Astronomy & Astrophysics

Generalized multi-plane gravitational lensing: time delays, recursive lens equation, and the mass-sheet transformation

Peter Schneider

Argelander-Institut für Astronomie, Universität Bonn, Auf dem Hügel 71, 53121 Bonn, Germany
e-mail: peter@astro.uni-bonn.de

ABSTRACT

We consider several aspects of the generalized multi-plane gravitational lens theory, in which light rays from a distant source are affected by several main deflectors, and in addition by the tidal gravitational field of the large-scale matter distribution in the Universe when propagating between the main deflectors. Specifically, we derive a simple expression for the time-delay function in this case, making use of the general formalism for treating light propagation in inhomogeneous spacetimes which leads to the characterization of distance matrices between main lens planes. Applying Fermat's principle, an alternative form of the corresponding lens equation is derived, which connects the impact vectors in three consecutive main lens planes, and we show that this form of the lens equation is equivalent to the more standard one. For this, some general relations for cosmological distance matrices are derived. The generalized multi-plane lens situation admits a generalized mass-sheet transformation, which corresponds to uniform isotropic scaling in each lens plane, a corresponding scaling of the deflection angle, and the addition of a tidal matrix (mass sheet plus external shear) to each main lens. The scaling factor in the lens planes exhibits a curious alternating behavior for odd and even numbered planes. We show that the time delay for sources in all lens planes scale with the same factor under this generalized mass-sheet transformation, thus precluding the use of time-delay ratios to break the mass-sheet transformation.

Key words. cosmological parameters – gravitational lensing: strong

1. Introduction

In strong gravitational lensing systems, in which a galaxy or a galaxy cluster causes multiple images or strong image distortions of background sources, one often neglects the inhomogeneities of the gravitational field between the observer and the lens, and between the lens and the source (see, e.g., Kochanek 2006; Treu 2010; Bartelmann 2010; Kneib & Natarajan 2011, for reviews on strong lensing systems). This usually provides a very good approximation, since the lensing strength of the main lens over the region where strong lensing effects occur is much larger than the typical distortion effects of matter along the line-of-sight. The latter is comparable to the typical strength of cosmic shear effects (see, e.g., Bartelmann & Schneider 2001; Schneider 2006; Munshi et al. 2008; Hoekstra 2013), and amounts to about 1 or 2% of the distortion in the strong-lens region.

Whereas the propagation effects are thus small, the interest of these weak distortions has been renewed, for at least two different reasons. The first is that strong lens systems may be biased toward showing relatively strong line-of-sight structures, thereby increasing their lensing probability. This effect is most likely affecting the lensing cross sections for the formation of giant arcs in clusters (see Bartelmann et al. 1998 for stating the "arc statistic problem", and Meneghetti et al. 2013 for a recent update of the issue). Cosmological simulations (Puchwein & Hilbert 2009, and references therein) indicate that line-of-sight structure can indeed substantially modify the lensing efficiency of clusters. More recently, Bayliss et al. (2014) found a significant overdensity of galaxy groups along the line-of-sight toward strong-lensing clusters, observationally supporting the presence of this bias.

The second reason for the renewed interest in intervening distortions of strong-lens systems is their use for a precise deter-mination of parameters, most noticible the Hubble constant from measured time delays (Refsdal 1964) in multiply imaged quasi-stellar objects (see, e.g., Kochanek 2003, 2006; Treu 2010; Suyu et al. 2010, 2013). The quality of modern imaging data and the accuracy of time delay estimates allows one to derive estimates of the Hubble constant with a formal error of ~6% (Suyu et al. 2013). At this level of precision, line-of-sight effects may become highly relevant in these strong-lensing systems (e.g., Wong et al. 2011; Collett et al. 2013; Greene et al. 2013, and references therein).

With additional deflectors along the line-of-sight, the lens equation, which relates the true source position to the observed positions of images, needs to be generalized to include deflection at more than one distance from the observer. Blandford & Narayan (1986) established the theory of multi-plane gravitational lensing (see also Chap. 9 of Schneider et al. 1992), where the mapping between images and their source is affected by the action of several deflectors at different redshifts. This multi-plane lensing is employed in ray-tracing simulations – see Refsdal (1970) for the earliest ray-tracing simulations, Jain et al. (2000), Hilbert et al. (2009), and references therein – where the three-dimensional mass distribution between observer and source is partitioned into separate slices, and the mass distribution in each slice is treated as a gravitational lens plane.

The full theory of multiple deflection lensing is required if there are two or more strong deflectors along the same line-of-sight toward a background source. For galaxy-scale lensing, such systems are rare, since such an alignment is not very probable. However, in the sample of several hundred galaxy-scale strong lens systems currently known, there are examples of multiple lenses at two different redshifts (Chae et al. 2001; Gavazzi et al. 2008). These systems may be of particular interest, since they

may be used in principle to determine cosmological distance ratios (Collett & Auger 2014), and therefore to constrain the cosmic expansion history, although in practice the mass-sheet transformation renders this a difficult task (Schneider 2014).

Far more common are situations where there is a single strong lens in the line-of-sight to a distant source, and several other deflectors at different redshifts situated sufficiently far away from the strong-lensing region of the main deflector such that their deflection angle can be linearized across this strong-lensing region. Kovner (1987) considered this case of one main lens, combined with linearized deflections at different redshifts. The effects of the linear deflectors can be summarized into a set of matrices which describe the mapping between angles at the vertices of light cones to the separation vectors along the light cone at the redshifts of lens and source. Schneider (1997) reconsidered this situation, using the general formalism of light propagation in an inhomogeneous universe (Seitz et al. 1994; hereafter SSE), and related this generalized gravitational lens equation to the one where the deflection occurs in only a single plane, with a tidal deflection matrix added (see also Keeton et al. 1997). McCully et al. (2014) further generalized this theory to consider several main lenses, together with linearized deflections between the main lens planes. In particular, McCully et al. (2014) emphasized the advantages of this hybrid framework for the modeling of strong lens systems with multiple main deflectors along the line-of-sight (see also McCully et al. 2017).

One of the prime motivations for the work of McCully et al. (2014) was the derivation of the time-delay function for such generalized lensing systems, as this is required for relating the time delay to the scale-length of the Universe, i.e., the Hubble radius. Using the Millennium Simulation, Jaroszynski & Kostrzewa-Rutkowska (2014) investigated the impact of the line-of-sight matter distribution on strong lensing properties, including the time delay; they concluded that the intermediate mass distribution leads to a spread of ~6% in the product of Hubble constant and time delay, for strong lensing systems with source redshifts $z_s \sim 2$. Arguably, the largest obstacle for the time-delay method to obtain accurate estimates for the Hubble constant is the degeneracy of the mass model due to the mass-sheet transformation (MST; Falco et al. 1985; see Schneider & Sluse 2013 for a recent discussion) or the more general source position transformation (Schneider & Sluse 2014). Recently, Schneider (2014) has shown that an MST also exists for the case of lenses at two different distances from us. In particular, this MST leads to a scaling of all time delays (from sources on both source planes) by the same factor, thus precluding that the degeneracy due to the MST can be broken by time-delay ratios.

In this paper, these results are further generalized to the case of arbitrarily many main lens planes, together with linear deflections between the main lens (and source) planes. After a short summary of the propagation equations in an inhomogeneous universe, as they apply to the case under consideration here, we derive the time-delay function for the case of several main lens planes. Whereas this has been done before by McCully et al. (2014), our expression for the light travel time function is expressed in a form which allows us to apply Fermat's theorem in gravitational lensing. Thus, the lens equation can be obtained from requiring that the light travel time function is stationary with respect to all impact vectors in the main lens planes. With this procedure, we derive an alternative form of the lens equation, which involves the impact vectors of three consecutive lens planes and the deflection angle in the middle one of them. We show that this iterative form of the lens equation is

equivalent to the more standard form; in order to do so, we first obtain a very general relation between distance matrices.

We then turn to the MST in this case, and find a curious property of its behavior: The uniform, isotropic scaling factor which characterizes the MST alternates between a free parameter λ and unity from one lens/source plane to the next. The corresponding modification of the deflection angle in the main lens planes corresponds to a scaling of the deflection, plus the addition of a tidal deflection matrix (which in the absence of linear deflections between lens planes reduces to the addition of a uniform mass sheet). Finally we show that all time delays, for sources located on any plane, scale by the same factor under an MST, precluding the possibility of breaking the degeneracy from the MST by measured time-delay ratios.

2. Generalized multi-plane lens equation

2.1. Optical tidal equation

The propagation of light rays follows from the geodesic equation, specialized to a perturbed Robertson–Walker metric (see Schneider et al. 1992; Seitz et al. 1994; Bartelmann 2010, and references therein). In particular, for infinitesimally small light bundles, the separation vector $\boldsymbol{\xi}$ of a ray from the reference (or "central") ray of the bundle is given by the optical tidal equation

$$\boldsymbol{\xi}''(\lambda) = \mathcal{T}(\lambda)\,\boldsymbol{\xi}(\lambda), \tag{1}$$

where $\mathcal{T}$ is the optical tidal matrix, evaluated at the affine parameter λ along the central ray. Here and throughout this paper, a prime denotes differentiation with respect to the affine parameter λ, which is related to the redshift z, the cosmic scale factor $a = 1/(1 + z)$, or cosmic time t through

$$d\lambda = -\frac{c\,dt}{(1 + z)} = -\frac{c\,da}{H} = \frac{c\,dz}{H\,(1 + z)^2}, \tag{2}$$

where we have assumed that λ increases with redshift, i.e., decreases with cosmic time,

$$H = \dot{a}/a = H_0 \sqrt{\Omega_{\rm m}(1 + z)^3 + (1 - \Omega_{\rm m} - \Omega_\Lambda)(1 + z)^2 + \Omega_\Lambda} \tag{3}$$

is the Hubble function, H_0 is the Hubble constant, and $\Omega_{\rm m}$ and Ω_Λ are the cosmic density parameters in matter and vacuum energy.

We consider $\mathcal{T}$ to consist of three separate components, $\mathcal{T} = \mathcal{T}_{\rm bg} + \mathcal{T}_{\rm sm} + \mathcal{T}_{\rm cl}$. The first is the optical tidal matrix of the homogeneous background universe, and is given by (see SSE)

$$\mathcal{T}_{\rm bg} = -\frac{3}{2} \left(\frac{H_0}{c}\right)^2 \Omega_{\rm m}(1 + z)^5 \mathcal{I}, \tag{4}$$

where $\mathcal{I}$ is the two-dimensional unit matrix. If we consider a light bundle with vertex at redshift $z = 0$ and affine parameter $\lambda = 0$, which is only subject to $\mathcal{T}_{\rm bg}$, then the solution of Eq. (1) is $\boldsymbol{\xi}(\lambda) = D(\lambda)\boldsymbol{\theta}$, where $\boldsymbol{\theta}$ is the angle that the light ray under consideration encloses with the fiducial ray at the vertex, and $D(\lambda)$ is the solution of the differential equation

$$\frac{d^2 D(\lambda)}{d\lambda^2} = -\frac{3}{2} \left(\frac{H_0}{c}\right)^2 \Omega_{\rm m}(1 + z)^5 D(\lambda), \tag{5}$$

with initial condition $D(0) = 0$ and $D' = 1$. Here, D is the angular-diameter distance of the homogeneous universe, as a function of affine parameter (or redshift).

