

# Light propagation in arbitrary spacetimes and the gravitational lens approximation

Stella Seitz, Peter Schneider and Jürgen Ehlers

Max-Planck-Institut für Astrophysik, Karl-Schwarzschild-Strasse 1, Pf 1523, D85740 Garching, Federal Republic of Germany

Received 9 February 1994

**Abstract.** We reformulate the transport equation which determines the size, shape and orientation of infinitesimal light beams in arbitrary spacetimes. The behaviour of such light beams near vertices and conjugate points is investigated, with special attention to the singular behaviour of the optical scalars. We then specialize the general transport equation to the case of an approximate metric of an inhomogeneous universe, which is a Friedmann metric 'on average' with superposed isolated weak matter inhomogeneities. In a series of well defined approximations, the equations of gravitational lens theory are derived. Finally, we derive a relative optical focusing equation which describes the focusing of light beams relative to the case that the beam is unaffected by matter inhomogeneities in the universe, from which it follows immediately that no beam can be focused less than one which is unaffected by matter clumps, before it propagates through its first conjugate point.

PACS numbers: 0425, 9530S, 9862S

## 1. Introduction

The propagation of light rays in curved spacetimes is described by the equation for null geodesics. Below, we consider congruences of light rays, so called light beams (for an exact definition, see section 2) and study their propagation in arbitrary spacetimes. Infinitesimal light beams are described by Jacobi's differential equation for deviation vectors. In this paper, we study some properties of the solutions of this propagation equation. In particular, we provide a detailed study of the behaviour of light beams near vertices and conjugate points. The behaviour of the optical scalars (Sachs 1961), which may diverge near conjugate points, is determined. We find the leading order behaviour of the convergence, shear and twist of light beams and their relation to the optical tidal matrix which represents the source of beam deformation.

We then specialize the propagation equation to the case where the metric can be described by that of a Friedmann universe, with superposed weak local inhomogeneities; this is the situation most relevant for light propagation in the universe. Here, the optical tidal matrix can be split into a contribution due to the background universe and one due to the local inhomogeneities, which is described in the first post-Minkowskian approximation. The background universe is assumed to have the overall geometry of a smooth Friedmann universe, but is locally modified due to matter inhomogeneities.

If the matter inhomogeneities along the light beam are well localized, i.e. if the spatial extent of the inhomogeneities is much smaller than the distance from the source to an observer, the contributions from the inhomogeneities can be described in the impulse

approximation, in which the contribution to the optical tidal matrix due to inhomogeneities is replaced by a sum of delta distributions. We will then show that this approximation leads to the gravitational lens equations, which are usually used to describe the influence of weak matter inhomogeneities on light propagation in the universe (for a review on gravitational lens theory and its applications, see Schneider, Ehlers and Falco (1992), hereafter SEF). Hence, the gravitational lens equations follow from the exact propagation equations for light beams with a series of well defined approximations.

The behaviour of the cross sectional area of an infinitesimally small light beam is described by the optical focusing equation (Sachs 1961), which contains the trace of the optical tidal matrix and the shear of the light beam as source terms. We will show that a *relative optical focusing equation* can be obtained which describes the cross sectional area of a beam relative to one which is unaffected by matter inhomogeneities. The uniquely determined independent variable for this relative focusing equation is the  $\chi$ -function introduced for other reasons in section 4.6 of SEF. From this relative focusing equation it follows directly that no light beam can be less focused than one which is unaffected by matter inhomogeneities before the beam propagates through its first conjugate point. In the frame of gravitational lens theory this fact has been proved earlier (Schneider 1984, Seitz and Schneider 1992, hereafter paper I, 1994).

## 2. Infinitesimal light beams

In this section we review some consequences of the fact that, according to the geometrical optics approximation to Maxwell's equations in an arbitrary spacetime  $(M, g_{\alpha\beta})$ , a locally nearly plane electromagnetic wave, propagating without interaction with matter, is associated with a hypersurface-orthogonal congruence of null geodesics representing *light rays*. We denote the corresponding phase function by  $S$  and the wave vector by  $k^\alpha = -g^{\alpha\beta} S_{,\beta}$ ; then  $k_\alpha k^\alpha = 0$  and  $\dot{k}^\alpha := k_{\alpha;\beta} k^\beta = 0$ . (For details concerning this section see, e.g., SEF chapter 3, and Wald (1984) sections 9.2 and 9.3, see also Blandford and Narayan (1992).)

We fix attention on one 'central' light ray  $\gamma_0$  and denote by  $Y^\alpha$  any *deviation vector field* (Jacobi field) 'connecting'  $\gamma_0$  to one of its neighbours. Then,  $k_\alpha Y^\alpha$  is constant on  $\gamma_0$ . Deviation vectors differing by a (constant) multiple of  $k^\alpha$  represent displacements to the same nearby ray. Given the 4-velocity  $U^\alpha$  of an observer at an event  $p$  on  $\gamma_0$ , one can always arrange that  $Y^\alpha$  is spatial for  $U^\alpha$ , i.e.  $U_\alpha Y^\alpha = 0$ .

Two events  $p, q$  on  $\gamma_0$  are said to be *conjugate* if there exists a not identically vanishing Jacobi field which is zero at  $p$  and  $q$ . For such a Jacobi field,  $k_\alpha Y^\alpha = 0$ . A deviation vector satisfying the last equation (whether it vanishes somewhere or not) connects rays contained in the same phase hypersurface  $S = \text{const}$ .

Henceforth we consider exclusively 2-parameter families of rays contained in one phase hypersurface which we call *beams*. Their deviation vectors obey  $k_\alpha Y^\alpha = 0$ , consequently the size, shape and orientation of an infinitesimal cross section of a beam is independent of the 4-velocity of the observer who measures it.

Given the 4-velocity  $U^\alpha$  of an observer at an event  $p$  on  $\gamma_0$ , one can choose deviation vectors to all neighbouring rays such that, besides  $k_\alpha Y^\alpha = 0$ , also  $U_\alpha Y^\alpha = 0$ . Such vectors  $Y^\alpha$  span a 2-dimensional, spacelike subspace of the tangent space  $M_p$  of  $p$  which we call a *screen* adapted to  $k^\alpha, U^\alpha$ .

In studying conjugate pairs on a ray  $\gamma_0$  it suffices to consider deviation vectors belonging to a beam surrounding  $\gamma_0$ .

For gravitational lensing, the important beams are those which are contained in either the future *null cone*  $C_s^+$  of an event  $s$ —flashes of light emitted from a source event  $s$ —or

the past null cone  $\mathcal{C}_o^-$  of an observation event  $o$ . (In the second case, the rays of a beam belong to different, usually mutually incoherent, locally plane waves emitted from different source events. This does not matter for the applications considered in this paper. It is often helpful to think of the rays as (classical models of) photons.) In the remainder of this paper we are concerned with such beams only.

$\mathcal{C}_o^-$  is generated by all null geodesic rays ending at  $o$ . The set of all events conjugate to the vertex  $o$  on those rays forms the *caustic* of  $\mathcal{C}_o^-$ .  $\mathcal{C}_o^-$  has the shape of a (hyper-) cone only between  $o$  and the first sheet of the caustic; thereafter in general it bifurcates and intersects itself. This is the (theoretical) reason for the phenomenon of multiple imaging in gravitational lensing.

Consider an observer at the event  $o$  with 4-velocity  $U_o^\alpha$ ,  $U_o^\alpha U_{o\alpha} = 1$ , and the past light cone  $\mathcal{C}_o^-$ . Choose the affine parameter  $\lambda$  of the rays ending at  $o$  such that (i)  $\lambda = 0$  at  $o$ , (ii)  $\lambda$  increases to the past, (iii) at  $o$ ,  $\tilde{k}_\alpha U_o^\alpha = -1$ . Then,  $\tilde{k}^\alpha = dx^\alpha/d\lambda$  is past-directed, and, for events on  $\mathcal{C}_o^-$  infinitely close to  $o$ ,  $d\lambda$  is the distance from  $o$  measured by the chosen observer. The 'new'  $\tilde{k}^\alpha$  is related to the wave vector introduced above by  $k^\alpha = -(\omega_o/c)\tilde{k}^\alpha$  if  $\omega_o$  is the frequency associated with  $k^\alpha$  at the observer.  $\tilde{k}^\alpha$  is purely kinematical, the same for all monochromatic waves which might be travelling in the direction  $-\tilde{k}^\alpha$ . Let  $\gamma_0$  be a ray, and let  $U^\alpha$  on  $\gamma_0$  be the result of parallelly propagating  $U_o^\alpha$ . Choose, along  $\gamma_0$ , orthonormal bases  $(E_1^\alpha, E_2^\alpha)$  on the screens adapted to  $\tilde{k}^\alpha$ ,  $U^\alpha$ , parallel on  $\gamma_0$ . The deviation vectors of the beam centred on  $\gamma_0$  can then be written as  $Y^\alpha = -\xi_1 E_1^\alpha - \xi_2 E_2^\alpha - \xi_o \tilde{k}^\alpha$ ; then the screen components  $\xi_i$  ( $i = 1, 2$ ) change according to the deformation equation

$$\dot{\xi}_i = S_{ij}\xi_j \quad S_{ij} = E_i^\alpha \tilde{k}_{\alpha;\beta} E_j^\beta$$

where a dot denotes differentiation with respect to the affine parameter. In matrix notation we write

$$\dot{\xi} = S\xi. \quad (2.1a)$$

The *optical deformation matrix*  $S$  is composed of Sachs' optical scalars of the beam (Sachs 1961), i.e. its rate of expansion

$$\theta(\lambda) := \frac{1}{2} \tilde{k}_{\alpha;\beta}^{\alpha}(\lambda)$$

and its (complex) rate of shear,

$$\sigma(\lambda) := \frac{1}{2} \tilde{k}_{\alpha;\beta}(\lambda) \epsilon^{*\alpha}(\lambda) \epsilon^{\beta}(\lambda) \quad \epsilon^\alpha := E_1^\alpha + iE_2^\alpha$$

according to

$$S(\lambda) := \begin{pmatrix} [\theta(\lambda) - \text{Re } \sigma(\lambda)] & [\text{Im } \sigma(\lambda)] \\ [\text{Im } \sigma(\lambda)] & [\theta(\lambda) + \text{Re } \sigma(\lambda)] \end{pmatrix}. \quad (2.1b)$$

