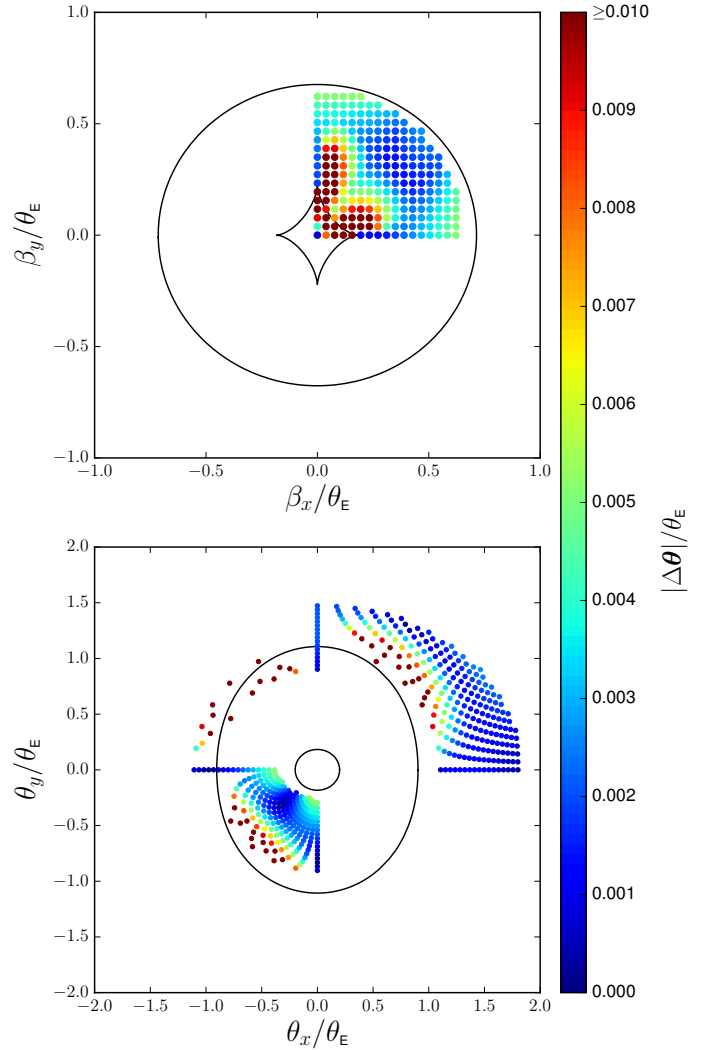
$$\Delta\alpha(\theta) = \hat{\beta}(\theta) - \tilde{\beta}(\theta) =: \Delta\beta. \quad (38)$$

Equation (38) shows that minimizing  $|\Delta\alpha(\theta)|$  is equivalent to minimizing  $|\Delta\beta|$  with no guarantee on  $|\Delta\theta|$ . Indeed, let us consider a position  $\theta$  close to a critical line for which  $|\Delta\alpha(\theta)| < \varepsilon_{\text{acc}}$ , for example  $(\theta_x/\theta_E, \theta_y/\theta_E) = (0.5, 1.0)$  (see Fig. 3 and bottom panel in Fig. 4). The corresponding source  $\beta(\theta)$  is necessarily close to a caustic, so is  $\hat{\beta}(\theta)$ . Thus, the source  $\hat{\beta}(\theta)$  lies in a region of the source plane where even small shifts  $|\Delta\beta|$  can lead to significantly different image positions. This explains why regions where  $|\Delta\theta| \gg \varepsilon_{\text{acc}}$  are those which surround the critical curves. Furthermore, whereas  $\theta_i - \hat{\alpha}(\theta_i) = \theta_j - \hat{\alpha}(\theta_j)$  is satisfied for all  $i \leq j$ , we have  $\theta_i - \tilde{\alpha}(\theta_i) \neq \theta_j - \tilde{\alpha}(\theta_j)$ , meaning that the  $\theta_i$  are not lensed images of a unique source under the deflection law  $\tilde{\alpha}$ . Thus, the criterion  $|\Delta\alpha(\theta_i)| < \varepsilon_{\text{acc}}$  for a lensed image configuration  $\theta_i$  is based upon positions that are not linked under the deflection law  $\tilde{\alpha}$ . These few simple arguments suggest with no loss of generality that the choice of Eq. (10) as a validity criterion may not be the most appropriate one.

Let us now evaluate  $\tilde{\alpha}$  at the position  $\tilde{\theta} = \hat{\beta} + \tilde{\alpha}(\tilde{\theta})$  instead of  $\theta$  and consider the difference  $\tilde{\alpha}(\tilde{\theta}) - \hat{\alpha}(\theta)$ . We readily find that

$$|\tilde{\alpha}(\tilde{\theta}) - \hat{\alpha}(\theta)| = |\tilde{\theta} - \theta| = |\Delta\theta|, \quad (39)$$

which corresponds exactly to the image shift induced by the SPT that we expect to be smaller than  $\varepsilon_{\text{acc}}$ . Assuming that  $|\Delta\theta|$  is



**Fig. 4.** *Top panel:* grid of source positions  $\beta$  covering the radial range  $0 \leq |\beta| \leq 0.66\theta_E$  in the first quadrant of the source plane. Each source produces a set of lensed images  $\theta$  (shown in the *bottom panel*) under the original deflection law  $\alpha$  and the corresponding set of  $\theta$  under the curl-free deflection law  $\tilde{\alpha}$ . The color-coding refers to the largest offset  $|\Delta\theta|_{\text{max}}$  associated with each source. The solid black curves locate the caustic curves (*top panel*) and the critical curves (*bottom panel*), respectively, for the NIS plus external shear. *Bottom panel:* set of mock lensed images  $\theta$  produced by the source positions shown in the *top panel* and lensed by an NIS plus external shear characterized by  $\theta_c = 0.1\theta_E$  and  $\gamma_p = 0.1$ . The color-coding represents the offsets  $|\Delta\theta|$  in units of  $\theta_E$  between  $\theta$  and the images  $\theta$ .

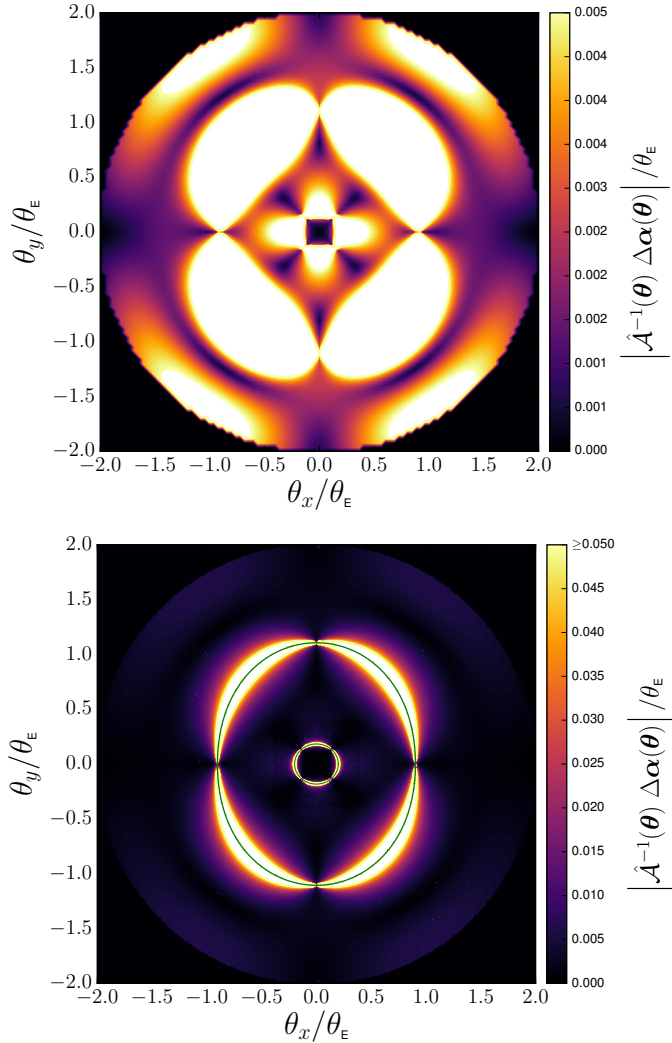
small, we can show to first order that

$$\begin{aligned} \hat{\beta} &= \tilde{\theta} - \tilde{\alpha}(\tilde{\theta}), \\ &= \theta + \Delta\theta - \hat{\alpha}(\theta + \Delta\theta) - \Delta\alpha(\theta + \Delta\theta), \\ &\approx \theta - \hat{\alpha}(\theta) + \left(1 - \frac{\partial\hat{\alpha}}{\partial\theta}\right) \Delta\theta - \Delta\alpha(\theta). \end{aligned} \quad (40)$$

Thus, for all positions  $|\theta| \leq 2\theta_E$  not located on a critical curve, Eq. (40) leads to