We consider two different kinds of inhomogeneities here. The first of them is related to small-scale density inhomogeneities, such as galaxies and their dark matter halos. For those, the optical tidal matrix $\mathcal{T}_{cl}$ is a strong function of position, and there is no longer a linear relation between enclosed angle $\boldsymbol{\theta}$ and separation $\boldsymbol{\xi}$ of a light ray, for finite $\boldsymbol{\theta}$. In strong-lensing applications, we are typically interested only in that region of these small-scale inhomogeneities where multiple images can be formed, which corresponds to a few tens of kiloparsecs for galaxy lenses. These small-scale inhomogeneities will be considered explicitly as "main lenses" in the following. The second kind of inhomogeneities is due to the large-scale mass distribution of the universe. We assume that the corresponding gravitational field is sufficiently smooth, so that the tidal effects can be considered approximately constant across the region where strong-lensing effects occur. Hence we assume that over such a region, $\mathcal{T}_{sm}$ can be considered to depend only on λ, not on the actual position of the light ray. According to SSE, this contribution of the optical tidal matrix is

$$(\mathcal{T}_{sm})_{ij} = -\frac{(1+z)^2}{c^2} \left(2 \frac{\partial^2 \phi}{\partial \xi_i \partial \xi_j} + \delta_{ij} \frac{\partial^2 \phi}{\partial \xi_3^2} \right), \tag{6}$$

where ϕ is the Newtonian potential sourced by the density inhomogeneity, i.e., satisfying the (three-dimensional) Poisson equation $\nabla_\xi^2 \phi = 4\pi G(\rho - \bar{\rho})$, where $\bar{\rho}(z)$ is the mean matter density in the Universe, and we assumed that the light ray propagates in the ξ_3-direction.

2.2. Generalized multi-plane lens equation

We consider a direction in the sky with has a set of main lenses along the line-of-sight to a distant source, located at redshifts z_i or affine parameters λ_i. As shown in SSE (see also Schneider 1997), the separation vector $\boldsymbol{\xi}(\lambda)$ then becomes

$$\boldsymbol{\xi}(\lambda) = \mathcal{D}(\lambda)\,\boldsymbol{\theta} - \sum_i \mathcal{D}_i(\lambda) \left[\hat{\boldsymbol{\alpha}}_i(\boldsymbol{\xi}_i) - \hat{\boldsymbol{\alpha}}_i^0 \right] H(\lambda - \lambda_i), \tag{7}$$

where $\boldsymbol{\theta}$ is the angle the light encloses with the fiducial ray at the observer. The distance matrix $\mathcal{D}(\lambda)$ solves the optical tidal equation

$$\frac{d^2 \mathcal{D}(\lambda)}{d\lambda^2} = \left[\mathcal{T}_{bg}(\lambda) + \mathcal{T}_{sm}(\lambda) \right] \mathcal{D}(\lambda), \tag{8}$$

with initial conditions $\mathcal{D}(0) = 0$ and $\mathcal{D}'(0) = \mathcal{I}$, and the distance matrices $\mathcal{D}_i(\lambda)$ solve the same differential equation, but with initial conditions $\mathcal{D}_i(\lambda_i) = 0$ and $(d\mathcal{D}_i/d\lambda)(\lambda_i) = (1 + z_i)\mathcal{I}$. They are the distance matrices which apply for light rays having their vertex at λ_i. In Eq. (7), $H(\lambda - \lambda_i)$ is the Heaviside step function. The deflection angle $\hat{\boldsymbol{\alpha}}_i(\boldsymbol{\xi}_i)$ is given in terms of the surface mass density $\Sigma_i(\boldsymbol{\xi}_i)$ as

$$\hat{\boldsymbol{\alpha}}_i(\boldsymbol{\xi}_i) = \frac{4G}{c^2} \int d^2\xi' \, \Sigma_i(\boldsymbol{\xi}') \frac{\boldsymbol{\xi}_i - \boldsymbol{\xi}'}{|\boldsymbol{\xi}_i - \boldsymbol{\xi}'|^2}, \tag{9}$$

and $\hat{\boldsymbol{\alpha}}_i^0$ denotes the deflection angle of the fiducial ray in the ith lens plane. If we now perform the translation $\boldsymbol{\xi}(\lambda) = \tilde{\boldsymbol{\xi}}(\lambda) + \boldsymbol{\eta}(\lambda)$, with

$$\boldsymbol{\eta}(\lambda) = \sum_i \mathcal{D}_i(\lambda) \, \hat{\boldsymbol{\alpha}}_i^0 \, H(\lambda - \lambda_i),$$

then $\tilde{\boldsymbol{\xi}}$ satisfies Eq. (7) with the $\hat{\boldsymbol{\alpha}}_i^0$ set to zero. In the following, we always assume this (unobservable) translation, and drop the

tilde on $\tilde{\boldsymbol{\xi}}$ henceforth. For the impact vectors $\boldsymbol{\xi}_j$ in the jth plane, we then obtain

$$\boldsymbol{\xi}_j = \mathcal{D}(\lambda_j)\,\boldsymbol{\theta} - \sum_{i=1}^{j-1} \mathcal{D}_i(\lambda_j)\, \hat{\boldsymbol{\alpha}}_i(\boldsymbol{\xi}_i) \equiv \mathcal{D}_j\,\boldsymbol{\theta} - \sum_{i=1}^{j-1} \mathcal{D}_{ij}\, \hat{\boldsymbol{\alpha}}_i(\boldsymbol{\xi}_i). \tag{10}$$

2.3. Calculation of the distance matrices

The distance matrices $\mathcal{D}_i(\lambda)$ depend on the large-scale matter distribution around the line-of-sight, which is dominated by dark matter and thus difficult to determine observationally. One possibility to estimate $\mathcal{T}_{sm}$ from observations is to assume that galaxies provide a good tracer of the total matter distribution on large scales. From an observed distribution of galaxies (or a particular kind of galaxies, like luminous red galaxies) around the line-of-sight, an estimate of the tidal field can be obtained; this is the strategy proposed in Collett et al. (2013) and Smith et al. (2014).

The propagation of light through the large-scale structure is the subject of cosmological weak lensing, or cosmic shear (see, e.g., Schneider 2006). In cosmic shear, one usually describes the propagation matrices in comoving coordinates. In order to connect the formulation given here, where the $\mathcal{D}_i$ relate angles to proper transverse separation vectors (which is appropriate, as the matter distribution of the main lenses – galaxies or clusters – are most conveniently described in physical scales), to that used in weak lensing, we show in the appendix that

$$\mathcal{D}_i(\chi) = D_i(\chi)\mathcal{I} - \frac{2}{c^2} \int_{\chi_i}^{\chi} d\chi' \frac{a(\chi)}{a(\chi')} f_k(\chi - \chi') H(\phi(\chi')) \mathcal{D}_i(\chi'), \tag{11}$$

where χ is the comoving distance, $f_k(\chi)$ is the comoving angular diameter distance, $D_i(\chi) = a(\chi) f_k(\chi - \chi_i)$, which satisfies the differential equation

$$\frac{d^2 f_k(\chi)}{d\chi^2} = -K f_k(\chi) \tag{12}$$

with initial conditions $f_k(0) = 0$ and $df_k(0)/d\chi = 1$. Furthermore, $K = (\Omega_m + \Omega_\Lambda - 1)H_0^2/c^2$ is the spatial curvature of the universe, and $H(\phi)$ is the two-dimensional Hessian of the gravitational potential ϕ, evaluated in comoving transverse coordinates. Provided the perturbations are small, so that $\mathcal{D}_i$ deviates only slightly from $D_i\mathcal{I}$, we can replace $\mathcal{D}_i(\chi')$ by $D_i(\chi')$ in the integrand,

$$\mathcal{D}_i(\chi) = D_i(\chi)\mathcal{I} - \frac{2}{c^2} \int_{\chi_i}^{\chi} d\chi' \frac{a(\chi)}{a(\chi')} f_k(\chi - \chi') H(\phi(\chi')) D_i(\chi'). \tag{13}$$

In weak lensing, this approximation is called "the neglect of lens-lens coupling", often also termed as "Born-approximation". This approximation is very accurate and certainly sufficient for the purposes discussed in the current context[1]. Thus, the deviation of $\mathcal{D}_i$ from $D_i\mathcal{I}$ is given as a line-of-sight integral over the tidal force field, with a distance-dependent weighting. We also note that in the approximation (13), the distance matrices $\mathcal{D}_i$ are symmetric, which is generally not the case if the exact expression (11) is used.

For statistical studies, instead of trying to obtain the tidal field along the line-of-sight toward the sources from observations of the galaxy distribution in those directions, one can also derive the probability distribution for the distance matrices from cosmological simulations, as has been done in Suyu et al. (2013).

---

[1] It must be stressed here that this Born-approximation only applies to the distance matrices $\mathcal{D}_i$ which are governed by the smooth part of $\mathcal{T}$ according to Eq. (8); only this smooth contribution is contained in Eq. (11). No such approximation is made with regards to the main deflectors.

3. Time-delay function and the iterative lens equation

In this section, we first derive the light travel time function (LTTF) corresponding to the generalized multi-plane lens equation considered in the previous section. An alternative form of the lens equation is then derived from Fermat's principle, which in the current context states that the lens equation is equivalent to setting the gradient of the LTTF with respect to all impact vectors equal to zero (i.e., real light rays correspond to stationary points of the LTTF). Then we show that this iterative form of the lens equation is equivalent to Eq. (10).

3.1. Time-delay function

We first consider a single lens plane at z_1 with a source at z_2, in which case the lens equation reads

$$\boldsymbol{\xi}_2 = \mathcal{D}_2\,\boldsymbol{\theta} - \mathcal{D}_{12}\,\hat{\boldsymbol{\alpha}}(\boldsymbol{\xi}_1) = \mathcal{D}_2\,\mathcal{D}_1^{-1}\boldsymbol{\xi}_1 - \mathcal{D}_{12}\,\hat{\boldsymbol{\alpha}}(\boldsymbol{\xi}_1), \tag{14}$$

where we set $\hat{\boldsymbol{\alpha}}(\boldsymbol{\xi}_1) \equiv \hat{\boldsymbol{\alpha}}_1(\boldsymbol{\xi}_1)$. We define the LTTF $\tau(\boldsymbol{\xi}_1, \boldsymbol{\xi}_2)$ to be the excess light travel time from a source at $\boldsymbol{\xi}_2$ to the observer caused by the deflection in the lens plane at $\boldsymbol{\xi}_1$. This excess travel time has two components, a geometrical one (a bent ray is longer than an unbent one), and a potential one caused by the retardation of photons in the gravitational potential of the deflector.