Since  $k_\alpha = -S_{,\alpha}$  is the gradient of the phase  $S$ ,  $\tilde{k}_{\alpha;\beta} = \tilde{k}_{\beta;\alpha}$ , and therefore  $S$  is a symmetric matrix. Differentiation of (2.1a) with respect to  $\lambda$  gives

$$\dot{\xi}(\lambda) = T(\lambda)\xi(\lambda) \quad (2.2a)$$

where

$$\dot{S} + S^2 = T. \quad (2.2b)$$

Combining the last equation with Sachs' transport equations for  $\theta$  and  $\sigma$ ,

$$\dot{\theta} + \theta^2 + |\sigma|^2 = \mathcal{R} \quad (2.2c)$$

$$\dot{\sigma} + 2\theta\sigma = \mathcal{F} \quad (2.2d)$$

shows that the *optical tidal matrix*  $T$  is given by

$$T(\lambda) := \begin{pmatrix} [\mathcal{R}(\lambda) - \text{Re } \mathcal{F}(\lambda)] & [\text{Im } \mathcal{F}(\lambda)] \\ [\text{Im } \mathcal{F}(\lambda)] & [\mathcal{R}(\lambda) + \text{Re } \mathcal{F}(\lambda)] \end{pmatrix} \quad (2.2e)$$

where

$$\mathcal{R}(\lambda) := -\frac{1}{2} R_{\beta\gamma}(\lambda) \tilde{k}^\beta(\lambda) \tilde{k}^\gamma(\lambda) \quad (2.2f)$$

$$\mathcal{F}(\lambda) := -\frac{1}{2} C_{\alpha\beta\gamma\delta}(\lambda) \epsilon^{\alpha*}(\lambda) \tilde{k}^\beta(\lambda) \epsilon^{\gamma*}(\lambda) \tilde{k}^\delta(\lambda). \quad (2.2g)$$

Similar equations have been derived by Blandford *et al* (1991) and Peebles (1993, chapter 14)†. The optical tidal matrix is symmetric due to the symmetry  $C_{\alpha\beta\gamma\delta} = C_{\delta\gamma\beta\alpha}$  of the conformal curvature tensor. Equations (2.2a), (2.2e)–(2.2g) exhibit how the Ricci and conformal curvatures govern the evolution of infinitesimal light beams; they are equivalent to the *geodesic deviation equation* (Jacobi equation) for screen vectors‡.

The linearity of the Jacobi equation (2.2a) implies that the solution  $\xi(\lambda)$  is related to its initial value  $\xi(0) =: \theta$  by a  $\lambda$ -dependent linear transformation

$$\xi(\lambda) = \mathcal{D}(\lambda)\theta. \quad (2.3)$$

With the choice of  $\lambda$  described above,  $\theta$  is the (vectorial) angle between  $\gamma_0$  and a neighbouring ray. Because of (2.2a),  $\xi(0) = 0$  and  $\dot{\xi}(0) = \theta$ ,  $\mathcal{D}(\lambda)$  is determined by

$$\ddot{\mathcal{D}}(\lambda) = T(\lambda)\mathcal{D}(\lambda) \quad (2.4a)$$

$$\mathcal{D}(0) = \mathcal{O} \quad \dot{\mathcal{D}}(0) = \mathcal{I} \quad (2.4b)$$

or, equivalently, by the linear integral equation

$$\mathcal{D}(\lambda) = \lambda\mathcal{I} + \int_0^\lambda d\lambda' (\lambda - \lambda') T(\lambda') \mathcal{D}(\lambda'). \quad (2.5)$$

The *Jacobi map* (2.3) takes infinitesimal changes of ray directions at the observer back to a screen at an event of  $\gamma_0$  given by the value of  $\lambda$ . If that event is taken on some source 'plane'  $z = \text{const}$ ,  $\mathcal{D}(\lambda)$  corresponds to the properly scaled *magnification matrix* (in the terminology of SEF) of lens theory. Note that in contrast to  $S$  and  $T$ ,  $\mathcal{D}$  is in general not symmetric.

Equation (2.4a) implies:

(1) If  $T(\lambda)$  is continuously differentiable  $k$  times,  $\mathcal{D}(\lambda)$  is continuously differentiable  $k+2$  times; assuming  $k$  sufficiently large (which is permissible) justifies our later use of Taylor polynomials to study the local properties of  $\mathcal{D}(\lambda)$  at special points.

† Note, however, that the component  $\alpha k^\alpha$  of  $Y^\alpha$  cannot be made to vanish for all  $\lambda$ , contrary to the claim in Peebles (consider equation (14.9) in his book, where  $\xi^i$  corresponds to 'our'  $Y^\alpha$ ).

‡ In equation (2.2g) one may write the full curvature tensor instead of  $C_{\alpha\beta\gamma\delta}$ ; the Ricci part does not contribute.

(2)  $\dot{\mathcal{D}}^T \mathcal{D} - \mathcal{D}^T \dot{\mathcal{D}}$  is a first integral of (2.4a). Since it vanishes in consequence of the initial conditions (2.4b), all solutions of (2.4) obey  $\dot{\mathcal{D}}^T \mathcal{D} = \mathcal{D}^T \dot{\mathcal{D}}$ , provided  $\mathcal{T}$  is continuous there. At discontinuities and  $\delta$ -type singularities of  $\mathcal{T}$  this relation is preserved.

According to the definitions given above,  $\lambda_c$  corresponds to a point  $p_c$  conjugate to the vertex (observer) if and only if  $\det \mathcal{D}(\lambda_c) = 0$ . If the rank of  $\mathcal{D}(\lambda_c)$  is equal to zero, i.e. if  $\mathcal{D}(\lambda_c) = \mathcal{O}$ , all rays arriving at  $o$  have been intersecting to first order at  $p_c$ ; if the rank of  $\mathcal{D}(\lambda_c)$  is equal to one, the cross section of the ray bundle has been degenerating into an infinitesimal line segment at  $p_c$ . In the first case,  $p_c$  is called a *focus* (or degenerate conjugate point) of the caustic of  $\mathcal{C}_0^-$ , in the second case, it is said to be a non-degenerate, or simple, conjugate point.

Comparison of (2.1a) and the derivative of (2.3) shows that

$$\dot{\mathcal{D}} = \mathcal{S} \mathcal{D} \quad (2.6)$$

thus  $\mathcal{S}$  can be obtained from  $\mathcal{D}$  (see below, section 3). With (2.6) we alternatively derive the symmetry of the  $\mathcal{S}$  matrix from the 'basic' differential equation (2.4a) and the vertex initial conditions (2.4b): at an affine parameter where  $\mathcal{D}^{-1}$  and thus  $\mathcal{S}$  exist,  $\dot{\mathcal{D}}^T \mathcal{D} = \mathcal{D}^T \dot{\mathcal{D}}$  is equivalent to  $\mathcal{S} = \mathcal{S}^T$ . This also implies that at points where  $\det \mathcal{D} = 0$ , the antisymmetric part of  $\mathcal{S}$  is equal to zero.

Consider now the determinant of the Jacobi map. From its definition in (2.3) it follows that its absolute value is equal to the area  $\delta A(\lambda)$  of the cross section of the light beam at this affine parameter, divided by the solid angle  $\delta \Omega$  which that cross section subtends at the observer:

$$|\det \mathcal{D}(\lambda)| = \frac{\delta A(\lambda)}{\delta \Omega}. \quad (2.7)$$

At a non-degenerate conjugate point the Jacobian determinant changes its sign; at a focus its sign is conserved, as will be shown in section 3. Thus,  $\det \mathcal{D}(\lambda)$  contains information about the area  $\delta A(\lambda)$  as well as the parity, i.e. the orientation of a beam at  $\lambda$  relative to that close to the vertex. Between the vertex and conjugate points, the area  $\delta A(\lambda)$  is governed by Sachs' focusing equation:

$$\left( \sqrt{A(\lambda)} \right)'' = [\mathcal{R}(\lambda) - |\sigma(\lambda)|^2] \sqrt{A(\lambda)}. \quad (2.8a)$$