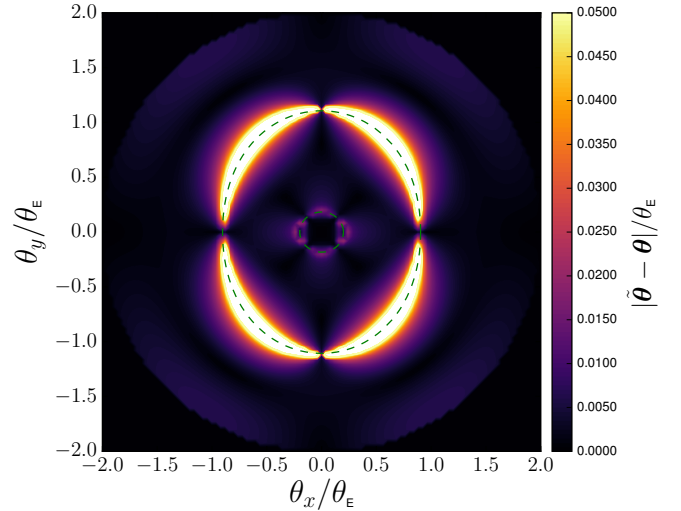
$$|\Delta\theta| \approx |\hat{\mathcal{A}}^{-1}(\theta) \Delta\alpha(\theta)|. \quad (41)$$

Equation (41) clearly shows that the offsets  $|\Delta\theta|$  are related to  $|\Delta\alpha(\theta)|$  through the SPT-modified Jacobi matrix  $\hat{\mathcal{A}}(\theta)$  whose



**Fig. 5.** Maps of  $|\hat{\mathcal{A}}^{-1}(\theta) \Delta\alpha(\theta)| \approx |\Delta\theta|$  in units of  $\theta_E$  for  $f_2 = 0.4$  and  $\gamma_p = 0.1$ . *Top panel:* we use the same color-coding as in Fig. 3 to explicitly show that  $|\Delta\alpha(\theta)| < 5 \times 10^{-3} \theta_E$  over a region of the lens plane does not guarantee the image offsets  $|\Delta\theta|$  to be smaller than  $5 \times 10^{-3} \theta_E$  over the same region. *Bottom panel:* we adjust the color-coding to bring out regions where the image offsets  $|\Delta\theta|$  are the largest, namely the critical curves.

impact become larger as we get closer to the critical curves. Figure 5 illustrates the quantity  $|\hat{\mathcal{A}}^{-1}(\theta) \Delta\alpha(\theta)|$  using two different color-coding. The upper panel shows the same color-coding as used in Fig. 3 for comparison. It shows that a significant part of the region  $|\theta| < 2\theta_E$  is characterized by  $|\Delta\theta| > \varepsilon_{\text{acc}}$  even though  $|\Delta\alpha(\theta)| < \varepsilon_{\text{acc}}$ . The lower panel adopts a color-coding which allows us to better visualize regions characterized by the largest offsets. These regions surround the two critical curves represented by the two green lines. We confirm the validity of the first order Eq. (41) by comparing Fig. 5 with Fig. 6, which represents explicitly the quantity  $|\tilde{\theta} - \theta|$ . As expected, small differences can be observed very close to the critical curves where higher order terms in Eq. (40) become significant and cannot be ignored. In addition, Fig. 6 is much more time consuming to obtain than Fig. 5. Indeed, a single  $|\tilde{\theta} - \theta|$  evaluation requires  $\tilde{\theta}$  to be calculated first, that is solving the lens equation  $\hat{\beta} = \tilde{\theta} - \tilde{\alpha}(\tilde{\theta})$  that implies numerous  $\tilde{\alpha}$  evaluations. In contrast, a single Eq. (41) evaluation requires only one  $\tilde{\alpha}$  evaluation. For this reason, the grid density in bottom panel in Fig. 5 is 20 times higher than in Fig. 6.



**Fig. 6.** Map of  $|\tilde{\theta} - \theta| =: |\Delta\theta|$  in units of  $\theta_E$  for  $f_2 = 0.4$  and  $\gamma_p = 0.1$ . Clear differences with Fig. 5 are observed for positions  $\theta$  located almost on the radial critical curve, where the approximation adopted in Eq. (41) is expected to become not valid. This map is much more time consuming to obtain than the ones represented in Fig. 5.

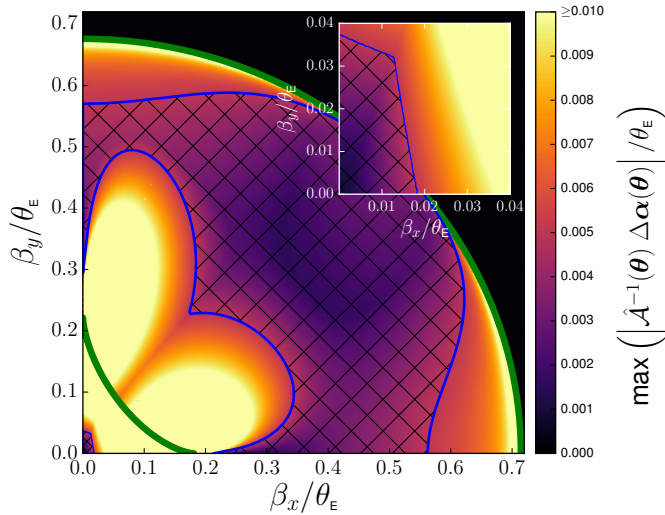
Equation (41) confirms that the criterion  $|\Delta\alpha(\theta)| < \varepsilon_{\text{acc}}$  for the validity of an SPT cannot guarantee the angular separation  $|\Delta\theta|$  to be smaller than the astrometric accuracy of current observations, at least in regions nearby critical curves. To construct the curl-free deflection field  $\tilde{\alpha}$ , USS17 have considered the action defined in Eq. (11) for which they found a minimum. This approach is based explicitly on the validity criterion (10), which is not satisfactory and should be reconsidered. A new appropriate criterion would of course imply the definition of a new action to be minimized, leading to a new definition for  $\tilde{\alpha}$  and  $\tilde{\psi}$ . Such a new approach is beyond the scope of this paper and will not be addressed here. Nevertheless, it remains possible to quantitatively estimate the impact of the SPT on time delays with the means available. In the next section, we first focus on the subset of source positions  $\beta$  depicted in the top panel in Fig. 4 that yields  $|\Delta\theta| < \varepsilon_{\text{acc}}$ .

#### 4.2. The SPT-modified time delays for valid configurations

In the previous section, we have shown that the criterion defined in Eq. (10) does not guarantee  $|\Delta\theta| < \varepsilon_{\text{acc}}$  for all  $|\theta| \leq 2\theta_E$ . However, the top panel in Fig. 4 also shows sources (mainly outside the tangential caustic curve) for which the corresponding largest offsets  $|\Delta\theta|_{\text{max}}$  between original and SPT-modified image configurations are smaller than  $\varepsilon_{\text{acc}}$ . Adopting the same original lens model as in the previous section, Fig. 7 shows the quantity

$$|\Delta\theta(\beta)|_{\text{max}} := \max(|\hat{\mathcal{A}}^{-1}(\theta(\beta)) \Delta\alpha(\theta(\beta))|), \quad (42)$$

in units of  $\theta_E$  over the first quadrant in the source plane. The region outside the radial caustic curve is irrelevant in our case because it does not lead to multiple image configurations. The two disjointed hatched regions denoted by  $B_1$  and  $B_2$  are defined such as all sources inside  $B_1 \cup B_2$  lead to image configurations characterized by  $|\Delta\theta|_{\text{max}} \leq \varepsilon_{\text{acc}}$ . For  $f_2 = 0.4$  and  $\gamma_p = 0.1$ , the region  $B_1 \cup B_2$  covers around 56% of the area enclosed by the radial caustic curve. Smaller values for  $f_2$  yield larger  $B_1 \cup B_2$  areas, up to 100% when  $f_2 = 0$  (SPT reduced to an MST) or  $\gamma_p = 0$  (axisymmetric lens). The regions  $B_1$  and  $B_2$  are situated on both sides of the tangential caustic curve. The very high



**Fig. 7.** Map of  $|\Delta\theta(\boldsymbol{\beta})|_{\max}$  in units of  $\theta_E$  for  $f_2 = 0.4$  and  $\gamma_p = 0.1$  (see Eq. (42)). The inner (resp. outer) green line represents the tangential (resp. radial) caustic curve. The two hatched regions ( $B_1$  and  $B_2$ ) delimited by blue curves demarcate parts of the source plane where  $|\Delta\theta(\boldsymbol{\beta})|_{\max} \leq 5 \times 10^{-3} \theta_E$ . The region  $B_1$  lies inside the tangential caustic curve while  $B_2$  lies outside. The inset highlight the region around the position  $\boldsymbol{\beta} = \mathbf{0}$ .

area ratio between these two regions (1–560 in this case) indicates that most of the valid image configurations are composed of two images (the fainter third central one is always omitted). Moreover, the few valid four component configurations are very symmetric, suggesting comparable time delays between opposite image pairs.