In standard lens theory, with unperturbed angular-diameter distances D_{ij}, the potential part of the time delay function takes the form $c\tau_{\mathrm{pot}} = -(1 + z_1)(D_1 D_2/D_{12})\psi(\boldsymbol{\theta})$, where $\psi(\boldsymbol{\theta})$ is the deflection potential, which satisfies $\nabla_{\boldsymbol{\theta}}\psi = \boldsymbol{\alpha}(\boldsymbol{\theta}) \equiv (D_{12}/D_2)\hat{\boldsymbol{\alpha}}(D_1\boldsymbol{\theta})$. Now we define the potential $\hat{\psi}(\boldsymbol{\xi}_1)$ such that it satisfies $\nabla_{\boldsymbol{\xi}_1}\hat{\psi} = \hat{\boldsymbol{\alpha}}$. This potential is a multiple of ψ, i.e., $\hat{\psi}(\boldsymbol{\xi}_1) = k\,\psi(\boldsymbol{\xi}_1/D_1)$, up to an irrelevant additive constant. To find k, we consider

$$\hat{\boldsymbol{\alpha}} = \nabla_{\boldsymbol{\xi}_1}\hat{\psi} = \frac{k}{D_1}\nabla_{\boldsymbol{\theta}}\psi = \frac{k}{D_1}\boldsymbol{\alpha} = \frac{k\,D_{12}}{D_1 D_2}\hat{\boldsymbol{\alpha}},$$

yielding

$$\hat{\psi}(\boldsymbol{\xi}_1) = \frac{D_1 D_2}{D_{12}}\psi(\boldsymbol{\xi}_1/D_1). \tag{15}$$

Hence, the potential part of the LTTF takes the form $c\tau_{\mathrm{pot}} = -(1 + z_1)\hat{\psi}(\boldsymbol{\xi}_1)$. As expected, it does not depend on the cosmological distances, since it is caused solely by the local effect of propagating through a gravitational field, and the corresponding time interval is then redshifted by a factor $(1 + z_1)$ to the observer. Accordingly, also in the case of perturbed light propagation between the lens planes, the potential part of the time delay must have the same form, since it is unaffected by propagation effects.

We find an explicit expression for $\tau(\boldsymbol{\xi}_1, \boldsymbol{\xi}_2)$ by requiring that $\nabla_{\boldsymbol{\xi}_1}\tau(\boldsymbol{\xi}_1, \boldsymbol{\xi}_2) = 0$ is equivalent to the lens Eq. (14). This fixes τ up to a multiplicative constant and terms which depend solely on $\boldsymbol{\xi}_2$. The multiplicative constant is fixed by the explicit expression given above for the potential part of τ, and the additive constant (which is irrelevant for time delay measurements) is fixed by requiring that the geometrical part of τ should vanish if the light ray is undeflected by the main lens. This then yields

$$c\tau(\boldsymbol{\xi}_1, \boldsymbol{\xi}_2) = \frac{1 + z_1}{2}\left(\mathsf{C}_{12}\boldsymbol{\xi}_1 - \mathcal{D}_{12}^{-1}\boldsymbol{\xi}_2\right)^{\mathrm{t}}\mathsf{C}_{12}^{-1}\left(\mathsf{C}_{12}\boldsymbol{\xi}_1 - \mathcal{D}_{12}^{-1}\boldsymbol{\xi}_2\right)$$
$$- (1 + z_1)\hat{\psi}(\boldsymbol{\xi}_1), \tag{16}$$

where we defined

$$\mathsf{C}_{ij} = \mathcal{D}_{ij}^{-1}\mathcal{D}_j\mathcal{D}_i^{-1}. \tag{17}$$

As shown in Schneider (1997) – see also Kovner (1987) – the matrix C_{ij} is symmetric. Indeed, we see that

$$\nabla_{\boldsymbol{\xi}_1}\tau(\boldsymbol{\xi}_1, \boldsymbol{\xi}_2) = \frac{1 + z_1}{c}\left(\mathsf{C}_{12}\boldsymbol{\xi}_1 - \mathcal{D}_{12}^{-1}\boldsymbol{\xi}_2 - \hat{\boldsymbol{\alpha}}\right) = 0 \tag{18}$$

is equivalent to the lens Eq. (14), as is easily verified by multiplying the foregoing expression by $\mathcal{D}_{12}$ from the left. We also note that $\tau = 0$ if $\hat{\psi} = 0$ and if the ray is unbent, i.e., if $\boldsymbol{\xi}_2 = \mathcal{D}_2\mathcal{D}_1^{-1}\boldsymbol{\xi}_1$, as we required for τ. McCully et al. (2014) obtained a somewhat different form for τ, which they showed to be equivalent to the expression given here. However, it must be pointed out that this equivalence applies only to physical light rays, i.e., those which satisfy the gravitational lensing Eq. (14). For those rays, we could write the light travel time as

$$c\tau = \frac{1 + z_1}{2}\hat{\boldsymbol{\alpha}}_1^{\mathrm{t}}\mathsf{C}_{12}^{-1}\hat{\boldsymbol{\alpha}}_1 - (1 + z_1)\hat{\psi}(\boldsymbol{\xi}),$$

where the lens Eq. (14) was used to eliminate $\boldsymbol{\xi}_2$. However, Eq. (16) is more general, as it yields the light travel time for all kinematically possible rays, not only for those for which the bend by the main lenses equals the actual deflection angle as calculated as the gradient of $\hat{\psi}$. This more general form of the τ is needed if the lens equation is to be derived from Fermat's principle.

In case of several main lens planes, the LTTF is obtained by considering the replacement of the actual light ray by successively straighter rays, i.e., by removing the bends of the ray and the gravitational potentials they traverse (see Sect. 9.2 of Schneider et al. 1992). Removing the bend and deflection potential in the first plane leads to the contribution (16) of the LTTF. Subsequent removal of the bend and potential in the second lens plane yields a similar contribution, with the indices (1,2) replaced by (2,3). Iterating this consideration, we obtain for the general case

$$c\tau(\boldsymbol{\xi}_1, \boldsymbol{\xi}_2, \ldots, \boldsymbol{\xi}_{N+1}) = \sum_{i=1}^{N}(1 + z_i)\left[\frac{1}{2}\left(\mathsf{C}_{i,i+1}\boldsymbol{\xi}_i - \mathcal{D}_{i,i+1}^{-1}\boldsymbol{\xi}_{i+1}\right)^{\mathrm{t}}\right.$$
$$\left.\times \mathsf{C}_{i,i+1}^{-1}\left(\mathsf{C}_{i,i+1}\boldsymbol{\xi}_i - \mathcal{D}_{i,i+1}^{-1}\boldsymbol{\xi}_{i+1}\right) - \hat{\psi}_i(\boldsymbol{\xi}_i)\right] \tag{19}$$

as the sum over terms of the form (16) for the individual planes.

Any ray connecting the source at $\boldsymbol{\xi}_{N+1}$ and the observer is fully characterized by the impact vectors $\boldsymbol{\xi}_i$, $1 \le i \le N$ in the lens planes, since between the planes, it follows the propagation Eq. (1) whose solution is uniquely determined by the two impact vectors at consecutive planes. Therefore, the actual light rays are singled out as those for which the LTTF is stationary, with respect to variations of the impact vectors in the N lens planes. Thus, the lens equation is obtained by setting the derivative of τ with respect to the $\boldsymbol{\xi}_j$ equal to zero. For each $j \ge 2$, two terms of the above sum contribute, namely the terms $i = j$ and $i = j - 1$. We obtain

$$\nabla_{\boldsymbol{\xi}_j}(c\tau) = (1 + z_j)\left[\mathsf{C}_{j,j+1}\boldsymbol{\xi}_j - \mathcal{D}_{j,j+1}^{-1}\boldsymbol{\xi}_{j+1} - \hat{\boldsymbol{\alpha}}_j(\boldsymbol{\xi}_j)\right]$$
$$+ (1 + z_{j-1})\left(\mathcal{D}_{j-1,j}^{-1}\right)^{\mathrm{t}}\left(\mathsf{C}_{j-1,j}\mathcal{D}_{j-1,j}^{-1}\boldsymbol{\xi}_j - \boldsymbol{\xi}_{j-1}\right)$$
$$= 0, \tag{20}$$

or

$$\boldsymbol{\xi}_{j+1} = \mathcal{D}_{j,j+1}\left[\mathsf{C}_{j,j+1} + \frac{1 + z_{j-1}}{1 + z_j}\left(\mathcal{D}_{j-1,j}^{-1}\right)^{\mathrm{t}}\mathsf{C}_{j-1,j}\mathcal{D}_{j-1,j}^{-1}\right]\boldsymbol{\xi}_j$$
$$- \frac{1 + z_{j-1}}{1 + z_j}\mathcal{D}_{j,j+1}\left(\mathcal{D}_{j-1,j}^{-1}\right)^{\mathrm{t}}\boldsymbol{\xi}_{j-1} - \mathcal{D}_{j,j+1}\hat{\boldsymbol{\alpha}}_j(\boldsymbol{\xi}_j). \tag{21}$$

This equation relates the position vectors in three consecutive planes to the deflection angle in the middle plane, quite in contrast to the lens Eq. (10) which contains all impact vectors $\boldsymbol{\xi}_i$ for a given $\boldsymbol{\xi}_j$, $1 \leq i \leq j - 1$. Hence, this new lens equation is more "local" than the original one.

In the following we explicitly show that these two forms of the lens equation are equivalent. For this, we first need to derive a general relation between distance matrices.

3.2. A relation between distance matrices

Consider the pairs of light rays sketched in Fig. 1, where the first has a vertex at λ_q and encloses an angle $\boldsymbol{\theta}$ with the fiducial ray. At the affine parameter λ_s, its separation vector from the fiducial ray is $\boldsymbol{\xi}_s$. The second light ray has its vertex at λ_r and intersects the first ray at λ_s; this then specifies its direction $\boldsymbol{\vartheta}$ relative to the fiducial ray. At the intersection point, the two rays enclose an angle $\boldsymbol{\varphi}$. From the geometry of the figure, we find

$$\boldsymbol{\theta} = \mathcal{D}_{qr}^{-1}\boldsymbol{\xi}_r = \mathcal{D}_{qs}^{-1}\boldsymbol{\xi}_s = \mathcal{D}_{qt}^{-1}\boldsymbol{\xi}_t;$$

$$\boldsymbol{\vartheta} = \mathcal{D}_{rt}^{-1}\left(\boldsymbol{\xi}_t + \Delta\boldsymbol{\xi}_t\right) = \mathcal{D}_{rs}^{-1}\boldsymbol{\xi}_s. \tag{22}$$

We use the latter equation to derive a relation between the $\mathcal{D}$'s, by expressing all vectors in terms of $\boldsymbol{\xi}_s$. Using the first of (22), we get $\boldsymbol{\xi}_t = \mathcal{D}_{qt}\mathcal{D}_{qs}^{-1}\boldsymbol{\xi}_s$. Furthermore, the figure shows that $\Delta\boldsymbol{\xi}_t = \mathcal{D}_{st}\boldsymbol{\varphi}$. On the other hand, $\boldsymbol{\xi}_r = -\mathcal{D}_s(\lambda_r)\boldsymbol{\varphi}$.