This ordinary differential equation has  $\mathcal{C}^2$  solutions in any  $\lambda$ -interval in which  $\mathcal{R}(\lambda) - |\sigma(\lambda)|^2$  is continuous. This is the case except if the interval contains simple conjugate points, see section 3. The initial conditions for the solution of (2.8a), which gives the area of the beam, are  $\sqrt{A(0)} = 0$  and  $(d\sqrt{A}/d\lambda)(0) = \Omega$ , where  $\Omega$  is the solid angle of the beam at the observer. If there is an odd number of non-degenerate conjugate points between the observer and  $\lambda$ , one has to take the negative root of  $A$ , otherwise the positive one. The driving term of the focusing equation,  $\mathcal{R} - |\sigma|^2$ , is non-positive: the Einstein field equation with an energy-momentum tensor of an ideal fluid yields a non-positive *source of convergence*  $\mathcal{R}$ ; this also holds for a cosmological constant. Hence, equation (2.8a) describes how a light beam is focused at  $\lambda$  due to the 'local' curvature (Ricci focusing) and due to its own shear rate at this affine parameter. Since this shear rate was produced by the *source of shear*  $\mathcal{F}$  at a smaller  $\lambda$ , this implies that both  $\mathcal{R}$  and  $\mathcal{F}$  yield a focusing of the light beam. Hence, as long as one considers only the area and not the shape of a light

beam, the actions of  $\mathcal{R}$  and  $\mathcal{F}$  are not distinguishable. In the following we do not consider the evolution of the area of a light beam, but that of

$$w(\lambda) := \mathcal{S}Q[\det \mathcal{D}](\lambda) \equiv \text{sign}(\det \mathcal{D}(\lambda)) \sqrt{|\det \mathcal{D}(\lambda)|} \quad (2.9)$$

the absolute value,  $|w|(\lambda) = \sqrt{|\det \mathcal{D}(\lambda)|} = \sqrt{\delta A(\lambda)/\delta \Omega}$ , of this function describes the angular diameter distance along the beam considered, and the sign is the parity of the Jacobi map. From (2.7) we obtain that  $w$  also satisfies the focusing equation

$$\ddot{w}(\lambda) = [\mathcal{R}(\lambda) - |\sigma(\lambda)|^2] w(\lambda) \quad (2.8b)$$

between conjugate points; the initial conditions for  $w$  are:  $w(0) = 0$  and  $\dot{w}(0) = 1$ . It is not clear *a priori* how to connect the solutions between conjugate points with each other, or whether one at all can integrate over conjugate points: the matrix  $\mathcal{S}$  of equation (2.6) and thus  $\theta$  and  $\sigma$  become singular at the vertex and at a conjugate point  $\lambda_c$ . We investigate the behaviour of a light beam near the vertex and a conjugate point in the next section and show that the solution of (2.8) is nevertheless well defined at conjugate points between source and observer.

### 3. The behaviour of light beams near vertices and conjugate points

*Preliminaries: parametrization of a  $2 \times 2$  matrix.* For our further discussion we parametrize a real  $2 \times 2$  matrix  $A$  in terms of† ‘convergence’  $\tau$ , ‘twist’  $\omega$  and ‘shear’  $\Gamma_1$  and  $\Gamma_2$  and write them as real and imaginary parts of complex numbers  $\Lambda$  and  $\Gamma$ , respectively:

$$\tau[A] := \frac{1}{2}(a_{11} + a_{22}) \quad \omega[A] := \frac{1}{2}(a_{12} - a_{21}) \quad \Lambda[A] = \tau[A] + i\omega[A] \quad (3.1)$$

$$\Gamma_1[A] := \frac{1}{2}(a_{11} - a_{22}) \quad \Gamma_2[A] := \frac{1}{2}(a_{12} + a_{21}) \quad \Gamma[A] = \Gamma_1[A] + i\Gamma_2[A]. \quad (3.2)$$

Then, the trace of  $A$  is  $\text{tr } A = 2 \text{Re } \Lambda[A]$  and its determinant is  $\det A = |\Lambda[A]|^2 - |\Gamma[A]|^2$ . Note that transforming  $A$  with a proper orthogonal matrix (rotation matrix)

$$R(\vartheta) = \begin{pmatrix} \cos \vartheta & \sin \vartheta \\ -\sin \vartheta & \cos \vartheta \end{pmatrix}$$

to  $A' = R^{-1}AR$  leaves  $\Lambda$  invariant ( $\Lambda[A'] = \Lambda[A]$ ) and transforms  $\Gamma$  to  $\Gamma[A'] = \Gamma[A]e^{2i\vartheta}$ .  $\Lambda$  and  $|\Gamma|$  have an intrinsic, coordinate-independent meaning for the map given by  $A$ , whereas the phase of  $\Gamma$  fixes the coordinate system to which  $A$  refers. We illustrate our definitions for  $\mathcal{S}$  and  $\mathcal{T}$ :

$$\Lambda[\mathcal{S}](\lambda) = \theta(\lambda) \in \mathbb{R} \quad \Gamma[\mathcal{S}](\lambda) = -\sigma^*(\lambda) \quad (3.3)$$

$$\Lambda[\mathcal{T}](\lambda) = \mathcal{R}(\lambda) \in \mathbb{R} \quad \Gamma[\mathcal{T}](\lambda) = -\mathcal{F}^*(\lambda). \quad (3.4)$$

If the argument of  $\Lambda$  is the Jacobian matrix  $\mathcal{D}$ , we simply write  $\Lambda[\mathcal{D}] =: \Lambda$ , and obtain for the derivatives with respect to  $\lambda$  that  $\Lambda[\dot{\mathcal{D}}](\lambda) = \dot{\Lambda}(\lambda)$  and  $\Lambda[\ddot{\mathcal{D}}](\lambda) = \ddot{\Lambda}(\lambda)$ ; analogous

† These names are chosen for convenience and are not intended to contain a geometrical meaning. In the case of the Jacobi matrix, the geometrical interpretation of  $\Lambda$  and  $\Gamma$  will be given below equation (3.5).

relations hold for  $\Gamma$ . This complex formalism is very convenient for matrix operations; e.g., we obtain for the multiplication of real  $2 \times 2$ -matrices  $A$  and  $B$

$$\begin{aligned}\Lambda[AB] &= \Lambda[A]\Lambda[B] + \Gamma^*[A]\Gamma[B] \\ \Gamma[AB] &= \Gamma[A]\Lambda[B] + \Lambda^*[A]\Gamma[B].\end{aligned}\quad (3.5)$$

To obtain the geometrical interpretation of  $\Lambda$  and  $\Gamma$  we consider the polar decomposition of  $\mathcal{D}$ . If  $\mathcal{D} \neq \mathcal{O}$ , there exist unique numbers  $b_1, b_2$ , with  $0 < b_1 \geq b_2$ , and unique angles  $\phi$  and  $\vartheta$ ,  $0 \leq \phi \leq \pi$ ,  $0 \leq \vartheta \leq 2\pi$ , such that

$$\mathcal{D} = R(\vartheta)B(b_1, b_2, \phi)$$

$R$  is the rotation matrix which has already been defined, and  $B$  is a symmetric matrix  $B = B^T$ :

$$B = R(-\phi) \begin{bmatrix} b_1 & 0 \\ 0 & b_2 \end{bmatrix} R(\phi).$$

In the polar decomposition  $b_1, b_2$  and  $\vartheta$  are the coordinate invariant numbers,  $\phi$  depends on the chosen coordinate system. The matrix  $B$  describes a rotation-free deformation, whereas  $R(\vartheta)$  rotates the plane by an angle  $\vartheta$ . The relation of  $\{\Lambda, \Gamma\}$  to  $\{b_1, b_2, \vartheta, \phi\}$  can be derived with (3.5); we obtain:

$$\begin{aligned}|\Lambda| \pm |\Gamma| &= b_{1,2} & \Lambda &= \frac{1}{2}(b_1 + b_2)e^{i\vartheta} \\ \Gamma &= \frac{1}{2}(b_1 - b_2)e^{i(2\phi - \vartheta)}.\end{aligned}$$

Inserting the values of  $b_1$  and  $b_2$  yields that the 'twist'  $\omega$  is related to the rotation angle  $\vartheta$  of the Jacobi map via

$$\frac{\Lambda}{|\Lambda|} = e^{i\vartheta} \Rightarrow \tan \vartheta = \frac{\omega}{\tau}. \quad (3.6)$$

### 3.1. Consequences of the symmetry of $\mathcal{S}$

(1) After (2.5) we have derived that  $\dot{\mathcal{D}}^T \mathcal{D} = \mathcal{D}^T \dot{\mathcal{D}}$ , or  $\omega[\dot{\mathcal{D}}^T \mathcal{D}] = 0$ , or that  $\mathcal{S}$  is symmetric. Evaluating the twist part of  $\dot{\mathcal{D}}^T \mathcal{D}$  yields:

$$\text{Im} \{ \dot{\Lambda}^* \Lambda + \dot{\Gamma}^* \Gamma \} = 0. \quad (3.7)$$

This constraint equation is valid at every affine parameter, and in particular at the vertex and at every conjugate point. Equation (3.7) illustrates that solving for  $\dot{\mathcal{D}} = \mathcal{T}\mathcal{D}$ , one has not 8 but only 7 free initial conditions. If one chooses the alternative way of solving for the light propagation—evaluating the optical scalars and then solving  $\dot{\mathcal{D}} = \mathcal{S}\mathcal{D}$ —then one has *a priori* only 7 free initial conditions, and the constraint equation (3.7) is hidden in the non-linear differential equations for the optical scalars.