Provided that  $\hat{\kappa}$  is physically meaningful, the curl-free deflection field  $\tilde{\alpha}$  yields indistinguishable image configurations for sources  $\boldsymbol{\beta} \in B_1 \cup B_2$  as compared to the original  $\boldsymbol{\alpha}$ . Although these valid image configurations are of limited interest for time delay cosmography<sup>3</sup>, the resulting model ambiguities may still prevent us from performing a robust lens modeling. Thus, even though the adopted SPT is not valid over all the region  $|\boldsymbol{\theta}| \leq 2\theta_E$ , we propose in this section to analyze the time delay ratios of these particular image configurations between the original and SPT-modified models. To this aim, we consider an original non-axisymmetric mass distribution which produces  $n$  lensed images  $\boldsymbol{\theta}_i$  of a source  $\boldsymbol{\beta} \in B_1 \cup B_2$ . The time delay  $\Delta t_{ij}$  between a pair of lensed images  $\boldsymbol{\theta}_i$  and  $\boldsymbol{\theta}_j$  is defined in Eq. (3). The corresponding SPT-modified time delay  $\tilde{\Delta t}_{ij}$  have to be evaluated at image positions  $\tilde{\boldsymbol{\theta}}_i$  and  $\tilde{\boldsymbol{\theta}}_j$ , respectively, leading to

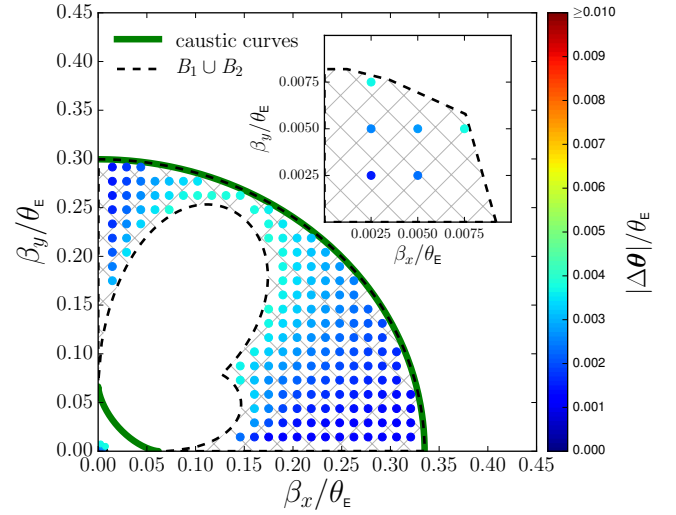
$$\tilde{\Delta t}_{ij} = \tilde{T}(\tilde{\boldsymbol{\theta}}_i) - \tilde{T}(\tilde{\boldsymbol{\theta}}_j) = \frac{D_{\Delta t}}{c} [\tilde{\tau}(\tilde{\boldsymbol{\theta}}_i) - \tilde{\tau}(\tilde{\boldsymbol{\theta}}_j)] =: \frac{D_{\Delta t}}{c} \Delta \tilde{\tau}_{ij}, \quad (43)$$

where the SPT-modified Fermat potential is defined by

$$\tilde{\tau}(\tilde{\boldsymbol{\theta}}) = \frac{1}{2} [\tilde{\boldsymbol{\theta}} - \hat{\boldsymbol{\beta}}(\tilde{\boldsymbol{\theta}})]^2 - \tilde{\psi}(\tilde{\boldsymbol{\theta}}) = \frac{1}{2} |\tilde{\boldsymbol{\alpha}}(\tilde{\boldsymbol{\theta}})|^2 - \tilde{\psi}(\tilde{\boldsymbol{\theta}}). \quad (44)$$

We present here the representative results obtained for two classes of models: the quadrupole NISg as defined in Eq. (36) and a non-singular isothermal elliptical lens (NIE). The NIE

<sup>3</sup> The sources located inside the region  $B_1$  produce very symmetric quadruply imaged configurations, while those located inside the region  $B_2$  produce only doubled image configurations. In both cases, only one relevant time delay can be inferred from these systems.



**Fig. 8.** Grid of sources  $\boldsymbol{\beta}$  located inside the region  $B_1 \cup B_2$  for the NIE with  $(\theta_c, \epsilon) = (0.1 \theta_E, 0.15)$ . The inner (resp. outer) green line represents the tangential (resp. radial) caustic curve. The color-coding refers to the offsets  $|\Delta\theta|/\theta_E$  between the lensed images  $\boldsymbol{\theta}$  of the sources  $\boldsymbol{\beta}$  and the lensed images  $\tilde{\boldsymbol{\theta}}$  of the SPT-modified sources  $\hat{\boldsymbol{\beta}}$ .

surface mass density  $\kappa$  is intrinsically nonaxisymmetric and is defined by (see e.g., Keeton 2001)

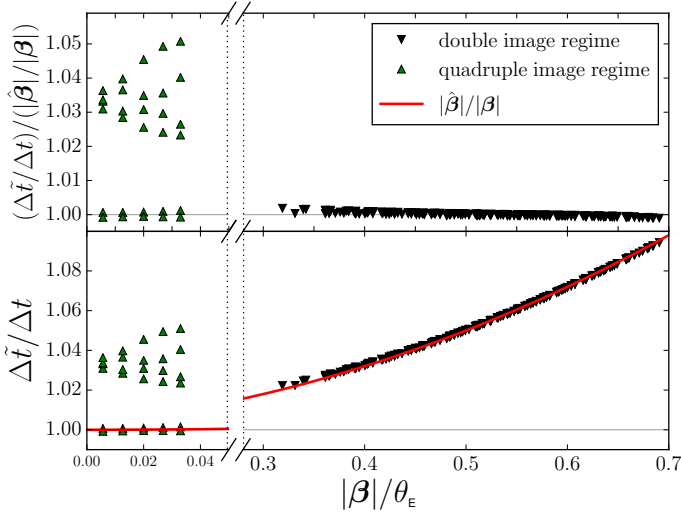
$$\kappa(\boldsymbol{\theta}) = \frac{\theta_E}{2\sqrt{\theta_c^2 + \rho^2}}, \quad (45)$$

where the variable  $\rho$ , constant on ellipses with axis ratio  $q = \sqrt{(1-\epsilon)/(1+\epsilon)}$ , is characterized by

$$\rho = \sqrt{\frac{\theta_x^2}{1-\epsilon} + \frac{\theta_y^2}{1+\epsilon}}. \quad (46)$$

For the rest of this section, the quadrupole model parameters are set to  $(\theta_c, \gamma_p) = (0.1 \theta_E, 0.1)$  and the NIE model parameters to  $(\theta_c, \epsilon) = (0.1 \theta_E, 0.15)$ . These models are deformed by an SPT corresponding to a radial stretching defined in Eq. (37) with  $f_0 = 0$  and  $f_2 = 0.4$ . In both cases, we used pySPT to create mock images  $\boldsymbol{\theta}$  of two separated grids of sources  $\boldsymbol{\beta}$ , which cover together the corresponding regions  $B_1 \cup B_2$ . The size and shape of  $B_1 \cup B_2$  are defined by both the model and SPT parameters, hence differ from the quadrupole to the NIE (see Fig. 7 for the quadrupole and Fig. 8 for the NIE). Making use of Eq. (16), we compute the corresponding images  $\tilde{\boldsymbol{\theta}}$  of the SPT-modified sources  $\hat{\boldsymbol{\beta}}$ . We finally derive the time delay  $\Delta t_{ij}$  and  $\tilde{\Delta t}_{ij}$  and represent their ratios in Fig. 9 (for the NISg) and Fig. 10 (for the NIE). Because of their similarities, we discuss the NISg and the NIE simultaneously.

In the double image regime, the time delay ratios scale almost like the ratios  $|\hat{\boldsymbol{\beta}}|/|\boldsymbol{\beta}|$ . The color-coding in Fig. 10 refers to the azimuth angle of  $\boldsymbol{\beta}$ , denoted as  $\angle\boldsymbol{\beta}$ . Even though the dispersion is slightly larger for the NIE, the deviations from  $|\hat{\boldsymbol{\beta}}|/|\boldsymbol{\beta}|$  do not exceed 0.5% in all cases. The inset in Fig. 10 clearly shows that the dispersion of the time delay ratios is the effect of the relative direction of  $\boldsymbol{\beta}$  with respect to the orientation of the axis of the elliptical iso-density contours (here equal to  $0^\circ$ ). The deviations from  $|\hat{\boldsymbol{\beta}}|/|\boldsymbol{\beta}|$  are maximum for  $\angle\boldsymbol{\beta} = 0^\circ$  and  $\angle\boldsymbol{\beta} = 90^\circ$ , and minimum for  $\angle\boldsymbol{\beta} \approx 45^\circ$ . A similar behavior is observed for the NISg, but with respect to the orientation of the external shear (also equal to  $0^\circ$ ). We suggest that Eq. (29), valid for



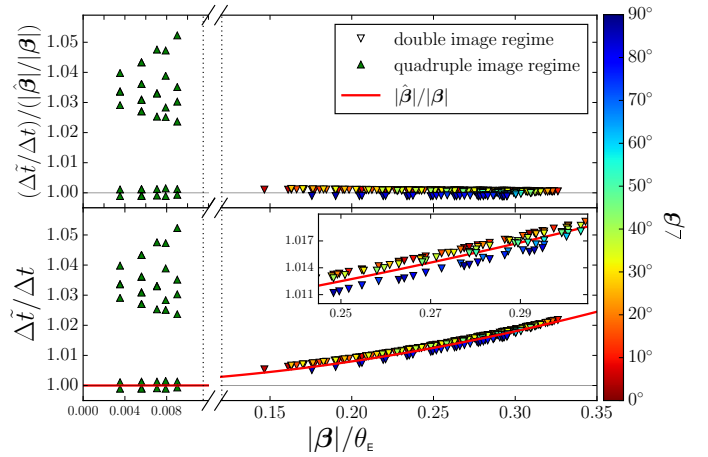
**Fig. 9.** Time delay ratios of image pairs between the NISg and the corresponding SPT-modified model. The model parameters are  $(\theta_c, \gamma_p) = (0.1\theta_E, 0.1)$  and the radial stretching is characterized by  $f_2 = 0.4$ . *Top panel:*  $\Delta\tilde{t}/\Delta t$  normalized by the ratio  $|\hat{\beta}|/|\beta|$  is close to 1 in the double image regime and for opposite images in the quadruple image regime. *Bottom panel:* the impact of the SPT on the time delays is around a few percent, reaching a maximum of 12% for the particular case of a source located almost on the radial caustic curve but still inside  $B_2$ .