We now have to relate $\mathcal{D}_s(\lambda_r)$, the backward extension of the solution $\mathcal{D}_s(\lambda)$ of Eq. (8), to $\mathcal{D}_{rs}$. For that, we consider two solutions of Eq. (8), $\mathcal{D}_r(\lambda)$ and $\mathcal{D}_s(\lambda)$, with their appropriate initial conditions at λ_r and λ_s, respectively, and define the matrix

$$\mathsf{W}(\lambda) = \frac{\mathrm{d}\mathcal{D}_r^{\mathrm{t}}}{\mathrm{d}\lambda}(\lambda)\,\mathcal{D}_s(\lambda) - \mathcal{D}_r^{\mathrm{t}}(\lambda)\,\frac{\mathrm{d}\mathcal{D}_s}{\mathrm{d}\lambda}(\lambda), \tag{23}$$

which is the Wronskian of the differential Eq. (8); here, the superscript "t" denotes the transpose of a matrix. The derivative of W vanishes, due to Eq. (8); hence, W is a constant. Evaluating Eq. (23) at $\lambda = \lambda_r$ and making use of the initial conditions of $\mathcal{D}_r(\lambda)$ yields $\mathsf{W} = (1 + z_r)\,\mathcal{D}_s(\lambda_r)$. Similarly, at $\lambda = \lambda_s$ we find $\mathsf{W} = -(1 + z_s)\,(\mathcal{D}_r^{\mathrm{t}})(\lambda_s) = -(1 + z_s)\,\mathcal{D}_{rs}^{\mathrm{t}}$. Thus, we obtain

$$\mathcal{D}_s(\lambda_r) = -\frac{1 + z_s}{1 + z_r}\,\mathcal{D}_{rs}^{\mathrm{t}}, \tag{24}$$

which is Etherington's theorem in matrix form (Etherington 1933). With this relation, we then find that

$$\Delta\boldsymbol{\xi}_t = \mathcal{D}_{st}\boldsymbol{\varphi} = -\mathcal{D}_{st}\mathcal{D}_s^{-1}(\lambda_r)\boldsymbol{\xi}_r = \frac{1 + z_r}{1 + z_s}\,\mathcal{D}_{st}\left(\mathcal{D}_{rs}^{\mathrm{t}}\right)^{-1}\boldsymbol{\xi}_r.$$

Using Eq. (22) and collecting terms,

$$\mathcal{D}_{rt}\mathcal{D}_{rs}^{-1}\boldsymbol{\xi}_s = \boldsymbol{\xi}_t + \Delta\boldsymbol{\xi}_t = \mathcal{D}_{qt}\mathcal{D}_{qs}^{-1}\boldsymbol{\xi}_s + \frac{1 + z_r}{1 + z_s}\,\mathcal{D}_{st}\left(\mathcal{D}_{rs}^{\mathrm{t}}\right)^{-1}\mathcal{D}_{qr}\mathcal{D}_{qs}^{-1}\boldsymbol{\xi}_s$$

follows. Since this relation is valid for all $\boldsymbol{\xi}_s$, a general relation between distance matrices is obtained:

$$\frac{1 + z_r}{1 + z_s}\,\mathcal{D}_{st}\left(\mathcal{D}_{rs}^{\mathrm{t}}\right)^{-1}\mathcal{D}_{qr} = \mathcal{D}_{rt}\mathcal{D}_{rs}^{-1}\mathcal{D}_{qs} - \mathcal{D}_{qt}, \tag{25}$$

where we multiplied the resulting equation by $\mathcal{D}_{qs}$ from the right[2]. This result generalizes the corresponding relation in

---

[2] We explicitly point out that the only geometrical relation used in this derivation is the one between angles and transverse separations, i.e., the definition of the distance matrices.
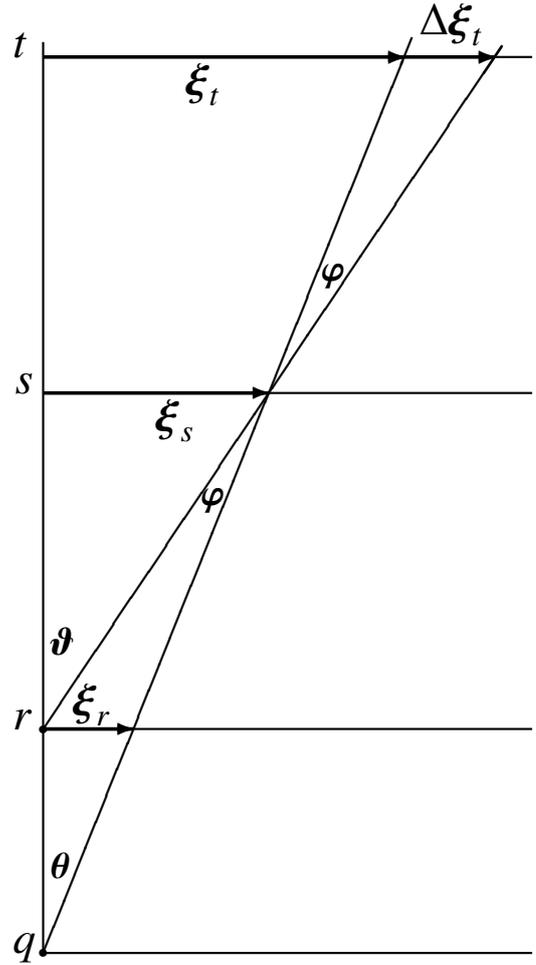
Fig. 1. Sketch of two light rays through four consecutive planes, with $0 \leq \lambda_q < \lambda_r < \lambda_s < \lambda_t$. The rays are not deflected in the lens planes. The first ray has its vertex at λ_q and encloses an angle $\boldsymbol{\theta}$ with the fiducial ray; the second ray with vertex at λ_r encloses an angle $\boldsymbol{\vartheta}$ with the fiducial ray. Both rays intersect at λ_s. The geometry of this figure yields the relation (25) between distance matrices.

Schneider (2016) derived for the case that the distance matrices $\mathcal{D}$ reduce to scalars.

Indeed, a relation of this kind is expected to hold: Consider $\lambda_t \equiv \lambda$ as a variable. The two matrix-valued functions $\mathcal{D}_q(\lambda)$ and $\mathcal{D}_r(\lambda)$ are linearly independent solutions of the transport Eq. (8), provided $\lambda_r \neq \lambda_q$. Therefore, the solution $\mathcal{D}_s(\lambda)$ can be written as a linear combination of the other two. This combination should be of the form

$$\mathcal{D}_s(\lambda) = \left[\mathcal{D}_r(\lambda)\mathcal{D}_r^{-1}(\lambda_s) - \mathcal{D}_q(\lambda)\mathcal{D}_q^{-1}(\lambda_s)\right]\mathsf{X}, \tag{26}$$

which satisfies one of the initial conditions, $\mathcal{D}_s(\lambda_s) = 0$. The matrix X is determined from the second initial condition; our result (25) shows that

$$\mathcal{D}_s(\lambda) = \left[\mathcal{D}_r(\lambda)\mathcal{D}_r^{-1}(\lambda_s) - \mathcal{D}_q(\lambda)\mathcal{D}_q^{-1}(\lambda_s)\right]\frac{1 + z_s}{1 + z_r}\,\mathcal{D}_{qs}\mathcal{D}_{qr}^{-1}\mathcal{D}_{rs}^{\mathrm{t}}. \tag{27}$$

3.3. Equivalence of Eqs. (10) and (21)

We shall now show that the two forms (10) and (21) of the lens equation are equivalent. As a first step, we rewrite Eq. (21) in a form that admits a simple geometrical interpretation. Specializing Eq. (25) to $q = 0$, $r = j - 1$, $s = j$, $t = j + 1$
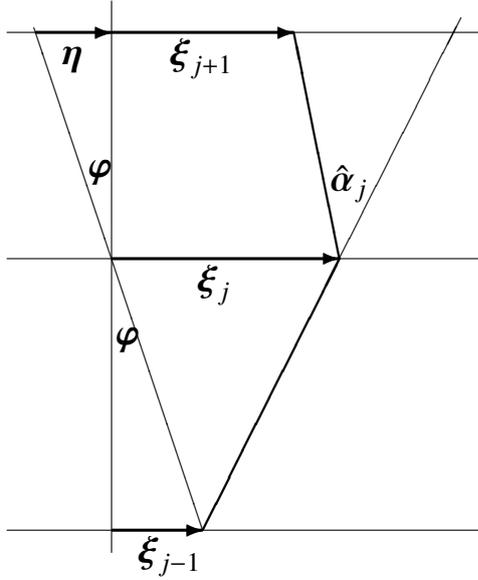
Fig. 2. Propagation of a light ray (thick bent line) between three consecutive planes. The vertical line is the optical axis, with respect to which the separation vectors $\boldsymbol{\xi}$ are measured. The geometry of this figure yields the lens Eq. (30) – see text.

yields

$$\frac{1 + z_{j-1}}{1 + z_j} \mathcal{D}_{j,j+1} \left(\mathcal{D}_{j-1,j}^t\right)^{-1} \mathcal{D}_{j-1} = \mathcal{D}_{j-1,j+1} \mathcal{D}_{j-1,j}^{-1} \mathcal{D}_j - \mathcal{D}_{j+1}. \quad (28)$$

We next consider the prefactor of $\boldsymbol{\xi}_j$ in Eq. (21). Using $\mathsf{C}_{j-1,j}^{-1} \mathcal{D}_{j-1,j}^{-1} = \mathcal{D}_{j-1} \mathcal{D}_j^{-1}$, which is obtained from the definition (17) of C, we find that this prefactor becomes

$$\mathcal{D}_{j+1}\mathcal{D}_j^{-1} + \frac{1 + z_{j-1}}{1 + z_j} \mathcal{D}_{j,j+1} \left(\mathcal{D}_{j-1,j}^{-1}\right)^t \mathcal{D}_{j-1}\mathcal{D}_j^{-1} = \mathcal{D}_{j-1,j+1}\mathcal{D}_{j-1,j}^{-1}, \quad (29)$$

where in the last step we made use of Eq. (28) and the fact that the inversion and transposition operations on a matrix commute. Thus, we can rewrite Eq. (21) in the form

$$\boldsymbol{\xi}_{j+1} = \mathcal{D}_{j-1,j+1}\mathcal{D}_{j-1,j}^{-1}\boldsymbol{\xi}_j$$
$$- \frac{1 + z_{j-1}}{1 + z_j} \mathcal{D}_{j,j+1}\left(\mathcal{D}_{j-1,j}^{-1}\right)^t \boldsymbol{\xi}_{j-1} - \mathcal{D}_{j,j+1}\hat{\boldsymbol{\alpha}}_j(\boldsymbol{\xi}_j). \quad (30)$$

We note that this generalizes equation (4.47) of SSE to the case of general distance matrices between main lens planes. This form of the lens equation can be immediately interpreted geometrically. For this, we consider Fig. 2, from which we read off

$$\boldsymbol{\eta} + \boldsymbol{\xi}_{j+1} = \mathcal{D}_{j-1,j+1}\mathcal{D}_{j-1,j}^{-1}\boldsymbol{\xi}_j - \mathcal{D}_{j,j+1}\hat{\boldsymbol{\alpha}}_j.$$

Furthermore, $\boldsymbol{\eta} = \mathcal{D}_{j,j+1}\boldsymbol{\varphi}$; on the other hand, $\boldsymbol{\xi}_{j-1} = -\mathcal{D}_{j,j-1}\boldsymbol{\varphi}$. Eliminating $\boldsymbol{\varphi}$ from these two relations and making use of Eq. (24), we find

$$\boldsymbol{\eta} = \frac{1 + z_{j-1}}{1 + z_j} \mathcal{D}_{j,j+1}\left(\mathcal{D}_{j-1,j}^{-1}\right)^t \boldsymbol{\xi}_{j-1}.$$

Together, these two equations reproduce Eq. (30). In this form, the equation not only is confined to three consecutive lens planes, but all distance matrices occurring here are those between these three planes.