(2) Consider a light beam in an interval  $\lambda \in [\lambda_n, \lambda_{n+1}]$  where  $\mathcal{T}(\lambda) = \mathcal{R}(\lambda)\mathcal{I}$ , i.e. where the source of shear vanishes. Then every component of  $\mathcal{D}$  satisfies the same differential

equation, and the general solution  $\mathcal{D}$  is a linear combination of two linearly independent solutions  $f$  and  $g$  of  $\ddot{x} = \mathcal{R}x$ :

$$\mathcal{D}(\lambda) = f(\lambda)\mathcal{D}_n + g(\lambda)\mathcal{D}_{n+1} \quad (3.8)$$

where  $\mathcal{D}(\lambda_i) =: \mathcal{D}_i$  and  $f_n = 1$ ,  $f_{n+1} = 0$ ,  $g_n = 0$  and  $g_{n+1} = 1$ †. Since  $g$  and  $f$  are linearly independent solutions, we also have  $\dot{g}_n \neq 0$  and  $\dot{f}_{n+1} \neq 0$ . Inserting (3.8) into  $\mathcal{D}^T \mathcal{D} = \mathcal{D}^T \bar{\mathcal{D}}$  and evaluating this matrix at  $\lambda_n$  yields  $\mathcal{D}_{n+1}^T \mathcal{D}_n = \mathcal{D}_n^T \mathcal{D}_{n+1}$ . Hence we have shown: if there is no source of shear between  $\lambda_n$  and  $\lambda_{n+1}$ , the matrix product

$$\mathcal{D}_{n+1}^T \mathcal{D}_n \quad (3.9a)$$

is symmetric. If one matrix (say  $\mathcal{D}_n$ ) is not singular, i.e. there is no point conjugate (to the vertex) at  $\lambda_n$ , then the symmetry of (3.9a) can be expressed as the statement that the matrix

$$\mathcal{D}_{n+1} \mathcal{D}_n^{-1} \quad (3.9b)$$

which carries connection vectors from  $\lambda_n$  to  $\lambda_{n+1}$  is symmetric. This property has been used extensively in the proof of the magnification theorem in gravitational lens theory in paper I.

### 3.2. The Jacobi map near a vertex

At the vertex,  $\Lambda = 0 = \Gamma$ ,  $\dot{\Gamma} = 0 = \text{Im } \dot{\Lambda}$  and  $\text{Re } \dot{\Lambda} = 1$ . In this subsection we do not investigate the behaviour of the optical scalars at the vertex since it is the same as that near a focus; this is due to the fact that locally a beam at a focus differs from that at a vertex only by the opening angle; this angle cancels out in the optical scalars because they are relative quantities. We investigate the Jacobi map in a Taylor expansion as a function of  $\epsilon = \lambda - \lambda_v = \lambda$ ; we put  $\mathcal{T}(0) =: \mathcal{T}_0$  and obtain with (2.4)

$$\mathcal{D}(\epsilon) = \mathcal{I}\epsilon + \frac{1}{6}\mathcal{T}_0\epsilon^3 + \mathcal{O}(\epsilon^4). \quad (3.10)$$

Equation (3.10) implies with the symmetry of  $\mathcal{T}_0$  that the shear of the Jacobi map is at least of third order near the vertex, and the twist increases even more slowly at the vertex. In other words, the cross-section of an initially circular light beam becomes distorted to an ellipse before it can get twisted. To compare the evolution of the shear of the Jacobian with its twist in more detail, we claim: *if the first non-vanishing contribution to  $\Gamma$  is of the order  $\epsilon^n$ ,  $n \geq 3$ , at a vertex, the leading term of  $\omega$  is at least of the order  $\epsilon^{2n}$  (generically,  $n = 3$ ).*

For the proof, we insert the Taylor expansions of  $\Gamma_1$  and  $\Gamma_2$  into the constraint equation (3.7); this yields that the first non-vanishing contribution of this term is of the order  $2n$ . Inserting the Taylor expansions of  $\tau$  and  $\omega$  and using that the first non-vanishing contribution to  $\tau$  is of order one we find that, in order to satisfy the constraint equation at every order of  $\epsilon$ , the leading order of  $\omega$  must be at least  $2n$ .

Therefore the twist  $\omega$  increases at the vertex very slowly compared with the shear; this explains that ‘not too far’ from the observer the light beam can not be twist dominated, i.e.  $\omega^2 < |\Gamma|^2$  holds. This slow increase also holds for the rotation angle  $\vartheta$  of the polar decomposition of the Jacobian matrix near the vertex, since with (3.6)  $\tan \vartheta = \omega/\tau$ . With  $\tau(\epsilon) = \epsilon + \mathcal{O}(\epsilon^3)$  and  $\omega = a\epsilon^6 + \mathcal{O}(\epsilon^7)$ , the rotation angle  $\vartheta = \arctan(\omega/\tau)$  becomes near the vertex  $\vartheta(\epsilon) = a\epsilon^5 + \mathcal{O}(\epsilon^6)$ .

† This is a Sturm boundary value problem: the functions  $f$  and  $g$  exist if and only if the solution with  $x_n = 0$  and  $\dot{x}_n = 1$  satisfies  $x_{n+1} \neq 0$ . This condition is violated if and only if  $\lambda_{n+1}$  and  $\lambda_n$  correspond to a pair of conjugate points.



### 3.3. The light beam near a conjugate point

Non-degenerate conjugate points  $\lambda_c$  are characterized by  $0 \neq |\Gamma(\lambda_c)| = |\Lambda(\lambda_c)|$ . Since  $\Lambda$ , but not  $\Gamma$ , is invariant under rotation of the coordinate system, we can orient the latter such that  $\Gamma(\lambda_c) = \Lambda(\lambda_c)$  at the conjugate point. At a focus  $\Gamma(\lambda_c) = 0 = \Lambda(\lambda_c)$ . In the following we describe the light beam, as before near the vertex, in a Taylor expansion about the conjugate point as a function of  $\epsilon := \lambda - \lambda_c$ . We first derive properties which are common to both kinds of conjugate points; investigating the local behaviour of beams at a conjugate point, we are only interested in solutions of (2.4a) which obey the initial conditions (2.4b).

**Theorem.** At a conjugate point  $x_c$  an eigenvector belonging to the eigenvalue zero of  $\mathcal{D}_c$  cannot be a zero eigenvector to  $\dot{\mathcal{D}}_c$ . In particular, this implies that at a focus the rank of  $\dot{\mathcal{D}}_c$  is two, and at a non-degenerate conjugate point the rank of  $\dot{\mathcal{D}}_c$  is at least equal to one.

**Proof.** Assume that there exists a conjugate point where  $\mathcal{D}_c x = 0$  and  $\dot{\mathcal{D}}_c x = 0$ . Let  $\xi(\lambda) = \mathcal{D}(\lambda)x$ . Then this Jacobi field obeys  $\xi_c = 0$ ,  $\dot{\xi}_c = 0$ , and hence  $\xi \equiv 0$  and also  $\dot{\xi}(0) = 0$ , which is in contradiction to  $\dot{\xi}(0) = \theta \neq 0$ .  $\square$

In order to derive Taylor expansions of  $\det \mathcal{D}$ ,  $\theta$  and  $\sigma$  near conjugate points, we consider the differential equation  $\dot{\mathcal{D}} = \mathcal{T}\mathcal{D}$ . Using (3.5) we rewrite this linear matrix differential equation as system of coupled differential equations for  $\Lambda$  and  $\Gamma$ :

$$\ddot{\Lambda} - \mathcal{R}\Lambda = -\mathcal{F}\Gamma \quad \ddot{\Gamma} - \mathcal{R}\Gamma = -\mathcal{F}^*\Lambda \quad (3.11)$$

which describe two coupled, planar oscillators with the same eigen-frequency and the same absolute coupling strength. Taking the  $n$ th derivative of (3.11), one can iteratively calculate the Taylor expansion coefficients of  $\Gamma$  and  $\Lambda$  in the  $(n+2)$ th order as a function of  $\Lambda_c$ ,  $\Gamma_c$ ,  $\dot{\Lambda}_c$  and  $\dot{\Gamma}_c$  (for the case of a non-degenerate conjugate point) or as a function of  $\dot{\Lambda}_c$  and  $\dot{\Gamma}_c$  (for the case of a focus). A conserved quantity of the differential equation system (3.11) is

$$\mathcal{L} := \dot{\Lambda}^* \Lambda - \Lambda \dot{\Lambda}^* + \Gamma^* \dot{\Gamma} - \dot{\Gamma}^* \Gamma. \quad (3.12)$$

Thus, if  $\mathcal{L}$  vanishes at one value  $\lambda$  it vanishes everywhere, for any  $C^2$  solution  $(\Lambda, \Gamma)$ . ( $\mathcal{R}$  and  $\mathcal{F}$  assumed in  $C^0$ .) Using that the real part of  $\mathcal{L}$  is zero and that  $\Gamma$  and  $\Lambda$  have to satisfy the constraint equation (3.7), yields  $\mathcal{L} = 0$  for a physical solution of (3.11). In terms of  $\Gamma$ ,  $\Lambda$  and their derivatives,  $\sigma$  and  $\theta$  can be written as

$$\theta = \frac{\dot{\Lambda} \Lambda^* - \dot{\Gamma}^* \Gamma}{|\Lambda|^2 - |\Gamma|^2} \quad \sigma = \frac{\dot{\Lambda} \Gamma^* - \dot{\Lambda}^* \Gamma}{|\Lambda|^2 - |\Gamma|^2} \quad (3.13)$$

provided the Jacobi map does not become singular; note that the reality of  $\theta$  is equivalent to  $\mathcal{L} = 0$ . Therefore, one can obtain the series expansions of  $\theta$  and  $\sigma$  by inserting the expansions of  $\Gamma$  and  $\Lambda$  which are derived from (3.11).