the axisymmetric case (see Sect. 3.1), may also be valid in the nonaxisymmetric case for sufficiently large values of  $|\beta|$ ,

$$\frac{\Delta\tilde{t}}{\Delta t} \approx 1 + f(|\beta|) = \frac{|\hat{\beta}(\beta)|}{|\beta|}. \quad (47)$$

Actually, even for the most unfavorable cases, Eq. (47) provides at least a fairly good estimate of  $\Delta\tilde{t}/\Delta t$ . It turns out that these two particular examples are representative of the time delay ratio behavior for double image configurations produced by a non-axisymmetric lens. Thus, the impact of the SPT in the double image regime comes mainly from the ratios  $|\hat{\beta}|/|\beta|$ , in the same way as for the axisymmetric case. In particular, the largest time delay ratios  $(\Delta\tilde{t}/\Delta t)_{\max}$  (considering only the double image configurations for now) is obtained for the source position  $\beta \in B_2$  characterized by the largest radial coordinate  $|\beta|$  and denoted as  $\beta_{\max}$ . Therefore,  $(\Delta\tilde{t}/\Delta t)_{\max}$  depends on  $\beta_{\max}$  and the latter depends on both the deformation function and the original lens model parameters, which define the size of  $B_2$ . For the NISg model depicted in Fig. 9, we find  $\beta_{\max}^{\text{NISg}} \approx (0.562, 0.413)\theta_E$ ,  $|\beta_{\max}^{\text{NISg}}| \approx 0.697\theta_E$ , leading to  $(\Delta\tilde{t}/\Delta t)_{\max}^{\text{NISg}} \approx 1.121$ , that is an impact of around 12% on  $H_0$ . For the NIE model depicted in Fig. 10, we find  $\beta_{\max}^{\text{NIE}} \approx (0.335, 0.0)\theta_E$ ,  $|\beta_{\max}^{\text{NIE}}| \approx 0.335\theta_E$ , leading to  $(\Delta\tilde{t}/\Delta t)_{\max}^{\text{NIE}} \approx 1.028$ , that is an impact of around 3% on  $H_0$ . Similar to the axisymmetric case, the impact of the SPT on time delays may substantially vary according to the nature of the original lens model.

A different behavior is observed for the case of the quadruple image regime. As first pointed out in SS13 from a pure empirical case, the time delay ratios of image pairs between the original and SPT-modified models are not conserved, that is  $(\Delta\tilde{t}/\Delta t)_{ij} \neq (\Delta\tilde{t}/\Delta t)_{ik}$  with  $i < k \leq 4$ . For this reason, even though only three independent time delays can be obtained from a quadruple image configurations, we represent in Figs. 9 and 10 the time delay ratios for all six image permutations

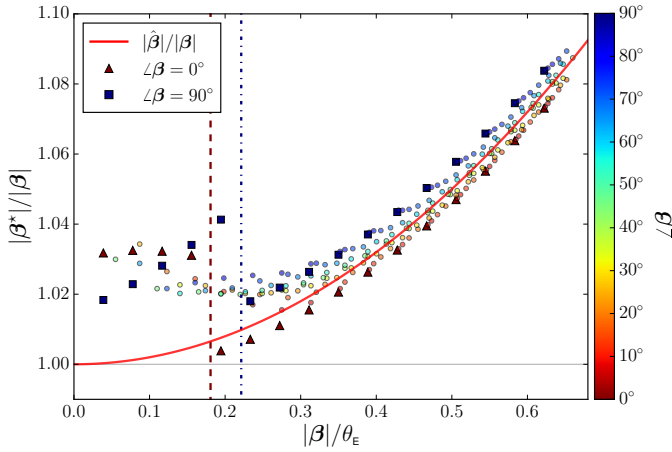


**Fig. 10.** Time delay ratios of image pairs between the NIE and the corresponding SPT-modified model. The model parameters are  $(\theta_c, \epsilon) = (0.1\theta_E, 0.15)$  and the radial stretching is characterized by  $f_2 = 0.4$ . The time delay ratios in the quadruple (resp. double) image regime are depicted with triangles (resp. inverted triangles). The red line shows the source position ratios  $|\hat{\beta}|/|\beta|$  and the color-coding refers to the azimuth angle  $\angle\beta$ . *Top panel:* even though the dispersion is slightly larger compared to the quadrupole model,  $\Delta\tilde{t}/\Delta t$  normalized by the ratio  $|\hat{\beta}|/|\beta|$  is still close to 1 in the double image regime and for opposite images in the quadruple image regime. *Bottom panel:* the time delay ratio dispersion is clearly related to  $\angle\beta$  (see the text for more details). The impact of the SPT on the time delays is also around a few percent, reaching a maximum of around 5% for a particular combination of non-opposite images of a source located inside  $B_1$ .

$(i, j) \in [(1, 2), (1, 3), (1, 4), (2, 3), (2, 4), (3, 4)]$ . We note that the criterion chosen for ordering the images (1–4) is the extra light travel time, from smallest to largest. The pair of opposite images, namely  $(\theta_1, \theta_2)$  and  $(\theta_3, \theta_4)$ , leads to  $\Delta\tilde{t}/\Delta t$  close to 1, regardless of the adopted original lens model we have tested (see the pairs of green triangles close to 1 in bottom panels in Figs. 9 and 10). Owing to the symmetry of the image configurations ( $B_1 \ni \beta \sim \mathbf{0}$ ), the time delays  $\Delta t_{12}$  and  $\Delta t_{34}$  are smaller than the time delays between other image combinations, tending towards 0 when  $\beta$  approaches  $\mathbf{0}$ . The same holds true for the SPT-modified time delays, while we note that  $\Delta\tilde{t}_{12} \approx \Delta t_{12}$  and  $\Delta\tilde{t}_{34} \approx \Delta t_{34}$ . For sources  $\beta \in B_1$ , the mean impact of the SPT, denoted as  $\langle \Delta\tilde{t}/\Delta t \rangle$ , is around of a few percent for both the NISg and the NIE. In contrast to the NISg, the impact of the SPT for the case of the NIE is larger in the quadruple image regime than in the double image regime. This only reflects that  $1 + f(|\beta_{\max}^{\text{NIE}}|) \lesssim \langle \Delta\tilde{t}/\Delta t \rangle^{\text{NIE}}$  while  $1 + f(|\beta_{\max}^{\text{NISg}}|) \gg \langle \Delta\tilde{t}/\Delta t \rangle^{\text{NISg}}$ .

### 4.3. The alternative SPT-modified time delays

In Sect. 4.1, we have shown that the capability of the deflection law  $\tilde{\alpha}$  to predict the same multiple images as predicted by  $\alpha$  (with an accuracy of  $\varepsilon_{\text{acc}}$ ) is very limited (see Eq. (41)). Based upon a representative example, Fig. 7 shows that only a very small region ( $B_1$ ) in the source plane leads to indistinguishable quadruple image configurations. In this section, we investigate a method to extend the region in the lens plane where the offsets  $|\Delta\theta|$  are smaller than  $\varepsilon_{\text{acc}}$ . The idea consists in finding source positions  $\beta^*$  in the vicinity of  $\hat{\beta}(\beta)$  that lead to new image positions  $\theta^* = \beta^* + \tilde{\alpha}(\theta^*)$  in such a way that the offsets  $|\Delta\theta^*| := |\theta^* - \theta|$  are as small as possible. The search for each  $\beta^*$  is based on the



**Fig. 11.** Ratios  $|\beta^*|/|\beta|$  plotted against  $\beta$  in units of  $\theta_E$ . The source positions  $\beta^*$  are obtained by means of a numerical optimization of the cost function  $h(\beta^*)$  defined in Eq. (48). The resulting source mapping  $\beta^*(\beta)$  is slightly anisotropic:  $|\beta^*|$  is larger than  $|\hat{\beta}|$  for  $\angle\beta > 45^\circ$  and smaller for  $\angle\beta < 45^\circ$ . The color-coding refers to the azimuth angle  $\angle\beta$  in the source plane. When a source crosses the tangential caustic curve, the mapping shows  $\beta^*(\beta)$  discontinuities. Two particular jumps are highlighted: (1) sources depicted with triangles pass by the cusp located on the  $\beta_x$ -axis (the corresponding  $\beta_x/\theta_E$  is identified by the dashed vertical line); and (2) sources depicted with squares pass by the cusp located on the  $\beta_y$ -axis (the corresponding  $\beta_y/\theta_E$  is identified by the dash-dotted line.)