In order to show the equivalence of Eqs. (10) and (21), it is useful to rewrite the prefactor of $\boldsymbol{\xi}_{j-1}$ in Eq. (30) in a different form. Making use again of Eq. (28), we obtain

$$\boldsymbol{\xi}_{j+1} = \mathcal{D}_{j-1,j+1}\mathcal{D}_{j-1,j}^{-1}\boldsymbol{\xi}_j$$
$$+ \left(\mathcal{D}_{j+1}D_{j-1}^{-1} - \mathcal{D}_{j-1,j+1}\mathcal{D}_{j-1,j}^{-1}\mathcal{D}_j\mathcal{D}_{j-1}^{-1}\right)\boldsymbol{\xi}_{j-1}$$
$$- \mathcal{D}_{j,j+1}\hat{\boldsymbol{\alpha}}_j(\boldsymbol{\xi}_j). \quad (31)$$

We prove the equivalence by induction; for $j = 1$, this equivalence is seen from Eq. (18). Hence we assume that it is true for all planes up to j. Then, taking the difference between Eq. (10) for $\boldsymbol{\xi}_{j+1}$ and Eq. (31),

$$\Delta = \mathcal{D}_{j-1,j+1}\mathcal{D}_{j-1,j}^{-1}\boldsymbol{\xi}_j$$
$$+ \left(\mathcal{D}_{j+1}D_{j-1}^{-1} - \mathcal{D}_{j-1,j+1}\mathcal{D}_{j-1,j}^{-1}\mathcal{D}_j\mathcal{D}_{j-1}^{-1}\right)\boldsymbol{\xi}_{j-1} - \mathcal{D}_{j,j+1}\hat{\boldsymbol{\alpha}}_j(\boldsymbol{\xi}_j)$$
$$- \mathcal{D}_{j+1}\boldsymbol{\theta} + \sum_{i=1}^{j} \mathcal{D}_{i,j+1}\,\hat{\boldsymbol{\alpha}}_i(\boldsymbol{\xi}_i), \quad (32)$$

we need to show that $\Delta = 0$. We first replace $\boldsymbol{\xi}_{j-1}$ and $\boldsymbol{\xi}_j$ by their expressions from Eq. (10), which holds because of the induction assumption,

$$\Delta = \mathcal{D}_{j-1,j+1}\mathcal{D}_{j-1,j}^{-1}\left(\mathcal{D}_j\boldsymbol{\theta} - \sum_{i=1}^{j-1} \mathcal{D}_{i,j}\,\hat{\boldsymbol{\alpha}}_i(\boldsymbol{\xi}_i)\right)$$
$$+ \left(\mathcal{D}_{j+1} - \mathcal{D}_{j-1,j+1}\mathcal{D}_{j-1,j}^{-1}\mathcal{D}_j\right)\left(\boldsymbol{\theta} - \sum_{i=1}^{j-2} \mathcal{D}_{j-1}^{-1}\mathcal{D}_{i,j-1}\,\hat{\boldsymbol{\alpha}}_i(\boldsymbol{\xi}_i)\right)$$
$$- \mathcal{D}_{j,j+1}\hat{\boldsymbol{\alpha}}_j(\boldsymbol{\xi}_j) - \mathcal{D}_{j+1}\boldsymbol{\theta} + \sum_{i=1}^{j} \mathcal{D}_{i,j+1}\,\hat{\boldsymbol{\alpha}}_i(\boldsymbol{\xi}_i). \quad (33)$$

From this equation, one sees immediately that the terms $\propto \boldsymbol{\theta}$ cancel each other. Second, the two terms $\propto \hat{\boldsymbol{\alpha}}_j$ add up to zero. Third, also the sum of the two terms $\propto \hat{\boldsymbol{\alpha}}_{j-1}$ is zero. Thus, what remains to be shown is that the prefactor of the terms $\propto \hat{\boldsymbol{\alpha}}_i$,

$$K_i = \mathcal{D}_{i,j+1} - \mathcal{D}_{j-1,j+1}\mathcal{D}_{j-1,j}^{-1}\mathcal{D}_{i,j}$$
$$- \left(\mathcal{D}_{j+1} - \mathcal{D}_{j-1,j+1}\mathcal{D}_{j-1,j}^{-1}\mathcal{D}_j\right)\mathcal{D}_{j-1}^{-1}\mathcal{D}_{i,j-1}, \quad (34)$$

for $i \leq j - 2$ vanish. For this, we consider again Eq. (25), setting $r = j - 1$, $s = j$, $t = j + 1$, once with $q = 0$, and once with $q = i$. This then yields

$$\frac{1 + z_{j-1}}{1 + z_j} \mathcal{D}_{j,j+1}\left(\mathcal{D}_{j-1,j}^t\right)^{-1} = \left(\mathcal{D}_{j-1,j+1}\mathcal{D}_{j-1,j}^{-1}\mathcal{D}_j - \mathcal{D}_{j+1}\right)\mathcal{D}_{j-1}^{-1}$$
$$= \left(\mathcal{D}_{j-1,j+1}\mathcal{D}_{j-1,j}^{-1}\mathcal{D}_{i,j} - \mathcal{D}_{i,j+1}\right)\mathcal{D}_{i,j-1}^{-1}. \quad (35)$$

After multiplying by $\mathcal{D}_{i,j-1}$, we see that the final equality shows that $K_i = 0$, which proves that $\Delta = 0$, and thus the equivalence of the two forms of the lens equation. The equation $K_i = 0$ itself provides an interesting relation between distance matrices.

4. Mass-sheet transformation

In standard gravitational lensing, with a single deflector between the source and observer, there is a transformation of the mass distribution of the lens which keeps most observables invariant, the mass-sheet transformation (MST, see Falco et al. 1985). Since this transformation is accompanied by a uniform isotropic scaling in the source plane, all magnifications are scaled by the same factor, so that magnification (and thus observable flux) ratios are unchanged. The MST changes the product of

time delay and Hubble constant, though, and the corresponding degeneracy can thus be broken by measuring the time delay in lens systems, assuming the Hubble constant to be known from other cosmological observations (see Schneider & Sluse 2013, and references therein). S14 has recently shown that a MST also exists in the case of two lens planes and two source planes. In this section, we show that also for perturbed gravitational lens systems, as considered in this paper, such a MST does exist.

4.1. Single main lens plane

We start with the case of a single lens plane, using the lens Eq. (14), and modify the deflection angle $\hat{\alpha}_1(\boldsymbol{\xi}_1)$ to the new form

$$\hat{\alpha}'_1(\boldsymbol{\xi}_1) = \lambda \hat{\alpha}_1(\boldsymbol{\xi}_1) + \mathsf{G}_1 \boldsymbol{\xi}_1, \tag{36}$$

where λ is a real number, and G_1 is a matrix[3]. Throughout this section, a prime denotes a mass-sheet transformed quantity. Thus, the modified deflection angle is a scaled version of the original one, plus a term linear in the impact vector. If G_1 is symmetric, this linear term corresponds to a tidal matrix, i.e., adding a uniform mass sheet to the scaled lens mass distribution, plus an external shear. The modified lens equation then becomes

$$\boldsymbol{\xi}'_2 = \mathcal{D}_2 \boldsymbol{\theta} - \mathcal{D}_{12} \hat{\alpha}'_1(\boldsymbol{\xi}_1) = \mathcal{D}_2 \boldsymbol{\theta} - \mathcal{D}_{12} \left[\lambda \hat{\alpha}_1(\boldsymbol{\xi}_1) + \mathsf{G}_1 \mathcal{D}_1 \boldsymbol{\theta} \right]. \tag{37}$$

As for the orignal MST, we require that the modified impact vector $\boldsymbol{\xi}'_2$ is related to the original one by a uniform, isotropic scaling, $\boldsymbol{\xi}'_2 = \nu_2 \boldsymbol{\xi}_2$, where ν_2 is the scaling factor. Thus we require

$$\mathcal{D}_2 \boldsymbol{\theta} - \mathcal{D}_{12} \left[\lambda \hat{\alpha}_1(\boldsymbol{\xi}_1) + \mathsf{G}_1 \mathcal{D}_1 \boldsymbol{\theta} \right] = \nu_2 \left[\mathcal{D}_2 \boldsymbol{\theta} - \mathcal{D}_{12} \hat{\alpha}_1(\boldsymbol{\xi}_1) \right]. \tag{38}$$

In order to have the terms $\propto \hat{\alpha}_1$ equal on both sides of Eq. (38), we need to set $\nu_2 = \lambda$, as is also the case for the MST in standard lensing – the scaling of the source plane (here plane number 2) is the same as that of the deflection angle. The remaining terms are all $\propto \boldsymbol{\theta}$, and setting them equal on both side leads to $\mathcal{D}_2 - \mathcal{D}_{12} \mathsf{G}_1 \mathcal{D}_1 = \lambda \mathcal{D}_2$, or

$$\mathsf{G}_1 = (1 - \lambda) \mathcal{D}_{12}^{-1} \mathcal{D}_2 \mathcal{D}_1^{-1} = (1 - \lambda) \mathsf{C}_{12}. \tag{39}$$

Since C_{12} is symmetric (see Schneider 1997), G_1 is indeed a tidal matrix. This single-main plane MST was also derived by McCully et al. (2014). Thus, in a generalized gravitational lens situation, the MST requires a shear in addition to a uniform mass sheet[4].

4.2. Two main lens planes

We now consider a second lens plane at z_2, with the source plane being located at z_3. The modified lens equation then reads

$$\boldsymbol{\xi}'_3 = \mathcal{D}_3 \boldsymbol{\theta} - \mathcal{D}_{13} \hat{\alpha}'_1(\boldsymbol{\xi}'_1) - \mathcal{D}_{23} \hat{\alpha}'_2(\boldsymbol{\xi}'_2), \tag{40}$$

---

[3] The use of the same symbol λ for the affine parameter and the MST parameter is due to the conventions in the literature, but should not lead to any confusion; in particular, in this section λ is exclusively used as MST parameter, and we use redshift z to label lens planes.