**The light beam at a focus.** At a focus,  $\Lambda_c = 0 = \Gamma_c$ ; from our theorem we know that  $|\Gamma_c| \neq |\dot{\Lambda}_c|$  (otherwise the rank of  $\dot{\mathcal{D}}_c$  would be smaller than two). We obtain from (3.11)

$$\Lambda(\epsilon) = \epsilon \dot{\Lambda}_c + \frac{1}{6} \epsilon^3 (\mathcal{R}_c \dot{\Lambda}_c - \mathcal{F}_c \dot{\Gamma}_c) + O(\epsilon^4) \quad (3.14a)$$

$$\Gamma(\epsilon) = \epsilon \dot{\Gamma}_c + \frac{1}{6} \epsilon^3 (\mathcal{R}_c \dot{\Gamma}_c - \mathcal{F}_c^* \dot{\Lambda}_c) + O(\epsilon^4) \quad (3.14b)$$

and thus the determinant of the Jacobian becomes

$$\det \mathcal{D}(\epsilon) = \epsilon^2 \left[ 1 + \frac{1}{3} \epsilon^2 \mathcal{R}_c \right] \det \dot{\mathcal{D}}_c + O(\epsilon^5). \quad (3.15)$$

Since  $\det \dot{\mathcal{D}}_c \neq 0$ , the leading term of  $\det \mathcal{D}(\epsilon)$  is of second order. The optical scalars become near the focus

$$\theta = \frac{1}{\epsilon} \left[ 1 + \frac{1}{3} \mathcal{R}_c \epsilon^2 \right] + \mathcal{O}(\epsilon^2) \quad (3.16a)$$

$$\sigma = \frac{1}{3} \mathcal{F}_c \epsilon + \mathcal{O}(\epsilon^2). \quad (3.16b)$$

The function  $w$  defined in (2.9) is equal to

$$w(\epsilon) = \text{sign}(\det \dot{\mathcal{D}}_c) |\epsilon| \sqrt{|\det \dot{\mathcal{D}}_c|} (1 + \mathcal{O}(\epsilon^2)) \quad (3.17a)$$

thus, it is continuous but not  $\mathcal{C}^1$  at the focus;  $w$  has a finite discontinuity. One obtains the expansions of  $\Gamma$  and  $\Lambda$  near the vertex by inserting the special values  $\dot{\Gamma}_v = 0 = \text{Im } \dot{\Lambda}_v$  and  $\text{Re } \dot{\Lambda}_v = 1$  into (3.14). As expected we obtain for  $w$  at the vertex:

$$w(\epsilon) = \epsilon (1 + \mathcal{O}(\epsilon^2)) \quad \epsilon > 0. \quad (3.17b)$$

As already claimed, the optical scalars (3.16) have the same structure at a vertex and at a focus, since the expansions of the light beam around these points differ only by the opening angle  $\det \dot{\mathcal{D}}_v = 1$  and  $\det \dot{\mathcal{D}}_c$  which cancels in the numerator and denominator of  $\theta$  and  $\sigma$ . Note that in lowest order ( $\epsilon^{-1}$ ) the behaviour of  $\theta$  and  $\sigma$  at the vertex (focus) is expected: the infinitesimal neighbourhood of such an event can be treated asymptotically as the flat Minkowski spacetime. In Minkowski spacetime, however,  $\theta(\lambda) = 1/\lambda$  and  $\sigma(\lambda) \equiv 0$  holds for all  $\lambda$ , in particular at the vertex. The first order terms in  $\theta$  and  $\sigma$  demonstrate that the source of convergence  $\mathcal{R}_0 \leq 0$  at the vertex (or focus) decreases the divergence of a beam, and that the source of shear produces a shear rate  $\sigma$ :

$$\dot{\sigma}(0) = \frac{1}{3} \mathcal{F}_0. \quad (3.18)$$

This implies  $\mathcal{F}_0 = 0 \iff \dot{\sigma}(0) = 0$ , and with (2.2d),  $\mathcal{F} \equiv 0 \iff \sigma \equiv 0$ . Thus, a beam centred on  $\gamma_0$  is shear free if and only if the tangent vector of  $\gamma_0$  is one of the at most 4 principal null directions of the conformal tensor, a rare, exceptional case. Thus generically  $\sigma \neq 0$ .

The fact that  $\sigma = 0$  at the vertex implies that the coefficient of the rhs of the focusing equation (2.8) is continuous at the vertex; thus its solution  $w$  is well defined at least from the observer to the first conjugate point.

*The light beam at a non-degenerate conjugate point.* At a non-degenerate conjugate point, the local expansion of the beam is determined by  $\Gamma_c = \Lambda_c \neq 0$ ,  $\dot{\Lambda}_c$  and  $\dot{\Gamma}_c$ . Since the constraint equation (3.7) has to be satisfied, there are only five free initial conditions: let  $a$  and  $b$  be the unique complex numbers which satisfy  $\dot{\Lambda}_c = a\Lambda_c$  and  $\dot{\Gamma}_c = b\Lambda_c$ ; then (3.7) yields  $\text{Im}[a + b] = 0$ ,  $\Rightarrow \text{Re}[a - b] = a - b^*$ . The zero eigenvector of  $\mathcal{D}_c$  is not a zero eigenvector of  $\dot{\mathcal{D}}_c$  if and only if  $\text{Re}[a - b] \neq 0$ ; therefore  $\text{Re}[a] \neq \text{Re}[b]$ . With (3.11), the expansions of  $\Lambda$  and  $\Gamma$  near the conjugate point can be written as

$$\Lambda(\epsilon) = \left[ 1 + a\epsilon + \frac{1}{2}\epsilon^2 (\mathcal{R}_c - \mathcal{F}_c) \right] \Lambda_c + \mathcal{O}(\epsilon^3) \quad (3.19a)$$

$$\Gamma(\epsilon) = \left[ 1 + b\epsilon + \frac{1}{2}\epsilon^2 (\mathcal{R}_c - \mathcal{F}_c^*) \right] \Lambda_c + \mathcal{O}(\epsilon^3). \quad (3.19b)$$

The determinant of the Jacobian matrix is equal to

$$\det \mathcal{D}(\epsilon) = 2 \text{Re}[a - b] |\Lambda_c|^2 \epsilon + [|a|^2 - |b|^2] |\Lambda_c|^2 \epsilon^2 + \mathcal{O}(\epsilon^3) \quad (3.20)$$

thus, the leading order of this expansion is equal to one. For the optical scalars, we obtain,

$$\theta(\epsilon) = \frac{1}{2\epsilon} \left[ 1 + \frac{|a|^2 - |b|^2}{2 \operatorname{Re}[a - b]} \epsilon + O(\epsilon^2) \right] \quad (3.21a)$$

$$\sigma(\epsilon) = \frac{1}{2\epsilon} \left[ 1 - \frac{|a|^2 - |b|^2}{2 \operatorname{Re}[a - b]} \epsilon + O(\epsilon^2) \right] \quad (3.21b)$$

hence, the rate of shear is real (in the chosen coordinate frame, in which  $\Lambda_c = \Gamma_c$ ) in zeroth order, it becomes imaginary in first order if and only if  $\mathcal{F}_c$  is not real. The function  $w$  is

$$w(\epsilon) = |\Lambda_c| \operatorname{sign}(\operatorname{Re}[a - b] \epsilon) \sqrt{|2 \operatorname{Re}[a - b] \epsilon|} (1 + O(\epsilon^2))$$

or with the abbreviation  $d_c = [(d/d\lambda) \det \mathcal{D}]_{\lambda_c}$ :

$$w(\epsilon) = \operatorname{sign}(d_c \epsilon) \sqrt{|d_c \epsilon|} (1 + O(\epsilon^2)).$$

Thus at a non-degenerate conjugate point  $w$  is continuous, changes its sign, and has an infinite first derivative.

Now we summarize the results for the behaviour of the determinant of the Jacobi map and the optical scalars near conjugate points:

(1) at a non-degenerate conjugate point,  $\det \mathcal{D} \propto \epsilon$ ,  $\theta = 1/2\epsilon$ ,  $\sigma \propto 1/2\epsilon$ ; in leading order;

(2) at a focus,  $\det \mathcal{D} \propto \epsilon^2$ ,  $\theta = 1/\epsilon$ ,  $\sigma = 0$ .

With our knowledge of the behaviour of the shear rate  $\sigma$  at a conjugate point, we can now prove that the focusing equation (2.8b) is integrable over the singularity at a conjugate point: in the worst case, the rhs of (2.8b) behaves like  $|\sigma|^2 \sqrt{\det \mathcal{D}} \propto \epsilon^{-3/2}$ ; this yields  $w(\epsilon) \propto \operatorname{sign}(\epsilon) \sqrt{|\epsilon|}$ . Thus the solution is well defined, even for the case where there is a conjugate point between source and observer. The behaviour of the determinant of the Jacobian map at the two different types of conjugate points also verifies that the sign of  $w$  from (2.8) changes only at a non-degenerate conjugate point, as was claimed in section 2. Our results also show that the points of  $\gamma_0$  conjugate to the vertex form a discrete set.

#### 4. The derivation of the gravitational lens equation from geometrical optics

So far, no approximation has been used. To evaluate the propagation equation (2.4) in an inhomogeneous universe requires several approximation assumptions. These will be stated in this section, and used to rederive the basic relations of the standard gravitational lens theory formalism from general relativity.