numerical minimization of the cost function  $h(\beta^*)$  defined by

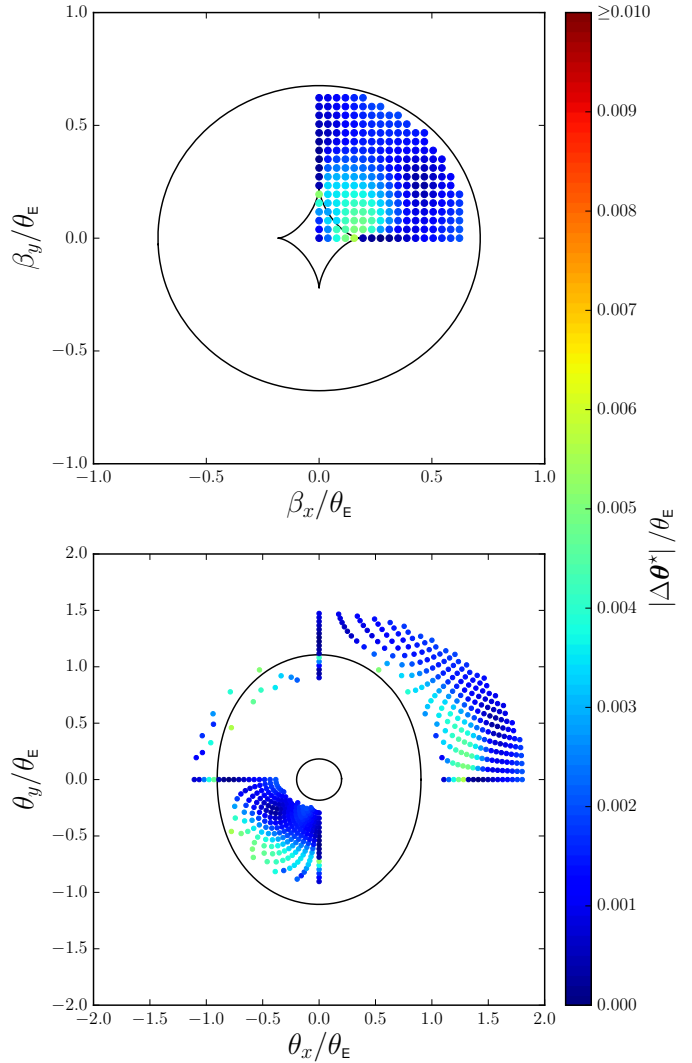
$$h(\beta^*) = \max\left(\left|\theta_i^*(\beta^*) - \theta_i(\hat{\beta})\right|\right), \quad (48)$$

using the Levenberg–Marquardt algorithm (Levenberg 1944). Because we expect  $\beta^*$  to be close to the corresponding  $\hat{\beta}$ , we always choose the latter as first guesses while we do not restrict  $\beta^*$  to share the same direction as  $\hat{\beta}$ . Thus, the resulting source mapping  $\beta^*(\hat{\beta})$ , hence  $\beta^*(\beta)$ , may not be isotropic as for the radial stretching.

As we shall see, this approach may drastically increase the region  $B_1 \cup B_2$  (in particular  $B_1$ ) while benefiting from a simple implementation. However, we must point out that this method suffers several flaws. First, the source mapping  $\beta^*(\beta)$  lacks a solid analytical basis. While  $\hat{\beta}(\beta)$  gives rise to  $\hat{\alpha}$  which is analytically connected to  $\tilde{\alpha}$  by definition<sup>4</sup>, there is no apparent link between  $\tilde{\alpha}$  and  $\beta^*$ . Moreover, the way it is obtained precludes any further analytical investigation. Second, there is no definitive guarantee for the source mapping  $\beta^*(\beta)$  to be one-to-one over  $\mathcal{U}$ . Finally, a successful minimization of the cost function  $h$  for a given  $\beta^*$  does not guarantee  $|\Delta\theta^*|$  to be smaller than  $\varepsilon_{\text{acc}}$ . Indeed, the solution  $\beta^*$  only corresponds to the one for which the cost function  $h(\beta^*)$  is the smallest in the vicinity of  $\hat{\beta}$ , being potentially larger than  $\varepsilon_{\text{acc}}$ . This is particularly true for sources which are located very close to the caustic curves. At least, we have  $|\Delta\theta^*| \leq |\Delta\theta|$  where the equality holds when  $\beta^* = \hat{\beta}$ . For these reasons, we point out that this numerical approach cannot supplant the analytical reconsideration of how the curl-free deflection law  $\tilde{\alpha}$  is defined. However, the combination of  $\beta^*(\beta)$  and  $\tilde{\alpha}$  constitutes a physically meaningful alternative to  $\beta$  and  $\alpha$ , and deserves to be considered.

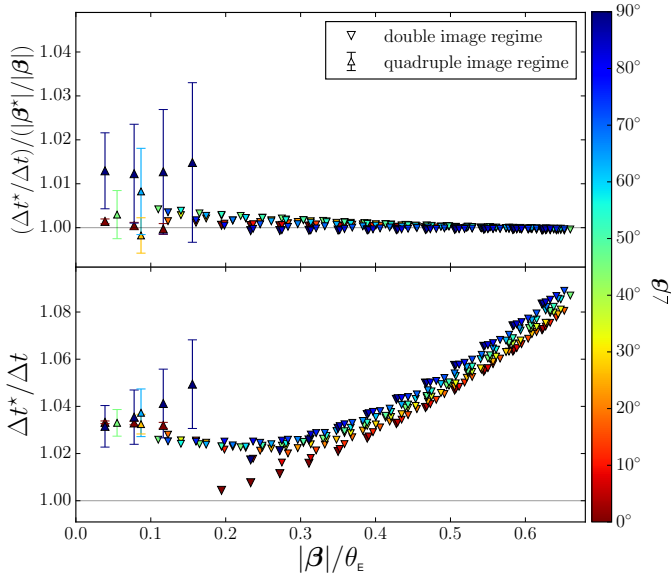
To illustrate the method, we adopt the same lens model and SPT as in Sect. 4.1. We also consider the same grid of sources

<sup>4</sup> We also recall the analytical relation  $\nabla \cdot \tilde{\alpha} = \nabla \cdot \hat{\alpha} = 2\hat{\kappa}$  over  $\mathcal{U}$ .



**Fig. 12.** *Bottom panel:* set of mock lensed images  $\theta$  produced by the source positions shown in the top panel and lensed by an NIS plus external shear characterized by  $\theta_c = 0.1\theta_E$  and  $\gamma_p = 0.1$ . The color-coding represents the image offsets  $|\Delta\theta^*|$  in units of  $\theta_E$  between  $\theta$  and  $\theta^*$ . The latter are the images of the source positions  $\beta^*$  that are lensed by the SPT-modified lens associated with  $\tilde{\alpha}$ . The sources  $\beta^*$  result from the numerical optimization of the cost function  $h(\beta^*)$ . Most of the new image configurations are now characterized by  $|\Delta\theta^*| \leq \varepsilon_{\text{acc}}$ . *Top panel:* grid of source positions  $\beta$  covering the radial range  $0 \leq |\beta| \leq 0.66\theta_E$  in the first quadrant of the source plane. Each source produces a set of lensed images  $\theta$  (shown in the *bottom panel*) under the original deflection law  $\alpha$  and the corresponding set of  $\theta^*$  under the curl-free deflection law  $\tilde{\alpha}$ . The color-coding refers to the largest offset  $|\Delta\theta^*|_{\text{max}} \equiv h(\beta^*)$  associated with each source  $\beta^*$ . The solid black curves locate the caustic curves (*top panel*) and the critical curves (*bottom panel*), respectively, for the NIS plus external shear.

$\beta$  covering the first quadrant of the source plane and restricted to  $0 \leq |\beta| \leq 0.66\theta_E$ . We illustrate the results of the numerical optimization in Figs. 11 and 12. The ratio  $|\beta^*|/|\beta|$  plotted against  $|\beta|/\theta_E$  in Fig. 11 clearly shows the slight anisotropy of the source mapping  $\beta^*(\beta)$  resulting from the numerical optimization. Indeed, for original sources located on a quarter circle with a radius  $|\beta|$ , the corresponding  $|\beta^*|$  depend on the azimuth angle  $\angle\beta$ ;  $|\beta^*|$  is larger than  $|\hat{\beta}|$  for  $\angle\beta > 45^\circ$  and smaller for  $\angle\beta < 45^\circ$ . For sources leading to double image configurations, the  $|\beta^*|$  scatter is around 1%. Furthermore, discontinuities in the mapping  $\beta^*(\beta)$  appear when a source crosses the tangential caustic curve. Two particular jumps are highlighted for sources passing through



**Fig. 13.** Time delay ratios of image pairs between the NISg and the corresponding SPT-modified model associated with the curl-free deflection field  $\tilde{\alpha}$ . The model parameters are  $(\theta_c, \gamma_p) = (0.1 \theta_E, 0.1)$  and the radial stretching defining the source positions  $\beta$  is characterized by  $f_2 = 0.4$ . The time delays  $\Delta t_{ij}^*$  are evaluated for pair of images  $\theta_i^*(\beta^*)$  and  $\theta_j^*(\beta^*)$ . Properties of the source positions  $\beta^*$  are shown in Fig. 11. *Top panel:*  $\Delta t^*/\Delta t$  normalized by the ratio  $|\beta^*|/|\beta|$  is close to 1 in the double image regime and for all combination  $i$  and  $j$  of images  $\theta$ , in the quadruple image regime, when the azimuth angle  $\angle\beta = 0^\circ$ . As  $\angle\beta$  increases, the time delay ratios  $\Delta t^*/\Delta t$  deviates from  $|\beta^*|/|\beta|$  and the corresponding  $\sigma(\Delta t^*/\Delta t)$  also increases. *Bottom panel:* the impact of the SPT on the time delays is around a few percent, reaching around 8% for source positions  $\beta$  located almost on the radial caustic curve.