[4] In case the distance matrices are proportional to the unit matrix, the transformation reduces to the known one in standard lensing. Note that the transformation (36) implies the transformation $\hat{\psi}'_1(\boldsymbol{\xi}_1) = \lambda \hat{\psi}_1(\boldsymbol{\xi}_1) + \boldsymbol{\xi}_1^t \mathsf{G}_1 \boldsymbol{\xi}_1 / 2$ for the deflection potential. For isotropic distance matrices, $\mathsf{G}_1 = (1 - \lambda) D_2 / (D_1 D_{12}) \mathcal{I}$, so that $\hat{\psi}'_1(\boldsymbol{\xi}_1) = \lambda \hat{\psi}_1(\boldsymbol{\xi}_1) + (1 - \lambda)(D_1 D_2 / D_{12}) |\boldsymbol{\theta}|^2 / 2$. According to (15), this then implies for the scaled deflection potential $\psi'(\boldsymbol{\theta}) = \lambda \psi(\boldsymbol{\theta}) + (1 - \lambda)|\boldsymbol{\theta}|^2 / 2$, as in standard lens theory.

and we require the modified deflection angle $\hat{\alpha}'_2$ to be chosen such that the 3-plane is just uniformly scaled relative to the original one, i.e., $\boldsymbol{\xi}'_3 = \nu_3 \boldsymbol{\xi}_3$. This condition then yields

$$\mathcal{D}_3 \boldsymbol{\theta} - \mathcal{D}_{13} \left[\lambda \hat{\alpha}_1(\boldsymbol{\xi}_1) + \mathsf{G}_1 \boldsymbol{\xi}_1 \right] - \mathcal{D}_{23} \hat{\alpha}'_2(\boldsymbol{\xi}'_2)$$
$$= \nu_3 \left[\mathcal{D}_3 \boldsymbol{\theta} - \mathcal{D}_{13} \hat{\alpha}_1(\boldsymbol{\xi}_1) - \mathcal{D}_{23} \hat{\alpha}_2(\boldsymbol{\xi}_2) \right]. \tag{41}$$

In order to account for the term $\propto \hat{\alpha}_2$ on the r.h.s. of Eq. (41), the modified deflection has to be of the form[5]

$$\hat{\alpha}'_2(\boldsymbol{\xi}'_2) = \nu_3 \hat{\alpha}_2(\boldsymbol{\xi}'_2 / \lambda) + \mathsf{G}_2 \boldsymbol{\xi}'_2$$
$$= \nu_3 \hat{\alpha}_2(\boldsymbol{\xi}'_2 / \lambda) + \lambda \mathsf{G}_2 \left[\mathcal{D}_2 \boldsymbol{\theta} - \mathcal{D}_{12} \hat{\alpha}_1(\boldsymbol{\xi}_1) \right]. \tag{42}$$

This choice then yields equal terms $\propto \hat{\alpha}_2$ on both sides. Equating the terms $\propto \hat{\alpha}_1$ leads to the condition $-\lambda \mathcal{D}_{13} + \lambda \mathcal{D}_{23} \mathsf{G}_2 \mathcal{D}_{12} = -\nu_3 \mathcal{D}_{13}$, or

$$\mathsf{G}_2 = (1 - \nu_3 / \lambda) \mathcal{D}_{23}^{-1} \mathcal{D}_{13} \mathcal{D}_{12}^{-1}. \tag{43}$$

Using the same arguments as in the Appendix of Schneider (1997), it is straightforward to show that any combination of distance matrices of the form

$$\mathcal{D}_{st}^{-1} \mathcal{D}_{rt} \mathcal{D}_{rs}^{-1} \quad \text{is symmetric,} \tag{44}$$

for $0 \leq z_r < z_s < z_t$. Hence, G_2 is symmetric, and thus corresponds to a tidal matrix. Equating the terms $\propto \boldsymbol{\theta}$ in Eq. (41) then leads to

$$(1 - \nu_3) \mathcal{D}_3 = (1 - \lambda) \mathcal{D}_{13} \mathcal{D}_{12}^{-1} \mathcal{D}_2 + \lambda (1 - \nu_3 / \lambda) \mathcal{D}_{13} \mathcal{D}_{12}^{-1} \mathcal{D}_2, \tag{45}$$

which has the unique solution $\nu_3 = 1$. Thus, as is the case for the standard multi-plane lens discussed in Schneider (2014), the MST does lead to no scaling in the plane $j = 3$. Therefore,

$$\mathsf{G}_2 = (1 - 1/\lambda) \mathcal{D}_{23}^{-1} \mathcal{D}_{13} \mathcal{D}_{12}^{-1}. \tag{46}$$

The implied scaling of the mass distribution in the plane $i = 2$ that follows from Eq. (42) is discussed in Schneider (2014); in short, the surface mass density distribution giving rise to $\hat{\alpha}_2(\boldsymbol{\theta}_2)$ needs to be scaled in amplitude and scale-length to yield a deflection $\hat{\alpha}_2(\lambda \boldsymbol{\theta}_2)$.

4.3. Arbitrary number of planes

Here we generalize the MST to an arbitrary number of source/lens planes. It turns out that the lens equation in the form (31) is better suited for that purpose. We write it in the form

$$\boldsymbol{\xi}_{j+1} = \mathcal{D}_{j-1,j+1} \mathcal{D}_{j-1,j}^{-1} \boldsymbol{\xi}_j + \mathsf{B}_j \boldsymbol{\xi}_{j-1} - \mathcal{D}_{j,j+1} \hat{\alpha}_j(\boldsymbol{\xi}_j), \tag{47}$$

where B_j is the term in parenthesis in Eq. (31). We now assume a scaling $\boldsymbol{\xi}'_j = \nu_j \boldsymbol{\xi}_j$ in every plane, and set the scaled deflection angles to be

$$\hat{\alpha}'_j(\boldsymbol{\xi}'_j) = \nu_{j+1} \hat{\alpha}_j(\boldsymbol{\xi}'_j / \nu_j) + \mathsf{G}_j \boldsymbol{\xi}'_j = \nu_{j+1} \hat{\alpha}_j(\boldsymbol{\xi}_j) + \nu_j \mathsf{G}_j \boldsymbol{\xi}_j. \tag{48}$$

Note that Eqs. (36) and (42) are special cases of the relation (48) for $j = 1, 2$, respectively, and $\nu_1 = 1$, $\nu_2 = \lambda$, $\nu_3 = 1$. Then, from $\boldsymbol{\xi}'_j = \nu_j \boldsymbol{\xi}_j$, we obtain from Eq. (47)

$$\boldsymbol{\xi}'_{j+1} = \nu_j \mathcal{D}_{j-1,j+1} \mathcal{D}_{j-1,j}^{-1} \boldsymbol{\xi}_j + \nu_{j-1} \mathsf{B}_j \boldsymbol{\xi}_{j-1}$$
$$\qquad - \nu_{j+1} \mathcal{D}_{j,j+1} \hat{\alpha}_j(\boldsymbol{\xi}_j) - \nu_j \mathcal{D}_{j,j+1} \mathsf{G}_j \boldsymbol{\xi}_j \tag{49}$$
$$\qquad = \nu_{j+1} \left[\mathcal{D}_{j-1,j+1} \mathcal{D}_{j-1,j}^{-1} \boldsymbol{\xi}_j + \mathsf{B}_j \boldsymbol{\xi}_{j-1} - \mathcal{D}_{j,j+1} \hat{\alpha}_j(\boldsymbol{\xi}_j) \right].$$

---

[5] This relation is unique for a general deflection law $\hat{\alpha}_2$. However, if the "strong lens plane" at z_2 only yields a deflection linear in the impact vector (i.e., if this lens plane is in fact just a "weak" deflector), then the form of $\hat{\alpha}'_2$ no longer is uniquely determined. This comment also applies to the more general case discussed below. Hence, we implicitly assume that all "strong lens planes" yield indeed a non-linear deflection law.

The terms $\propto \hat{\alpha}_j$ cancel each other. The terms $\propto \boldsymbol{\xi}_{j-1}$ yield the condition $\nu_{j+1} = \nu_{j-1}$, and equating the terms $\propto \boldsymbol{\xi}_j$ yields

$$\mathsf{G}_j = (1 - \nu_{j+1}/\nu_j)\mathcal{D}_{j,j+1}^{-1}\mathcal{D}_{j-1,j+1}\mathcal{D}_{j-1,j}^{-1}, \tag{50}$$

which is symmetric according to Eq. (44) and thus represents a tidal matrix. Thus, we obtain $\nu_j = \lambda$ for j even, and $\nu_j = 1$ for j odd. Correspondingly,

$$\mathsf{G}_j = (1 - 1/\lambda)\mathcal{D}_{j,j+1}^{-1}\mathcal{D}_{j-1,j+1}\mathcal{D}_{j-1,j}^{-1} \quad \text{for } j \text{ even};$$
$$\mathsf{G}_j = (1 - \lambda)\mathcal{D}_{j,j+1}^{-1}\mathcal{D}_{j-1,j+1}\mathcal{D}_{j-1,j}^{-1} \quad \text{for } j \text{ odd.} \tag{51}$$

We note that Eqs. (39) and (43) are special cases of Eq. (51).

Hence we find that the MST in multiple (lens and source) plane gravitational lensing exhibits a curious behavior: The scaling factor in every second plane is just unity, whereas it is λ in the other half of the planes. In particular that means that a "standard candle" or "standard rod" in one of the planes with j odd cannot be used to break the degeneracy related to the MST, as the images of these sources are unaffected by the MST. The prefactor in the tidal matrices G_j are positive on every other plane, and negative on the remaining ones. If one disregards the perturbations between lens planes, so that the distance matrices $\mathcal{D}_{ij}$ reduce to angular diameter distances D_{ij}, then the G_j become scalars proportional to the density of uniform mass sheets; in this case, positive and negative densities of these sheets alternate.

4.4. Transformation of the time delay

We next consider how the MST affects the time delays. For that, we assume to have a source on plane number $N + 1$, with its light being deflected in N main lens planes. The corresponding LTTF is given in Eq. (19), where $\boldsymbol{\xi}_{N+1}$ is the position of the source in its source plane.