*The Friedmann universe.* If one assumes that the universe is isotropic and homogeneous, then its metric is given by the Robertson–Walker metric. The only non-vanishing components of the metric tensor are then  $g_{tt} = c^2$ ,  $g_{ii} = -R^2(t) \tilde{g}_{ii}$ , with  $\tilde{g}_{rr} = 1/(1 - kr^2)$ ,  $\tilde{g}_{\theta\theta} = r^2$  and  $\tilde{g}_{\phi\phi} = r^2 (\sin \theta)^2$ ; the value of  $k = 0, +1, -1$  determines whether the space is flat, spherical or hyperbolic;  $t$  is the cosmic time. A fundamental observer with four velocity  $U^\alpha(\lambda)$  at an event  $\lambda$  on the central ray of a beam measures the frequency  $\omega(\lambda) := ck^\alpha(\lambda) U_\alpha(\lambda) = -\omega_0 \tilde{k}^\alpha(\lambda) U_\alpha(\lambda) =: \omega_0 (1 + z(\lambda))$ ;  $k^\alpha$  is the wavevector of the central ray,  $\omega_0$  is the frequency at the vertex of the beam and  $z(\lambda)$  is (by definition) the (red)shift. In a Robertson–Walker metric, the redshift is isotropic and is related to the scale

factor of the metric by  $R(z) = R_0/(1+z)$ , where  $R_0$  is the scale factor at the vertex of the beam ( $z=0$ ,  $t=t_0$ ). The affine parameter redshift differential equation is

$$\frac{dz}{d\lambda} = \frac{1}{c} \frac{\frac{d}{dt}R(t)}{R(t)} [1+z(\lambda)]^2 = \frac{1}{c} \frac{\frac{d}{dt}R(t)}{R(t_0)} [1+z(\lambda)]^3. \quad (4.1)$$

Note that this yields a proper distance affine parameter relation at redshift  $z$  of

$$dD_{\text{proper}} = (1+z) d\lambda \quad (4.2)$$

which is consistent with our convention that the affine parameter equals the proper length at the vertex at  $\lambda=0=z$ . For a *Friedmann universe* with zero cosmological constant and an energy momentum tensor of a matter-dominated ideal fluid,  $p \ll \rho c^2$ , equation (4.1) can be solved by inserting the Friedmann equation for  $\dot{R}(t)/R(t)$ :

$$\lambda(z) = \frac{c}{H_0} \int_0^z \frac{dz'}{(1+z')^3 \sqrt{\Omega_0 z' + 1}} \quad (4.3)$$

$H_0$  is the Hubble parameter  $d(\ln R)/dt$  at the observation event  $t_0$ .

*Parallel transport in a Robertson-Walker spacetime.* To calculate the source of shear defined in (2.2g), we need the screen vectors  $E_i^\alpha$ ,  $i=1,2$ , and  $\tilde{k}^\alpha$  along the central ray. We choose the centre of the spatial coordinate system  $(r, \theta, \phi)$  at the observer, and the central ray  $\gamma_0$  connecting source and observer in the direction of  $\theta = \pi/2$ . Consider the dimensionless function

$$\Upsilon(t, r) := \int_{t_0}^t \frac{c d\tau}{R(\tau)} + \int_0^r \frac{dx}{\sqrt{1 - kx^2}}.$$

It solves the eikonal equation; the hypersurface  $\Upsilon(t, r) = \Upsilon(t', 0)$  defines the past null cone of  $(t', 0)$ . Therefore, the phase functions converging on the world line  $r=0$  are all given by  $S(t, r) = f(\Upsilon(t, r))$ , where  $f$  depends on the phase  $S(t, 0)$ . The vector  $\tilde{k}_\alpha$  (which is on  $\mathcal{C}_0^-$ ) has to be a constant multiple of  $\Upsilon_{,\alpha} = (R^{-1}(t), 1/\sqrt{1 - kr^2}, 0, 0)^\dagger$ ; since  $\tilde{k}_0 = -1$  at the vertex, we obtain

$$\tilde{k}_\alpha(z) = -(1+z) \left( 1, \frac{R}{\sqrt{1 - kr^2}}, 0, 0 \right)$$

and thus

$$\tilde{k}^\alpha(z) = (1+z) \left[ -1, \frac{1}{\sqrt{-g_{rr}(z)}}, 0, 0 \right]. \quad (4.4)$$

The spacelike screen vectors  $E_1^\alpha$  and  $E_2^\alpha$  adapted to  $\tilde{k}^\alpha$  can be chosen at the observer proportional to  $[0, 0, 1, 0]$  and  $[0, 0, 0, 1]$ . For general  $z$  we then obtain

$$E_1^\alpha(z) = \frac{1}{\sqrt{-g_{\theta\theta}(z)}} [0, 0, 1, 0] \quad E_2^\alpha(z) = \frac{1}{\sqrt{-g_{\phi\phi}(z)}} [0, 0, 0, 1]. \quad (4.5)$$

$\dagger$  The components of a 4-vector  $x^\alpha$  are  $x^0 = ct, r, \theta, \phi$ .

The components of the vectors  $E_i^\alpha(0)$  become singular at the observer at  $z = 0$ . This is due to the choice of the coordinate system; the vectors themselves and their inner products are regular.

*The on-average Friedmann universe.* Of course, a homogeneous universe is not realistic. A better model must take into account that only a fraction  $0 \leq \tilde{\alpha} \leq 1$  of the matter is distributed homogeneously, whereas the rest is concentrated in clumps. Imagine a model universe that is inhomogeneous on small scales and homogeneous on large scales (some 100 Mpc) such that this clumping of matter does not affect 'global' (or large scale) functions like  $R(t)$ ,  $R(z)$ ,  $\lambda(z)$  and the parallelly transported fields  $\tilde{k}^\alpha(z)$ ,  $E_{(i)}^\alpha(z)$ . This means that, on average, this universe behaves like a Friedmann universe with density  $\rho_F$  which has the same total matter content as the actual universe. Thus, such a model is called an *on-average Friedmann universe* (see, e.g., Zeldovich 1964, Dyer and Roeder 1973).

This picture of the matter distribution in our universe is a realistic one if one is interested in the light deflection caused by 'strong', isolated matter inhomogeneities, such as galaxies and clusters of galaxies, the deflectors which produce multiply-imaged QSOs, radio rings, and luminous arcs. In these situations it seems to be a fair approximation to consider the light beams between us and the deflector, and between the deflector and the source to be nearly unperturbed by matter inhomogeneities; if there is more than one deflector along the line of sight, this can be accounted for in the present prescription. An alternative view of the matter distribution in the universe is provided by considering larger scales, at which the density inhomogeneities are linear or quasi-linear. Then it is more realistic to model the matter distribution as a field  $\delta\rho$  which is superposed on the Friedmann density  $\rho_F$ , such that  $\langle \delta\rho \rangle = 0$ , and the average is taken at spatial scales which are small compared with the Hubble length, but larger than the largest scale at which the density fluctuations  $\delta\rho$  still have appreciable power (see, e.g., Gunn 1967, Blandford *et al* 1991, Kaiser 1992 for studies of light propagation in such a weakly inhomogeneous universe). In the following we adopt the first view, that of a clumpy universe; we note, however, that most of our results derived below also apply for the weakly inhomogeneous universe. In particular, the (multiple deflection) gravitational lens equation can also be used in the latter case, if the universe is 'sliced' into redshift bins and the matter inhomogeneities are projected onto 'lens planes' in the bins, since the multiple deflection gravitational lens equation can be considered just as a discretization of the exact propagation equation (2.4). The only modification that has to be applied in the case of a weakly inhomogeneous universe is that  $\mathcal{R}_{cl}$  no longer is non-positive, and the projected surface mass density  $\Sigma$  in each lens plane can attain positive and negative values. Furthermore, since the magnification, defined in section 5 below, is defined relative to the Friedmann-Lemaître universe, the mean magnification relative to that must be unity (see the discussion in section 4.5.1 of SEF), and the focusing theorem of section (5.3) no longer holds, since  $\mathcal{R}_{cl}$  can have either sign.

#### 4.1. The sources of shear and convergence for weak, isolated inhomogeneities

*Weak, isolated inhomogeneities.* We assume that inhomogeneities like galaxies or clusters of galaxies are isolated from each other such that in each domain containing an inhomogeneity, small compared with the Hubble distance, the metric can be approximated by a post-Minkowskian line element

$$ds^2 = \left(1 + 2\frac{\Phi}{c^2}\right)c^2 dt^2 - \left(1 - 2\frac{\Phi}{c^2}\right)d\mathbf{x}^2.$$

The relative velocities of its mass distribution are small,  $v \ll c$ , and its Newtonian gravitational potential  $\Phi$  is weak,  $\Phi \ll c^2$ . If the density outside such regions is  $\tilde{\alpha}\rho_F$

and we write for the density inside a clump  $\tilde{\alpha}\rho_F + \rho_{cl}$ , such that  $\rho_{cl}$  is localized in the region, Poisson's equation  $\Delta_3\Phi = 4\pi G\rho_{cl}$  holds within the region†. The metric does not change appreciably on the time scale light needs to propagate through the inhomogeneity. We therefore call such inhomogeneities *quasi-static, weak inhomogeneities*.