the two cusps. The ones located on the  $\beta_x$ -axis and  $\beta_y$ -axis are explicitly depicted in Fig. 11. Due to these discontinuities, an extended source which crosses the tangential caustic curve is not mapped smoothly onto an SPT-modified extended source. This effect propagates to the image plane, but the impact on the corresponding lensed image is not observable, as shown in Fig. 12. Similar to Fig. 4, the bottom panel in Fig. 12 represents the image positions  $\theta$  and the color-coding shows the offsets  $|\Delta\theta^*|$  in unit of  $\theta_E$ . The same color-coding is also applied to the sources where only the largest offset between the corresponding pairs of lensed images are shown (see top panel in Fig. 12). Figure 12 shows that almost all the offsets  $|\Delta\theta^*|$  are now smaller than  $\varepsilon_{\text{acc}}$ , even for sources located inside the tangential caustic curve. It is worth stating that a finer source grid would have led to a larger number of sources located very close to the caustic curves, for which the optimized cost function may be larger than  $\varepsilon_{\text{acc}}$ . Compared to Fig. 4, some image positions depicted in Fig. 12 show an offset  $|\Delta\theta^*|$  (after the optimization process) larger than the offset  $|\Delta\theta|$  (before the optimization process). For example, the image position  $\theta_k = (1.242, 0) \theta_E$  of the source  $\beta = (0.117, 0) \theta_E$  is characterized by  $|\Delta\theta_k| = 0.001 \theta_E$  while  $|\Delta\theta_k^*| = 0.005 \theta_E$ . This behavior stems from the fact that, for a given  $n$ -image configuration, the optimization process minimizes only the largest offset but not all the  $n$  offsets simultaneously (because of the  $\max(\cdot)$  function in Eq. (48)). Thus, while the largest offset becomes smaller, the offset  $|\Delta\theta_k^*|$  also varies during the optimization process, leading at the end to  $|\Delta\theta_k| < |\Delta\theta_k^*| \leq h(\beta^*)$ .

Now that we have obtained a large set of indistinguishable image configurations, we derive the time delays  $\Delta t_{ij}$  and  $\Delta t_{ij}^*$  between image pairs where  $\Delta t_{ij}^*$  is defined by

$$\Delta t_{ij}^* = \tilde{T}(\theta_i^*) - \tilde{T}(\theta_j^*) = \frac{D\Delta t}{c} [\tilde{\tau}(\theta_i^*) - \tilde{\tau}(\theta_j^*)]. \quad (49)$$

Figure 13 plots the time delay ratios  $\Delta t^*/\Delta t$  (bottom panel) and the time delay ratios normalized by  $|\beta^*|/|\beta|$  (top panel), both against  $|\beta|/\theta_E$ . The same color-coding as in Fig. 11 is also applied to the time delay ratios. In the double image regime, the normalized time delay ratios  $(\Delta t^*/\Delta t)/(|\beta^*|/|\beta|)$  is close to 1. This suggests once again that the time delay ratios scale almost like the source ratios  $|\beta^*|/|\beta|$ , even though the source mapping is not perfectly isotropic. Thus, the equation

$$\frac{\Delta t^*}{\Delta t} \approx \frac{|\beta^*(\beta)|}{|\beta|} \quad (50)$$

still holds for the slight anisotropic source mapping  $\beta^*(\beta)$  and is particularly true for large  $|\beta|$  inside the radial caustic curve. In the quadruple image regime, the time delay ratios of image pairs between the models are also not perfectly conserved. In either panels in Fig. 13, we represent the mean between the six time delay ratios and the corresponding standard deviation, denoted as  $\sigma(\Delta t^*/\Delta t)$  and depicted with error bars. However,  $\sigma(\Delta t^*/\Delta t)$  is around a minimum of 0.001 for  $\angle\beta = 0^\circ$  when  $\beta$  points towards the same direction as the external shear. Conversely, it reaches a maximum of 0.02 for  $\angle\beta = 90^\circ$  when  $\beta$  points perpendicularly to the direction of the external shear. In addition for  $\angle\beta = 0^\circ$ , the time delay ratios  $\Delta t^*/\Delta t$  scale like  $|\beta^*|/|\beta|$ , regardless of the image combinations. Thus, as the azimuth angle  $\angle\beta$  increases, the time delay ratios  $\Delta t^*/\Delta t$  deviates from  $|\beta^*|/|\beta|$  and the corresponding  $\sigma(\Delta t^*/\Delta t)$  also increases. We recall that prior to any numerical optimization, the time delay ratios  $\Delta\tilde{t}/\Delta t$  were close to 1 only for pairs of opposite images. Quantitatively, the impact of the SPT on the time delays agrees with what we have obtained in Figs. 9 and 10, that is an impact of a few percent. It turns out that most of the conclusions drawn in Sect. 4.2 seem to apply in this case as well.

In Sect. 4.1 we suggest to reconsider the criterion for the validity of an SPT as formulated in Unruh et al. (2017). This shall lead to more appropriate definition of the SPT-transformed deflection law and potential, enlarging the scope where the SPT is a valid invariance transformation. Because Eq. (50) holds within the numerical approach described in this section, we expect that Eq. (47) will also remain true when considering new definitions for  $\tilde{\alpha}$  and  $\tilde{\psi}$ , and an isotropic SPT. Thus, most of the conclusions drawn in this paper should remain unchanged.

## 5. Conclusions

The SPT corresponds to a global invariance transformation of the source plane. It gives rise to a new deflection law, which leaves almost all the lensing observables invariant. We have studied the impact of the SPT on the time delays between pairs of lensed images produced by both axisymmetric and nonaxisymmetric lenses. Although we have mainly considered the case of an isotropic SPT described by a radial stretching of the form  $\hat{\beta} = [1 + f(|\beta|)]\beta$ , we have also addressed a particular case for which the SPT is slightly anisotropic.

Owing to its simplicity, we were able to deal with the axisymmetric case in an analytical way. We have shown that the time delay ratios of image pairs between the SPT-modified and original models approximate as the ratios between the SPT-modified and original source positions, namely  $\Delta\hat{t}/\Delta t \approx \hat{\beta}/\beta$ . Based on simple analytical arguments, we have demonstrated that this relation holds for any axisymmetric lenses and even

deformation function  $f$ . In particular, this relation is exact when the SPT reduces to an MST or when the lens is described by an SIS model, regardless of the deformation function  $f$ . For an NIS model deformed by a radial stretching of the form  $\hat{\beta}(\beta) = [1 + f_2 \beta^2 / (2 \theta_E^2)] \beta$ , we have shown that  $\Delta \hat{t} / \Delta t$  deviates no more than a few hundredth of percent from  $\hat{\beta} / \beta$ . In addition, we have demonstrated that the source mapping can be expressed in terms of the mean surface mass densities, that is  $\hat{\beta} / \beta = (1 - \langle \hat{\kappa} \rangle) / (1 - \langle \kappa \rangle)$ .

Quantitatively, the impact of the SPT on time delays may reach a few percent for particular image configurations, and depends on various factors. Indeed,  $\hat{\beta}$  depends on the choice made for the deformation function  $f$  and its corresponding deformation parameters. Not all combinations of an original mass profile  $\kappa$ , a function  $f$ , and a set of deformation parameters yield a physically meaningful SPT-modified mass profile  $\hat{\kappa}$ . It means that the parameter validity range of a given SPT, hence the validity range of  $\hat{\beta}$  or  $\langle \hat{\kappa} \rangle$ , needs to be studied on a case-by-case basis.

When we drop the axisymmetry assumption for the original lens model, the SPT-modified deflection angle  $\hat{\alpha}$  is not a curl-free field caused by a mass distribution corresponding to a gravitational lens (SS14). An alternative deflection field  $\tilde{\alpha}$  was proposed in USS17, namely the closest curl-free approximation to  $\hat{\alpha}$  which fulfills the validity criterion  $|\Delta \alpha(\theta)| \equiv |\tilde{\alpha}(\theta) - \hat{\alpha}(\theta)| \leq \varepsilon_{\text{acc}}$  over a region  $\mathcal{U}$  in the lens plane where multiple images occur. We have studied in detail the relevance of this criterion using  $\varepsilon_{\text{acc}} = 5 \times 10^{-3} \theta_E$  as suggested in USS17. We have shown that this criterion is not appropriate, in particular for positions close to the critical lines. Indeed, we have demonstrated to first order that the offsets  $|\Delta \theta|$  between the original and SPT-modified lensed images depends on the SPT-modified Jacobi matrix whose impact become larger as we get closer to the critical curves,  $|\Delta \theta| \approx |\hat{\mathcal{A}}^{-1}(\theta) \Delta \alpha(\theta)|$ . Thus, the criterion  $|\Delta \alpha(\theta)| \leq \varepsilon_{\text{acc}}$  over  $\mathcal{U}$  does not guarantee the image offsets  $|\Delta \theta|$  to be smaller than  $\varepsilon_{\text{acc}}$  over the same region. In those cases, the deflection field  $\tilde{\alpha}$  produces image configurations which can be observationally distinguished from the original ones. As a result, we suggest that the criterion  $|\Delta \alpha(\theta)| \leq \varepsilon_{\text{acc}}$  proposed in USS17 should be reconsidered, which also means a revision of how the curl-free deflection law  $\tilde{\alpha}$  and  $\tilde{\psi}$  are defined.