To obtain the corresponding function $c\tau'$ after the MST, we first need to consider the transformation of the potential time delay. That is, we need to find the transformed deflection potential $\hat{\psi}'_i(\boldsymbol{\xi}'_i)$, which needs to satisfy $\nabla_{\boldsymbol{\xi}'_i}\hat{\psi}'_i(\boldsymbol{\xi}'_i) = \hat{\alpha}'_i(\boldsymbol{\xi}'_i)$. For this, we make the ansatz $\hat{\psi}'_i(\boldsymbol{\xi}'_i) = a\hat{\psi}_i(\boldsymbol{\xi}'_i/\nu_i) + \boldsymbol{\xi}'^{\mathrm{t}}_i\mathsf{G}_i\boldsymbol{\xi}'_i/2$. Taking the gradient yields $\nabla_{\boldsymbol{\xi}'_i}\hat{\psi}'_i(\boldsymbol{\xi}'_i) = (a/\nu_i)\hat{\alpha}_i(\boldsymbol{\xi}'_i/\nu_i) + \mathsf{G}_i\boldsymbol{\xi}'_i$. This is seen to agree with $\hat{\alpha}'_i(\boldsymbol{\xi}'_i)$ in Eq. (48), provided $a = \nu_i\nu_{i+1} = \lambda$. Thus,

$$\hat{\psi}'_i(\boldsymbol{\xi}'_i) = \lambda\hat{\psi}_i(\boldsymbol{\xi}'_i/\nu_i) + \frac{1}{2}\boldsymbol{\xi}'^{\mathrm{t}}_i\mathsf{G}_i\boldsymbol{\xi}'_i. \tag{52}$$

We then obtain for the transformed LTTF

$$c\tau' = \sum_{i=1}^{N}(1 + z_i)\left[\frac{1}{2}\left(\nu_i\mathsf{C}_{i,i+1}\boldsymbol{\xi}_i - \nu_{i+1}\mathcal{D}_{i,i+1}^{-1}\boldsymbol{\xi}_{i+1}\right)^{\mathrm{t}}\mathsf{C}_{i,i+1}^{-1}\right.$$
$$\times \left(\nu_i\mathsf{C}_{i,i+1}\boldsymbol{\xi}_i - \nu_{i+1}\mathcal{D}_{i,i+1}^{-1}\boldsymbol{\xi}_{i+1}\right)$$
$$\left. - \lambda\hat{\psi}_i(\boldsymbol{\xi}_i) - \frac{\nu_i^2}{2}\boldsymbol{\xi}_i^{\mathrm{t}}\mathsf{G}_i\boldsymbol{\xi}_i\right]. \tag{53}$$

We now show that

$$\tau' = \lambda\tau + F(\boldsymbol{\xi}_{N+1}), \tag{54}$$

which means that the transformed LTTF just scales by a factor λ, independent of the plane on which the source is located, plus a function which only depends on the location of the source, and thus cancels when considering time delays, i.e., differences between τ for pairs of multiple images of the source.

Taking the difference of $\tau' - \lambda\tau$, we first note that the terms $\propto \hat{\psi}_i$ drop out. Second, we note that the terms containing products of $\boldsymbol{\xi}_i$ and $\boldsymbol{\xi}_{i+1}$ also cancel, since $\nu_i\nu_{i+1} = \lambda$. Thus we find that

$$c(\tau' - \lambda\tau) = \sum_{i=1}^{N}\frac{(1 + z_i)}{2}\left\{\boldsymbol{\xi}_i^{\mathrm{t}}\left[(\nu_i^2 - \lambda)\mathsf{C}_{i,i+1} - \nu_i^2\mathsf{G}_i\right]\boldsymbol{\xi}_i\right.$$
$$\left. + (\nu_{i+1}^2 - \lambda)\boldsymbol{\xi}_{i+1}^{\mathrm{t}}\left(\mathcal{D}_{i,i+1}^{-1}\right)^{\mathrm{t}}\mathsf{C}_{i,i+1}^{-1}\mathcal{D}_{i,i+1}^{-1}\boldsymbol{\xi}_{i+1}\right\}. \tag{55}$$

The term $i = 1$ of the first sum vanishes, since $\nu_1 = 1$, and Eq. (39) holds. The final term ($i = N$) of the second sum depends only on the source position, and thus corresponds to the function $F(\boldsymbol{\xi}_{N+1})$ previously mentioned. Since this term is of no interest, we simply drop it from now on. Relabeling the index of the second sum as $i \to i - 1$, we then get

$$c(\tau' - \lambda\tau) = \boldsymbol{\xi}_i^{\mathrm{t}}\left[\sum_{i=2}^{N}\frac{(1 + z_i)}{2}(\nu_i^2 - \lambda)\mathsf{K}_i\right]\boldsymbol{\xi}_i, \tag{56}$$

where the matrices K_i are given as

$$\mathsf{K}_i = \frac{1 + z_{i-1}}{1 + z_i}\left(\mathcal{D}_{i-1,i}^{-1}\right)^{\mathrm{t}}\mathsf{C}_{i-1,i}^{-1}\mathcal{D}_{i-1,i}^{-1} + \mathsf{C}_{i,i+1} - \left(\frac{\nu_i^2}{\nu_i^2 - \lambda}\right)\mathsf{G}_i. \tag{57}$$

Since, according to the definition (17), $\mathsf{C}_{i-1,i}^{-1}\mathcal{D}_{i-1,i}^{-1} = \mathcal{D}_{i-1}\mathcal{D}_i^{-1}$, we can rewrite K_i as

$$\mathsf{K}_i = \frac{1 + z_{i-1}}{1 + z_i}\left(\mathcal{D}_{i-1,i}^{-1}\right)^{\mathrm{t}}\mathcal{D}_{i-1}\mathcal{D}_i^{-1} + \mathsf{C}_{i,i+1} - \left(\frac{\nu_i^2}{\nu_i^2 - \lambda}\right)\mathsf{G}_i. \tag{58}$$

From Eq. (29),

$$\frac{1 + z_{i-1}}{1 + z_i}\left(\mathcal{D}_{i-1,i}^{-1}\right)^{\mathrm{t}}\mathcal{D}_{i-1}\mathcal{D}_i^{-1} = \mathcal{D}_{i,i+1}^{-1}\mathcal{D}_{i-1,i+1}\mathcal{D}_{i-1,i}^{-1} - \mathcal{D}_{i,i+1}^{-1}\mathcal{D}_{i+1}\mathcal{D}_i^{-1} \tag{59}$$

is obtained. Noting that the final term is $\mathsf{C}_{i,i+1}$, we obtain with Eq. (50) that

$$\mathsf{K}_i = \left[1 - \left(\frac{\nu_i^2}{\nu_i^2 - \lambda}\right)\left(1 - \frac{\nu_{i+1}}{\nu_i}\right)\right]\mathcal{D}_{i,i+1}^{-1}\mathcal{D}_{i-1,i+1}\mathcal{D}_{i-1,i}^{-1}. \tag{60}$$

However, the prefactor vanishes: if i is odd, $\nu_i = 1$, $\nu_{i+1} = \lambda$, and thus

$$1 - \left(\frac{\nu_i^2}{\nu_i^2 - \lambda}\right)\left(1 - \frac{\nu_{i+1}}{\nu_i}\right) = 1 - \left(\frac{1}{1 - \lambda}\right)(1 - \lambda) = 0.$$

If i is even, $\nu_i = \lambda$, $\nu_{i+1} = 1$, and

$$1 - \left(\frac{\nu_i^2}{\nu_i^2 - \lambda}\right)\left(1 - \frac{\nu_{i+1}}{\nu_i}\right) = 1 - \left(\frac{\lambda^2}{\lambda^2 - \lambda}\right)\left(1 - \frac{1}{\lambda}\right) = 0.$$

Therefore, $\mathsf{K}_i = 0$ for all i, which completes our proof of the validity of Eq. (54). Since the result is independent of the plane on which the source is located – the time delay scales for all source planes with λ – we see that all time delays are scaled by the factor λ under the MST. In particular this implies that the degeneracy due to the MST cannot be broken from measuring time delay ratios.

5. Discussion

In this paper we have considered several aspects of the generalized multi-plane gravitational lensing equation. In contrast to the treatment in Schneider et al. (1992) and more recent papers (e.g., McCully et al. 2014), we treat the light propagation between main lens planes with a continuous formalism, offered by the optical tidal equation, instead of slicing up the matter into several "weak-lensing" lens planes[6]. For this, we made use of the formalism of light propagation in arbitrary spacetimes, as given in SSE. As a result, the distance matrices between lens planes are not written in terms of recursion relations, but as solutions of the optical tidal equation; the explicit solution in terms of an integral over the tidal field caused by large-scale density inhomogeneities along the line-of-sight is provided in Eq. (11).

The time-delay function for generalized multi-plane lensing was derived, using the same arguments as employed in Schneider et al. (1992) for the derivation of the time delay in ordinary multi-plane lensing. The explicit form deviates from that obtained in McCully et al. (2014), in that our result depends only on the impact vectors in the various main planes, but not on the deflection angles. In other word, our expression for τ yields the light travel time of a kinematically possible ray with specified impact vectors ξ_i in the main lens planes, up to an additive constant. Physical light rays are those for which the light-travel time is stationary; this allows the derivation of an iterative lens equation which relates the impact vectors of three consecutive main lens planes to the deflection angle in the middle one of those. We have shown that this form of the lens equation is equivalent to the more standard one which contains the impact vectors and deflection angles of all earlier lens planes. This consecutive lens equation is probably preferable for the use in ray-tracing simulations (see, e.g. Petkova et al. 2014; Jaroszyński & Skowron 2016)[7].

Finally, we showed that the generalized multi-plane lensing admits a mass-sheet transformation (MST) which leaves all observables but the time delay invariant. In contrast to ordinary lensing, the MST corresponds to adding a tidal matrix in each main lens plane. We obtained the curious behavior that the uniform isotropic scaling of the source/lens planes, which is the key aspect of the MST, alternates between planes; in every second plane, the scaling corresponds to the MST parameter λ, in the other half of the planes, the scaling is unity. In particular, this implies that the magnification of sources living on the odd planes with scaling factor unity, is unaffected by the MST. All time delays – i.e., for sources in all main planes – scale as λ under the MST.

This curious behavior of the MST in multi-plane lensing may indeed offer a way to break the corresponding degeneracy, at least in a statistical way. As we discussed before, in the case of vanishing perturbations between the main lens planes, the MST corresponds to mass sheets of alternating sign from plane to plane. Since such a mass sheet changes the slope of the total mass distribution, it means that this slope change also alternates. If one now makes the perhaps plausible assumption that the shape of the mean mass profiles of lenses is the same, this alternating slope change would violate the universality of the mean mass profile. Thus, in multi-plane lensing, the mass-sheet generacy may be more easily lifted than in the case of a single lens plane only.

We hope that the results obtained here will be useful for further theoretical studies of generalized multi-plane lensing, as well as for modeling lens systems in which more than one main deflector affects the imaging properties between observer's sky and the source plane.

---

[6] Whereas these two treatments are equivalent (indeed, as shown in SSE, the slicing into weak-lensing planes corresponds to a discretized version of the optical tidal equation), the continuous formalism is more convenient for analytical calculations – for example, obtaining a result such as (25) using the discretized version is probably extremely tedious.