*The source of convergence.* First we consider the source of convergence  $\mathcal{R}$ , defined in (2.2f). Inserting the field equations with an energy-momentum tensor of an ideal fluid yields:

$$\mathcal{R} = -\frac{4\pi G}{c^2} \tilde{\rho} \tilde{U}_\alpha \tilde{U}_\beta \tilde{K}^\alpha \tilde{K}^\beta. \quad (4.6)$$

In this equation,  $\tilde{U}^\alpha$  is the 4-velocity of the ideal fluid, which deviates from the velocity in a pure Friedmann universe  $U^\alpha$  by the peculiar velocity  $U_{pec}^\alpha$ ,  $\tilde{U}^\alpha = U^\alpha + U_{pec}^\alpha$ , and  $\tilde{K}^\alpha$  is the wave vector of the central ray of the beam considered, which deviates from the wavevector  $\tilde{k}^\alpha$  in a Friedmann universe due to deflection within the inhomogeneity by a vector  $\delta\tilde{k}^\alpha$ ,  $\tilde{K}^\alpha = \tilde{k}^\alpha + \delta\tilde{k}^\alpha$ . The matter density  $\tilde{\rho} = \rho_{bg} + \rho_{cl}$  is given as a sum of the reduced background density in the on-average Friedmann universe,  $\rho_{bg} = \tilde{\alpha}\rho_F$ , and the matter density of the clump  $\rho_{cl}$ . If we use that peculiar velocities of inhomogeneities (e.g. galaxies) are small,  $v_{pec} \lesssim 10^{-3}c$ , and that their gravitational fields are also small,  $\delta z_g \approx 2\Phi/c^2 \ll 1$ , we can neglect the contributions from  $U_{pec}^\alpha$  and  $\delta\tilde{k}^\alpha$  and obtain from (4.6) that in lowest order, with  $\mathcal{R} = \mathcal{R}_{bg} + \mathcal{R}_{cl}$ , the contribution of the clumps is given by

$$\mathcal{R}_{cl} \approx -\frac{4\pi G}{c^2} \rho_{cl} U_\alpha U_\beta \tilde{k}^\alpha \tilde{k}^\beta. \quad (4.7)$$

Consider an inhomogeneity along a ray  $\gamma_0$  localized in the affine parameter interval  $[\lambda_{min}, \lambda_{min} + \Delta\lambda]$  which is small compared with its distance to us:  $\Delta\lambda \ll \lambda_{min}$ ; let  $z_d$  be an element of the corresponding redshift interval  $[z_{min}, z_{min} + \Delta z]$ . Since the inhomogeneity must not change significantly during the time the light beam traverses it, we can calculate (4.7) for one instant of time,  $t(z_d)$ . The line element in the asymptotically flat neighbourhood  $\mathcal{U}$  of  $\lambda(z_d)$  is  $ds^2 = (1 + 2\Phi/c^2)(c dt)^2 - (1 - 2\Phi/c^2)(d\zeta^2)$ , with  $t = R_d\eta$ ,  $R_d^2 d\sigma_k^2 \approx (d\zeta)^2$  and  $R(z_d) = R_d$ ;  $(t, \zeta)$  denote post-Minkowski coordinates centred on  $\lambda(z_d)$  and oriented such that  $\zeta_3$  is parallel to the spatial direction of  $\gamma_0$  there. We calculate  $\mathcal{R}$  and  $\mathcal{F}$  not only on the central ray  $\gamma_0$  of the beam considered, but for all spatial positions  $\zeta$  in  $\mathcal{U}$ . This yields  $\mathcal{R}$  and  $\mathcal{F}$  for all rays traversing  $\mathcal{U}$ , where the spatial paths of the rays are parametrized by  $\zeta(\lambda)$ ; note that the rays in  $\mathcal{U}$  do not have to be infinitesimally near to  $\gamma_0$  in the sense of (2.3). The source of convergence on a ray in  $\mathcal{U}$  is the sum of

$$\mathcal{R}_{bg}(\lambda) = -\frac{4\pi G}{c^2} \rho_{bg}(z) [1+z]^2 \quad \mathcal{R}_{cl}(\lambda) = -\frac{4\pi G}{c^2} \rho_{cl}(\zeta(\lambda)) [1+z]^2 \quad (4.8a)$$

where we have written  $z$  instead of  $z(\lambda)$ . If one uses the Poisson equation  $\Delta_3\Phi = 4\pi G\rho_{cl}$  for the quasi-stationary Newtonian gravitational potential  $\Phi(t_d, \zeta) \approx \Phi(\zeta)$  of the inhomogeneity, this yields for  $\mathcal{R}_{cl}$ :

$$\mathcal{R}_{cl}(\lambda) = -\frac{(1+z)^2}{c^2} \Delta_3\Phi(\zeta(\lambda)). \quad (4.8b)$$

† Concerning the difficult problem of constructing approximate solutions of Einstein's equations containing quasi-static, weak inhomogeneities separated by 'empty regions' and being Friedmannian on a large scale, see Futamase and Sasaki (1989), Jacobs et al (1993); see also Kasai (1993).

Up to now we have considered weak inhomogeneities which are small in size compared with their distance from us. Now we will restrict ourselves to those which are sufficiently thin, such that one can replace the wavevector and the vectors  $E^\alpha$  in (2.2g) by (4.4) and (4.5) evaluated at the redshift of the clump. (That is, for the calculation of the source term for the evolution of the light beam, one can neglect the deflection relative to the unperturbed light beam). Thus we approximate

$$\zeta(\lambda) \approx (\zeta_1(\lambda_d), \zeta_2(\lambda_d), \zeta_3(\lambda)) \quad (4.9)$$

for rays which are roughly parallel to  $\gamma_0$  at  $\lambda(z_d)$ ; the deviation of rays from the parallel direction must be small, as well as the typical deflection angle caused by an inhomogeneity, otherwise the approximation (4.9) would break down†. With our choice of the coordinate system,  $\zeta_1$  and  $\zeta_2$  are orthogonal coordinates on the screen defined in section 2. Therefore we write  $(\zeta_1(\lambda_d), \zeta_2(\lambda_d)) = \xi$  in the following;  $\xi$  is a parameter for labelling rays. Equations (4.8) hold for an infinitesimal beam with central ray  $\gamma_0$  ( $\xi = 0$ ), and for any other ray which is in  $\mathcal{U}$  and roughly parallel to  $\gamma_0$  at  $\lambda_d$ .

The approximation (4.9) is equivalent to one on which gravitational lens theory is based: there, the source term for the light bending along the deflected light ray is approximated locally by that evaluated along the path of the unperturbed ray.

*The source of shear.* Outside the matter inhomogeneities, where  $\rho_{bg} = \bar{\alpha}\rho_F$ , we neglect the source of shear due to clumps; i.e. we neglect the long range gravitational action of the weak inhomogeneities, and put  $\mathcal{F} = 0$ . At the inhomogeneity we evaluate (2.2g) in post-Minkowskian coordinates, hence we have to transform the coordinates from  $(x^0, r, \theta, \phi)$  to  $(x^0, \zeta_1, \zeta_2, \zeta_3)$ . Note that we have chosen the  $\zeta_3$ -direction of the new coordinate system parallel to the spatial direction of the central ray. Since the normalization of all vectors stays invariant under the transformation of the coordinate system and since the norm in the local Minkowski system is built with  $\eta = \text{diag}(1, -1, -1, -1)$ , we have to replace the metric tensor  $g$  by  $\eta$  in (4.4) and (4.5) and we obtain

$$\tilde{k}^\alpha(z_d) = -(1 + z_d)[1, 0, 0, -1] \quad E_1^\alpha(z_d) = [0, 1, 0, 0] \quad E_2^\alpha(z_d) = [0, 0, 1, 0]. \quad (4.10)$$

The Riemann tensor in the post-Minkowskian approximation in  $ct$  and  $\zeta$  coordinates is equal to

$$R_{\alpha\beta\gamma\delta} = -\frac{1}{c^2} \{ \delta_{\alpha\gamma} \Phi_{,\beta\delta} - \delta_{\beta\gamma} \Phi_{,\alpha\delta} - \delta_{\alpha\delta} \Phi_{,\beta\gamma} + \delta_{\beta\delta} \Phi_{,\alpha\gamma} \}. \quad (4.11)$$

Thus, (4.10) and (4.11) yield that there are only contributions to the source of shear in (2.2g) if  $\alpha, \gamma \in \{1, 2\}$  and  $\beta, \delta \in \{0, 3\}$ ; hence the summation contains only 16 non-vanishing contributions. Using the quasi-stationarity of the metric,  $\Phi_{,0} \ll \Phi_{,i}$ , yields that in lowest order of  $v/c$ , only the following eight components of the curvature tensor contribute to (2.2g):

$$R_{1010} = -\frac{1}{c^2} \Phi_{,11} \quad R_{1020} = R_{2010} = -\frac{1}{c^2} \Phi_{,12} \quad R_{1313} = -\frac{1}{c^2} (\Phi_{,33} + \Phi_{,11}) \quad (4.12a)$$

$$R_{2020} = -\frac{1}{c^2} \Phi_{,22} \quad R_{1323} = R_{2313} = -\frac{1}{c^2} \Phi_{,12} \quad R_{2323} = -\frac{1}{c^2} (\Phi_{,33} + \Phi_{,22}). \quad (4.12b)$$

† In astrophysically relevant situations the beams under consideration have an opening angle of  $\sim 1$  arcminute  $\sim 3 \cdot 10^{-4}$  for galaxy clusters, and of  $\sim 10$  arcseconds  $\sim 5 \times 10^{-5}$  for lensing by individual galaxies; the corresponding typical deflection angles are of the same order or smaller.

Inserting  $\Phi_{,12} = \Phi_{,21}$ , (4.12) and (4.10) in (2.2g) and using (4.9) yields

$$\mathcal{F}_{\text{cl}}(\xi; \lambda) = \frac{1}{c^2} (1+z)^2 \{ \Phi_{,11} - \Phi_{,22} - 2i\Phi_{,21} \} (\xi; \xi_3(\lambda)). \quad (4.13)$$

Therefore we obtain with (4.13), (4.8a) and (4.8b) that the optical tidal matrix along a family of rays traversing an asymptotically flat neighbourhood of an event  $\lambda_d$  localized in a weak geometrically thin clump in an on-average Friedmann universe, such that their spatial directions are roughly parallel to the  $\xi_3$ -direction at  $\lambda_d$ , is  $\mathcal{T}(\xi; \lambda) = \mathcal{T}_{\text{bg}}(z) + \mathcal{T}_{\text{cl}}(\lambda)$  with  $\mathcal{T}_{\text{bg}}(z) = \mathcal{R}_{\text{bg}}(z)\mathcal{I}$  and

$$(\mathcal{T}_{\text{cl}})_{ik}(\xi; \lambda) = -\frac{(1+z)^2}{c^2} [2(\Phi_{,ik}) + (\delta_{ik}\Phi_{,33})] (\xi; \xi_3(\lambda)) \quad i, k \in \{1, 2\}. \quad (4.14)$$

Thus, the optical tidal matrix is simply related to the ordinary tidal matrix, i.e., the matrix of the second derivatives of the Newtonian potential. In these equations,  $z = z(\lambda)$ , and  $\xi$  is the screen position of the ray considered at  $\lambda_d$  relative to one chosen ray  $\gamma_0$  of the family;  $\xi_3$  is the direction in the post-Minkowski coordinate system parallel to the rays at  $\lambda_d$ , hence with (4.2)

$$d\xi_3 = (1+z)d\lambda. \quad (4.15)$$

If one evaluates the mapping of an infinitesimally thin beam (i.e. one needs the value of (4.14) on one ray  $\gamma_0$  only), one puts  $\xi = 0$  in (4.14).