For the indistinguishable image configurations produced by  $\tilde{\alpha}$ , we have studied how the time delays are affected by the SPT. For a quadrupole (NIS + shear) and a NIE models, we have shown that, once again, the time delay ratios scale like the source position ratios, i.e.  $\Delta \hat{t} / \Delta t \approx |\hat{\beta}| / |\beta|$ , when two images are produced. This result holds for opposite image pairs when four images are produced. For other image pair combinations, we confirm that the time delay ratios are not conserved. Thus, accurate time delay ratios measurements should help to reduce the degeneracy between SPT-generated models. However, the impact of the SPT remains low with deviations not larger than a few percent for the illustrative examples we have considered.

To extend the range of indistinguishable image configurations produced by  $\tilde{\alpha}$ , we have slightly modified the source mapping by means of a numerical optimization,  $\hat{\beta}(\beta) \rightarrow \beta^*(\beta)$ . We tested this method on a quadrupole model (NIS + shear) deformed by a radial stretching. The new source mapping differs from a radial stretching in the sense that  $\beta^*(\beta)$  is slightly anisotropic and shows discontinuities when a source crosses the tangential caustic curve. In the same way as for the previous cases, the time delay ratios of image pairs scale like the source position ratios, i.e.  $\Delta \hat{t}^* / \Delta t \approx |\beta^*| / |\beta|$ , when two images are produced. When four images are produced, the time delay ratios are

sensitive to the azimuth angle of  $\beta$ . We have shown that they scale like the source position ratios when the azimuth angle is parallel to the external shear direction and deviate to a few percent when the azimuth angle is perpendicular to the external shear direction. As a general conclusion, the impact of the SPT on time delay cosmography seems not be as crucial as initially suspected, leading to deviations that do not exceed a few percent.

In a future work, we aim to reconsider the validity criterion in more detail by redefining the curl-free deflection field  $\tilde{\alpha}$  and the corresponding potential  $\tilde{\psi}$ . Although new definitions for  $\tilde{\alpha}$  and  $\tilde{\psi}$  will affect the time delays, we expect that it will not modify substantially the results presented in this paper, likely leaving most of the conclusions unchanged.

*Acknowledgements.* We thank Dominique Sluse and Sandra Unruh for useful discussions. This work was supported by the Humboldt Research Fellowship for Postdoctoral Researchers.

## References

- Akhunov, T. A., Wertz, O., Elyiv, A., et al. 2017, *MNRAS*, **465**, 3607  
 Bar-Kana, R. 1996, *ApJ*, **468**, 17  
 Birrer, S., Amara, A., & Refregier, A. 2016, *JCAP*, **8**, 020  
 Blandford, R. D., & Narayan, R. 1992, *ARA&A*, **30**, 311  
 Bonvin, V., Courbin, F., Suyu, S. H., et al. 2017, *MNRAS*, **465**, 4914  
 Burke, W. L. 1981, *ApJ*, **244**, L1  
 Coe, D., & Moustakas, L. A. 2009, *ApJ*, **706**, 45  
 Courbin, F., Chantry, V., Revaz, Y., et al. 2011, *A&A*, **536**, A53  
 Dyer, C. C., & Roeder, R. C. 1980, *ApJ*, **238**, L67  
 Eulaers, E., Tewes, M., Magain, P., et al. 2013, *A&A*, **553**, A121  
 Falco, E. E., Gorenstein, M. V., & Shapiro, I. I. 1985, *ApJ*, **289**, L1  
 Fassnacht, C. D., Gal, R. R., Lubin, L. M., et al. 2006, *ApJ*, **642**, 30  
 Finet, F., & Surdej, J. 2016, *A&A*, **590**, A42  
 Finet, F., Elyiv, A., & Surdej, J. 2012, *Mem. Soc. Astron. It.*, **83**, 944  
 Finet, F., Elyiv, A., Melnyk, O., et al. 2015, *MNRAS*, **452**, 1480  
 Fohlmeister, J., Kochanek, C. S., Falco, E. E., et al. 2013, *ApJ*, **764**, 186  
 Gorenstein, M. V., Shapiro, I. I., & Falco, E. E. 1988, *ApJ*, **327**, 693  
 Jean, C., Claeskens, J.-F., & Surdej, J. 2001, *ASP Conf. Ser.*, **237**, 423  
 Keeton, C. R. 2001, ArXiv e-prints [arXiv:astro-ph/0102341]  
 Keeton, C. R. 2003, *ApJ*, **584**, 664  
 Kochanek, C. S. 2002, *ApJ*, **578**, 25  
 Kochanek, C. S. 2006, in *Gravitational Lensing: Strong, Weak and Micro, Saas-Fee Advanced Course 33*, eds. G. Meylan, et al. (Berlin: Springer), 91  
 Koopmans, L. V. E., Treu, T., Fassnacht, C. D., Blandford, R. D., & Surpi, G. 2003, *ApJ*, **599**, 70  
 Levenberg, K. 1944, *Quart. J. Appl. Math.*, **2**, 164  
 Liao, K., Treu, T., Marshall, P., et al. 2015, *ApJ*, **800**, 11  
 Liesenborgs, J., & De Rijcke S. 2012, *MNRAS*, **425**, 1772  
 McCully, C., Keeton, C. R., Wong, K. C., & Zabludoff, A. I. 2014, *MNRAS*, **443**, 3631  
 McCully, C., Keeton, C. R., Wong, K. C., & Zabludoff, A. I. 2016, *AAS Meeting Abstracts*, **227**, 338.02  
 Oguri, M., & Marshall, P. J. 2010, *MNRAS*, **405**, 2579  
 Paraficz, D., & Hjorth, J. 2010, *ApJ*, **712**, 1378  
 Rathna Kumar, S., Tewes, M., Stalin, C. S., et al. 2013, *A&A*, **557**, A44  
 Refsdal, S. 1964, *MNRAS*, **128**, 307  
 Saha, P. 2000, *AJ*, **120**, 1654  
 Schneider, P. 2006, in *Gravitational Lensing: Strong, Weak and Micro, Saas-Fee Advanced Course 33*, eds. G. Meylan, et al. (Berlin: Springer), 1  
 Schneider, P. 2014a, *A&A*, **568**, L2  
 Schneider, P. 2014b, *A&A*, submitted [arXiv:1409.0015]  
 Schneider, P., & Sluse, D. 2013, *A&A*, **559**, A37  
 Schneider, P., & Sluse, D. 2014, *A&A*, **564**, A103  
 Seljak, U. 1994, *ApJ*, **436**, 509  
 Suyu, S. H., Marshall, P. J., Auger, M. W., et al. 2010, *ApJ*, **711**, 201  
 Suyu, S. H., Auger, M. W., Hilbert, S., et al. 2013, *ApJ*, **766**, 70  
 Tewes, M., Courbin, F., Meylan, G., et al. 2013, *A&A*, **556**, A22  
 Treu, T., & Marshall, P. J. 2016, *A&ARv*, **24**, 11  
 Unruh, S., Schneider, P., & Sluse, D. 2017, *A&A*, **601**, A77  
 Vuissoz, C., Courbin, F., Sluse, D., et al. 2008, *A&A*, **488**, 481  
 Wertz, O., & Orthen, B. 2018, ArXiv e-prints [arXiv:1801.04151]  
 Wong, K. C., Keeton, C. R., Williams, K. A., Momcheva, I. G., & Zabludoff, A. I. 2011, *ApJ*, **726**, 84  
 Wong, K. C., Suyu, S. H., Auger, M. W., et al. 2017, *MNRAS*, **465**, 4895  
 Wucknitz, O. 2002, *MNRAS*, **332**, 951  
 Xu, D., Sluse, D., Schneider, P., et al. 2016, *MNRAS*, **456**, 739

## Appendix A: Proof that $\varepsilon_{AB} = 0$ for an SIS and a radial stretching of the form (19)

For the sake of clarity, we first recall the definition of  $\varepsilon_{AB}$  given in Eq. (28):

$$\varepsilon_{AB} = \int_{|\theta_B|}^{\theta_A} \beta(\theta) [f(\beta(\theta)) - f(\beta_s)] d\theta. \quad (\text{A.1})$$