[7] The advantage of Eqs. (21) and (10) is twofold: First, in order to calculate the impact vectors in all N planes requires of order $N^2/2$ multiplications for each light ray when Eq. (10) is used, compared to about $3N$ multiplications for Eq. (21). More significant, however, is the fact that Eq. (21) allows one to save on memory: whereas Eq. (10) requires the information of all lens planes for each ray, one can treat with Eq. (21) a large set of rays, tracing them from plane to plane, and in each step require only the information on a single lens plane.

References

Bartelmann, M. 2010, Classical Quantum Gravity, 27, 233001
Bartelmann, M., & Schneider, P. 2001, Phys. Rep., 340, 291
Bartelmann, M., Huss, A., Colberg, J. M., Jenkins, A., & Pearce, F. R. 1998, A&A, 330, 1
Bayliss, M. B., Johnson, T., Gladders, M. D., Sharon, K., & Oguri, M. 2014, ApJ, 783, 41
Blandford, R., & Narayan, R. 1986, ApJ, 310, 568
Chae, K.-H., Mao, S., & Augusto, P. 2001, MNRAS, 326, 1015
Collett, T. E., & Auger, M. W. 2014, MNRAS, 443, 969
Collett, T. E., Marshall, P. J., Auger, M. W., et al. 2013, MNRAS, 432, 679
Etherington, I. M. H. 1933, Philos. Mag., 15, 761
Falco, E. E., Gorenstein, M. V., & Shapiro, I. I. 1985, ApJ, 289, L1
Gavazzi, R., Treu, T., Koopmans, L. V. E., et al. 2008, ApJ, 677, 1046
Greene, Z. S., Suyu, S. H., Treu, T., et al. 2013, ApJ, 768, 39
Hilbert, S., Hartlap, J., White, S. D. M., & Schneider, P. 2009, A&A, 499, 31
Hoekstra, H. 2013, ArXiv e-prints [arXiv:1312.5981]
Jain, B., Seljak, U., & White, S. 2000, ApJ, 530, 547
Jaroszynski, M., & Kostrzewa-Rutkowska, Z. 2014, MNRAS, 439, 2432
Jaroszyński, M., & Skowron, J. 2016, MNRAS, 462, 1405
Keeton, C. R., Kochanek, C. S., & Seljak, U. 1997, ApJ, 482, 604
Kneib, J.-P., & Natarajan, P. 2011, A&ARv, 19, 47
Kochanek, C. S. 2003, ApJ, 583, 49
Kochanek, C. S. 2006, in Saas-Fee Advanced Course 33: Gravitational Lensing: Strong, Weak and Micro, eds. G. Meylan, P. Jetzer, P. North, et al., 91
Kovner, I. 1987, ApJ, 316, 52
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. 2017, ApJ, 836, 141
Meneghetti, M., Bartelmann, M., Dahle, H., & Limousin, M. 2013, Space Sci. Rev., 177, 31
Munshi, D., Valageas, P., van Waerbeke, L., & Heavens, A. 2008, Phys. Rep., 462, 67
Petkova, M., Metcalf, R. B., & Giocoli, C. 2014, MNRAS, 445, 1954
Puchwein, E., & Hilbert, S. 2009, MNRAS, 398, 1298
Refsdal, S. 1964, MNRAS, 128, 307
Refsdal, S. 1970, ApJ, 159, 357
Schneider, P. 1997, MNRAS, 292, 673
Schneider, P. 2006, in Saas-Fee Advanced Course 33: Gravitational Lensing: Strong, Weak and Micro, eds. G. Meylan, P. Jetzer, P. North, et al., 269
Schneider, P. 2014, A&A, 568, L2
Schneider, P. 2016, A&A, 592, L6
Schneider, P., & Sluse, D. 2013, A&A, 559, A37
Schneider, P., & Sluse, D. 2014, A&A, 564, A103
Schneider, P., Ehlers, J., & Falco, E. E. 1992, Gravitational Lenses (Berlin: Springer-Verlag)
Seitz, S., & Schneider, P. 1994, A&A, 287, 349
Seitz, S., Schneider, P., & Ehlers, J. 1994, Classical Quantum Gravity, 11, 2345
Smith, M., Bacon, D. J., Nichol, R. C., et al. 2014, ApJ, 780, 24
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
Treu, T. 2010, ARA&A, 48, 87
Wong, K. C., Keeton, C. R., Williams, K. A., Momcheva, I. G., & Zabludoff, A. I. 2011, ApJ, 726, 84

Appendix A: Distance matrices in terms of peculiar gravitational potential

In this appendix we derive the expression (11) for the distance matrices in an inhomogeneous Universe. As usual in cosmological weak lensing, we work in comoving coordinates, and therefore replace the affine parameter λ by the comoving distance χ, where $d\chi/da = c/(a^2 H)$, and a is the cosmic scale factor normalized to unity today. From the Robertson–Walker metric and the condition for radial null geodesics, we have $a\,d\chi = -c\,dt$, and with Eq. (2) follows $d\chi = d\lambda/a^2$. These relations then imply that

$$\frac{d}{d\lambda} = \frac{d\chi}{d\lambda}\frac{d}{d\chi} = \frac{1}{a^2}\frac{d}{d\chi},$$
$$\frac{d^2}{d\lambda^2} = \frac{1}{a^4}\frac{d^2}{d\chi^2} - \frac{2}{a^3}\frac{H}{c}\frac{d}{d\chi}, \qquad (A.1)$$
$$a\frac{d^2}{d\lambda^2} + \frac{2H}{c}\frac{d}{d\lambda} = \frac{1}{a^3}\frac{d^2}{d\chi^2}.$$

The angular-diameter distance D_i is related to the comoving angular diameter distance by $D_i = a\,f_k(\chi - \chi_i)$, where χ_i is the comoving distance corresponding to the affine parameter λ_i. Applying Eq. (5), we get

$$\frac{d^2 D_i}{d\lambda^2} = a\frac{d^2 f_k}{d\lambda^2} + 2\frac{da}{d\lambda}\frac{df_k}{d\lambda} + \frac{d^2 a}{d\lambda^2}f_k = -\frac{3}{2}\left(\frac{H_0}{c}\right)^2\frac{\Omega_m}{a^4}f_k. \qquad (A.2)$$

With $da/d\lambda = (d\chi/d\lambda)(da/d\chi) = H/c$, and

$$\frac{d^2 a}{d\lambda^2} = \frac{dH}{c\,d\lambda} = \frac{da}{d\lambda}\frac{dH}{c\,da} = \frac{1}{2c^2}\frac{dH^2}{da}$$
$$= \frac{H_0^2}{2c^2}\left(2\frac{\Omega_m + \Omega_\Lambda - 1}{a^3} - \frac{3\Omega_m}{a^4}\right),$$

and making use of the relations (A.1), Eq. (A.2) reduces to Eq. (12).

Next we write the distance matrix as $\mathcal{D}_i = D_i\mathsf{B}_i$, so that B_i describes the deviation of $\mathcal{D}_i$ from the unperturbed distance matrix $D_i\mathcal{I}$[8]. This factorization transforms Eq. (8) into

$$\frac{d^2 \mathsf{B}_i}{d\lambda^2}D_i + 2\frac{d\mathsf{B}_i}{d\lambda}\frac{dD_i}{d\lambda} + \mathsf{B}_i\frac{d^2 D_i}{d\lambda^2} = \left(\mathcal{T}_{bg} + \mathcal{T}_{sm}\right)\mathsf{B}_i D_i. \qquad (A.3)$$

Subtracting from this the transport Eq. (5) for D_i, we are left with

$$\frac{d^2 \mathsf{B}_i}{d\lambda^2}D_i + 2\frac{d\mathsf{B}_i}{d\lambda}\frac{dD_i}{d\lambda} = \mathcal{T}_{sm}\mathsf{B}_i D_i. \qquad (A.4)$$

Inserting $D_i(\lambda) = a(\lambda)f_k(\chi - \chi_i)$, and using the differentiation rules (A.1), this turns into

$$\frac{d^2 \mathsf{B}_i}{d\chi^2}f_k(\chi - \chi_i) + 2\frac{d\mathsf{B}_i}{d\chi}\frac{df_k}{d\chi} = \mathcal{T}_{sm}\mathsf{B}_i a^4 f_k(\chi - \chi_i). \qquad (A.5)$$

We this relation, we find for $\mathsf{X}_i(\chi) := f_k(\chi - \chi_i)\mathsf{B}_i(\chi)$:

$$\frac{d^2 \mathsf{X}_i}{d\chi^2} = \frac{d^2 \mathsf{B}_i}{d\chi^2}f_k(\chi - \chi_i) + 2\frac{d\mathsf{B}_i}{d\chi}\frac{df_k}{d\chi} - K\mathsf{X} = -K\mathsf{X}_i + \mathcal{T}_{sm}a^4\mathsf{X}_i, \qquad (A.6)$$

where we made use of Eq. (12). This differential equation can be transformed into an integral equation, using the method of Green's functions, to read

$$\mathsf{X}_i(\chi) = f_k(\chi - \chi_i)\mathcal{I} + \int_{\chi_i}^{\chi} d\chi'\,f_k(\chi - \chi')\mathcal{T}_{sm}(\chi')a^4(\chi')\mathsf{X}_i(\chi'). \qquad (A.7)$$

By differtiating twice, one can easily show that Eq. (A.7) indeed is a formal solution of Eq. (A.6), with the correct initial condition $\mathsf{X}_i(\chi_i) = 0$, $d\mathsf{X}_i(\chi_i)/d\chi = \mathcal{I}$. If we now use the expression (6) for $\mathcal{T}_{sm}$, neglecting the final term which is a derivative along the line-of-sight and thus cancels in the integration, and replacing the derivatives w.r.t. $\boldsymbol{\xi}$ by those w.r.t. comoving transverse coordinates, we get

$$\mathsf{X}_i(\chi) = f_k(\chi - \chi_i)\mathcal{I} - \frac{2}{c^2}\int_{\chi_i}^{\chi} d\chi'\,f_k(\chi - \chi')\mathsf{H}(\phi(\chi'))\mathsf{X}_i(\chi'). \qquad (A.8)$$

Using $\mathcal{D}_i = a\mathsf{X}_i$, we arrive at Eq. (11). We also note that the large-scale structure component of the optical tidal matrix can alternatively written in the form (see Seitz & Schneider 1994)

$$\mathcal{T}_{sm} = -\frac{2}{c^2 a^4}\left[2\pi G a^2(\rho - \bar{\rho}) + \begin{pmatrix} \Gamma_1 & \Gamma_2 \\ \Gamma_2 & -\Gamma_1 \end{pmatrix}\right], \qquad (A.9)$$

with $\Gamma_1 = (\phi_{,11} - \phi_{,22})/2$ and $\Gamma_2 = \phi_{,12}$, where the partial derivatives are with respect to transverse comoving coordinates. Thus, one can replace $\mathsf{H}(\phi)$ in Eq. (A.8) by the bracket in Eq. (A.9).

---

[8] We note that $\mathsf{B} \equiv \mathsf{B}_0$ corresponds to the matrix $\boldsymbol{B}$ in McCully et al. (2014), and the B_i corresponds, up to a prefactor, to their matrices $\boldsymbol{C}$.