In this section, we will prove that  $\varepsilon_{AB} = 0$  for an SIS and for any deformation function  $f(\beta)$  which satisfies a few reasonable conditions. By definition,  $f(\beta)$  must be even to preserve the symmetry and  $1 + f(\beta) + \beta df(\beta)/d\beta > 0$  guarantees the mapping to be one-to-one. Let us assume that  $f(\beta)$  is a real analytic function, its Maclaurin series expansion thus exists and is simply given by

$$f(\beta) = \sum_{n=0}^{+\infty} f_n \frac{\beta^n}{n!}, \quad (\text{A.2})$$

where  $f_n := d^n f/d\beta^n$  is evaluated in  $\beta = 0$  and  $f_{2k+1} = 0$  for all  $k \in \mathbb{Z}^*$  to preserve the symmetry. Because of the linearity of integration and recalling the lens equation  $\beta(\theta) = \theta - \theta_E \theta/|\theta|$  for an SIS, Eq. (A.1) transforms into

$$\varepsilon_{AB} = \sum_{n=0}^{+\infty} \frac{f_n}{n!} \left( \int_{|\theta_B|}^{\theta_A} (\theta - \theta_E)^{n+1} d\theta \right) - f(\beta_s) \int_{|\theta_B|}^{\theta_A} (\theta - \theta_E) d\theta, \quad (\text{A.3})$$

where  $n$  is now a positive even integer or 0. Keeping in mind that  $n + 2$  is even, the linearity of the integrand guarantees that

$$\int_{|\theta_B|}^{\theta_A} (\theta - \theta_E)^{n+1} d\theta = \left[ \frac{(\theta - \theta_E)^{n+2}}{n+2} \right]_{|\theta_B|}^{\theta_A} = 0, \quad (\text{A.4})$$

where in the last step we used  $\theta_A - \theta_E = \beta_s$  and  $|\theta_B| - \theta_E = -\beta_s$ . Because the latter expression is valid for all positive even integer  $n$  and for  $n = 0$ , we deduce that all the successive terms in the series vanish, as the second integral in Eq. (A.3), leading to  $\varepsilon_{AB} = 0$ . Furthermore, the accuracy of (A.2) is not affected by the value at which the function is evaluated as long as the series converges. Thus, we finally require that the radius of convergence  $r$  of the Maclaurin series (A.2) satisfies the condition  $r \geq \theta_E$ . If satisfied, this very plausible assumption ensures the series expansion (A.2) to be exact for all  $\beta < \theta_E$ .

## Appendix B: Proof that $\varepsilon_{AB} \approx 0$ for any axisymmetric model and radial stretching of the form (19)

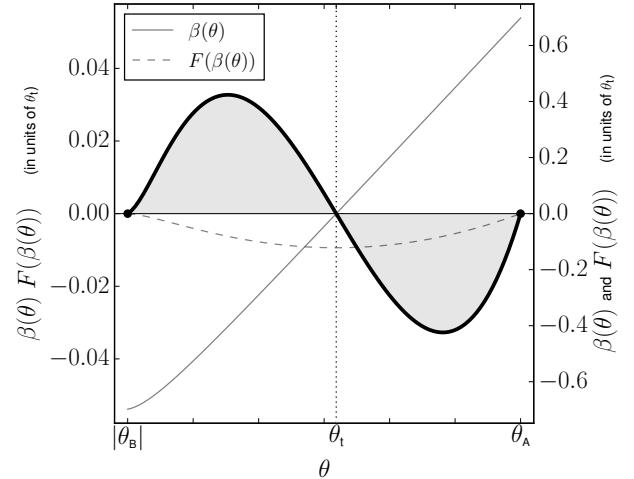
To convince the reader, we consider separately the cases when  $\beta_s$  tends to 0 ( $\beta_s \rightarrow 0$ ) and when  $\beta_s < \beta(\theta_r)$ , where  $\theta_r$  corresponds to the angular radius of the radial critical curve.

For  $\beta_s \rightarrow 0$ , we see from Eq. (28) that the integration interval degenerates into  $\{\theta_t\}$ , where  $\theta_t$  corresponds to the angular radius of the tangential critical curve. Thus, the integrand also tends to 0 and  $\varepsilon_{AB} \rightarrow 0$ . In addition,  $\varepsilon_{AB} \rightarrow 0$  faster than  $\Delta\tau_{AB} \rightarrow 0$  due to the first term in Eq. (26) while  $1 + f(\beta_s) \rightarrow 1$ . Thus, we deduce that the  $\varepsilon_{AB}$ -term in Eq. (27) can be neglected when  $\beta_s \rightarrow 0$ .

For  $\beta_s < \beta(\theta_r)$ , we define the quantity

$$\eta_{AB} = \frac{\varepsilon_{AB}/\Delta\tau_{AB}}{\hat{\beta}(\beta_s)/\beta_s}, \quad (\text{B.1})$$

and show that  $|\eta_{AB}| \ll 1$ , which is sufficient to guarantee that the  $\varepsilon_{AB}$ -term in Eq. (27) can be neglected. With this aim in mind, we analyze the graph of the integrand



**Fig. B.1.** Graph of the integrand  $\beta(\theta)F(\beta(\theta))$  (thick black curve) defined in Eq. (28) for the “unfavorable” case  $|\theta_B| \rightarrow \theta_r$ . This integrand is composed of the functions  $\beta(\theta)$  (solid gray curve) and  $F(\beta(\theta))$  (dashed gray curve). The (almost) symmetrical graph of  $\beta(\theta)$  with respect to  $(\theta_t, 0)$  and  $F(\beta(\theta))$  with respect to the axis  $\theta = \theta_t$  (dotted vertical line) implies the (almost) symmetry of the integrand with respect to  $(\theta_t, 0)$ . As a consequence, the two gray areas (almost) compensate each other implying that  $|\varepsilon_{AB}| \ll \theta_t$ , hence  $|\eta_{AB}| \ll 1$ , and confirms the validity of the Eq. (29).

$\beta(\theta)[f(\beta(\theta)) - f(\beta_s)] =: \beta(\theta)F(\beta(\theta))$  defined in Eq. (28) and for which an example is shown in Fig. B.1. Based on general considerations, we expect the graph of the integrand to be almost symmetric with respect to the point  $(\theta_t, 0)$  over the interval  $[[\theta_B], \theta_A]$ . First, the graph of  $\beta(\theta)$  monotonically increases over  $[[\theta_B], \theta_A]$  and always crosses the  $\theta$ -axis at the position  $\theta = \theta_t \approx (\theta_A + |\theta_B|)/2$ . These statements stem from the general properties of axisymmetric lenses in the case of a single lens plane (see e.g. Sect. 3.1 in Schneider 2006). In addition, because of  $d\beta(\theta)/d\theta$  is almost constant ( $\approx 1$ ) for most  $\theta \in [[\theta_B], \theta_A]$ , the graph of  $\beta(\theta)$  is almost symmetric with respect to the point  $(\theta_t, 0)$ . The largest asymmetry occurs for  $\beta_s \rightarrow \beta(\theta_r)$  for which  $|\theta_B| \rightarrow \theta_r$  (this “unfavorable” case is actually the one shown in Fig. B.1). Second, the graph of  $F(\beta(\theta))$  can only cross the  $\theta$ -axis at the positions  $\theta_A$  and  $\theta_B$  over  $[[\theta_B], \theta_A]$ , and is almost symmetric with respect to the axis  $\theta = \theta_t$ . We easily confirm that  $\theta_A$  and  $\theta_B$  are  $\theta$ -intercepts from  $F(\beta(\theta_A)) = F(\beta(\theta_B)) = F(\beta_s) = 0$ . To show there exists no other  $\theta$ -intercept within  $[[\theta_B], \theta_A]$ , we use a reduction ad absurdum argument. If the graph of  $F(\beta(\theta))$  crosses the  $\theta$ -axis at a third position  $\theta_t \in [[\theta_B], \theta_A]$ , then  $F(\beta(\theta_t)) = 0$  implies that  $\theta_t$  corresponds to the position of a lensed image of the source  $\beta_s$ . Since the third image  $\theta_C$  always satisfies the condition  $0 < |\theta_C| < |\theta_B| < \theta_t < \theta_A$ , hence  $\theta_C \notin [[\theta_B], \theta_A]$ , the existence of this fourth lensed image violates the so-called “odd number theorem” (Dyer & Roeder 1980; Burke 1981) in the case of a single lens plane. In addition, because of the symmetry of  $\beta(\theta)$  for  $\theta \in [[\theta_B], \theta_A]$  and recalling that  $f$  is an even function of  $\beta$ , we have  $F(\beta(\theta_t + \delta\theta)) \approx F(\beta(\theta_t - \delta\theta))$ , with  $\delta\theta \in [0, \min(\theta_A - \theta_t, \theta_t - |\theta_B|)]$ , which guarantees  $F(\beta(\theta))$  to be almost symmetric with respect to the axis  $\theta = \theta_t$ . Combining all these statements leads to the conclusion that the integrand  $\beta(\theta)F(\beta(\theta))$  is also almost symmetric with respect to  $(\theta_t, 0)$ . Thus, the integral (28) consists in differencing the (almost identical) gray areas displayed in Fig. B.1, which tends to compensate each other, leading to  $|\varepsilon_{AB}| \ll \theta_t$ . Finally, from Eq. (26) we deduce that  $|\Delta\tau_{AB}| \approx 2\beta_s \theta_t$  due to the symmetry of  $\beta(\theta)$  over  $[[\theta_B], \theta_A]$ . As a conclusion, we find  $|\eta_{AB}| \ll 1$  from Eq. (B.1), which confirms that the  $\varepsilon_{AB}$  term in Eq. (27) can be in general neglected.