#### 4.2. The thin lens approximation

One of the simplifying assumptions underlying lens theory is that the inhomogeneities are geometrically thin. Thus one approximates the inhomogeneities by two-dimensional *surface mass densities*  $\Sigma$ . Let one of the distributions be situated on the 'plane'  $\xi_3 = 0$ ,

$$\rho_{\text{cl}}(\xi, \xi_3) \approx \delta(\xi_3)\Sigma(\xi) \quad (4.16)$$

where

$$\Sigma(\xi) := \int_{-\infty}^{+\infty} d\xi_3 \rho_{\text{cl}}(\xi, \xi_3). \quad (4.17)$$

The Newtonian potential of this distribution is

$$\Phi(\xi, \xi_3) = -G \int \frac{\Sigma(\eta)d^2\eta}{\sqrt{(\xi - \eta)^2 + \xi_3^2}}. \quad (4.18)$$

The derivatives  $\Phi_{,ik}$ ,  $\Phi_{,33}$  which occur in the tidal matrix (4.14) decrease as the inverse third power of the distance from the plane of the mass distribution. It is, therefore, reasonable to approximate the optical tidal matrix for a clump of matter as a delta-distributional source term in  $\lambda$ :

$$\mathcal{T}_{\text{cl}}^p(\xi; \lambda) := \delta(\lambda - \lambda_d) \int_{-\infty}^{+\infty} \mathcal{T}_{\text{cl}}(\xi; \xi_3(\lambda')) d\lambda'. \quad (4.19)$$



The deflection potential  $\tilde{\Psi}$ . The deflection potential  $\tilde{\Psi}(\xi)$  of an inhomogeneity is defined as usual by

$$\tilde{\Psi}(\xi) = \frac{4G}{c^2} \int d^2\xi' \Sigma(\xi') \ln \left\{ \frac{|\xi - \xi'|}{D_d} \right\} \quad (4.20)$$

(see SEF, sections 4.3 and 5.1). In the deflection potential the denominator in the argument of the logarithm is an arbitrary length, to make this argument dimensionless; we have chosen it equal to the so called empty cone angular diameter distance  $D_d := D(z_d)$  from the observer to the redshift  $z_d$ . Under a change of this length scale, the value of (4.20) changes only by an unimportant additive constant. It is straightforward to see that  $\tilde{\Psi}$  and  $\Sigma$  are related to each other by the Poisson equation for the surface mass density

$$\Delta_{(2)} \tilde{\Psi}(\xi) = \frac{8\pi G}{c^2} \Sigma(\xi) \quad (4.21)$$

where  $\Delta_2$  is the two-dimensional Laplace operator.

We now show that the approximate tidal matrix of equation (4.19) can be expressed in terms of the second derivatives of the deflection potential rather than in terms of the  $\Phi$  derivatives. In fact, using equations (4.18), (4.20) and (4.15) one verifies by a straightforward calculation that, for  $i, k \in \{1, 2\}$ ,

$$\int_{-\infty}^{\infty} d\zeta_3 \Phi_{,ik}(\xi, \zeta_3) = \frac{1}{2} c^2 \tilde{\Psi}_{,ik}(\xi) \quad (4.22)$$

$$\int_{-\infty}^{\infty} d\zeta_3 \Phi_{,33}(\xi, \zeta_3) = 0. \quad (4.23)$$

Therefore equation (4.18) leads to

$$\begin{aligned} T_{cl}^p(\xi; \lambda) &= -(1 + z_d) \delta(\lambda - \lambda_d) \begin{bmatrix} \tilde{\Psi}_{,11}(\xi) & \tilde{\Psi}_{,12}(\xi) \\ \tilde{\Psi}_{,21}(\xi) & \tilde{\Psi}_{,22}(\xi) \end{bmatrix} \\ &= -(1 + z_d) \delta(\lambda - \lambda_d) \tilde{U}(\xi). \end{aligned} \quad (4.24)$$

In the last step, we have defined the *deflection matrix*  $\tilde{U}(\xi)$  as the Hessian of the deflection potential  $\tilde{U}(\xi) =: \mathcal{H}[\tilde{\Psi}](\xi)$ .

We can generalize the result (4.24) to the case of several inhomogeneities, i.e. for the following case, which also is the 'standard situation' in gravitational lens theory: given an observer at redshift zero in an on-average Friedmann universe, a source at redshift  $z_s =: z_{N+1}$  and an arbitrary number  $N$  of geometrically thin, weak inhomogeneities between source and observer, situated at  $\lambda_1, \dots, \lambda_N$  with corresponding redshifts of  $z_1, \dots, z_N$ . Then, if we indicate the two-dimensional screen positions of a ray (relative to one ray  $\gamma_0$  of the family) in the inhomogeneities with  $\xi_j$  and the deflection matrices at those positions as  $\tilde{U}_j(\xi_j)$ , the optical tidal matrix is equal to

$$T^p(\xi_1, \dots, \xi_N; \lambda) = \mathcal{R}_{bg}(\lambda) \mathcal{I} - \sum_{i=1}^N (1 + z_i) \tilde{U}_i(\xi_i) \delta(\lambda - \lambda_i) \quad (4.25a)$$

the different rays considered must be roughly parallel to each other before the first inhomogeneity, then, the same holds at every following inhomogeneity provided the deflection angles are small. Again, considering only one infinitesimal beam with central ray  $\gamma_0$ , one has to consider

$$T^p(\lambda) = \mathcal{R}_{bg}(\lambda) \mathcal{I} - \sum_{i=1}^N (1 + z_i) \tilde{U}_i(0) \delta(\lambda - \lambda_i). \quad (4.25b)$$

### 4.3. The recurrence relation for the mapping of the light beam

The equations (4.25) result from well defined assumptions and approximations. Hence we can solve the differential equation (2.4a) with (4.25) as source term. We again consider not only a single beam, but a family of beams with (nearly) parallel central rays, and label a beam by the screen position  $\xi_n$  of its central ray relative to one reference ray  $\gamma_0$ . Defining  $\dot{D}_n^+(\xi_n) := \lim_{\lambda \searrow \lambda_n} \dot{D}(\xi_n; \lambda)$  and  $\dot{D}_n^-(\xi_n) := \lim_{\lambda \nearrow \lambda_n} \dot{D}(\xi_n; \lambda)$ , this yields:

$$\dot{D}_n^+(\xi_n) - \dot{D}_n^-(\xi_n) = -(1 + z_n) \ddot{U}_n(\xi_n) D_n(\xi_n) \quad (4.26)$$

thus the Jacobi matrix, but not its derivative, is continuous at an inhomogeneity in the lens approximation. On the lhs of (4.26) we want to express the derivatives of the Jacobi matrices as functions of the values of the Jacobi matrices at redshifts  $z_{n-1}$ ,  $z_n$  and  $z_{n+1}$ . In order to do this, we first have to determine the evolution of an infinitesimal light beam outside clumps.

*The evolution of a beam outside clumps, Dyer–Roeder differential equation.* We now investigate the evolution of a beam outside clumps, which we call empty beam or *empty cone* in the following. Since outside of clumps the source of shear vanishes, the differential equation (2.4a) simplifies with the first of (4.8a) to

$$\ddot{D}(\lambda) = \mathcal{R}_{bg}(\lambda) D(\lambda) = -\frac{4\pi G}{c^2} \rho_{bg}(z) [1 + z]^2 D(\lambda).$$

If we insert the evolution of the density with redshift,  $\rho_{bg}(z) = \tilde{\alpha} \rho_0 (1 + z)^3$ , the definition of the density parameter  $\Omega = \rho_0 / \rho_{crit}$  with  $\rho_{crit} = 3H_0^2 / 8\pi G$ , we find that each component of  $D$  satisfies the differential equation

$$\frac{d^2}{d\lambda^2} B(\lambda) = -\frac{3}{2} \left( \frac{c}{H_0} \right)^{-2} \tilde{\alpha} \Omega [1 + z(\lambda)]^5 B(\lambda).$$

Using the affine parameter redshift relation (4.3), this finally transforms to the *Dyer–Roeder differential equation* (Dyer and Roeder 1973)

$$(\Omega z + 1)(1 + z) \frac{d^2 B(z)}{dz^2} + \left[ \frac{7}{2} \Omega z + \frac{1}{2} \Omega + 3 \right] \frac{dB(z)}{dz} + \frac{3}{2} \tilde{\alpha} \Omega B(z) = 0. \quad (4.27)$$

This second order differential equation has two linearly independent solutions; two solutions  $B_1$  and  $B_2$  are independent if and only if the Wronskian  $W(z) := \dot{B}_1 B_2 - \dot{B}_2 B_1(z)$  is different from zero at one value of  $z$  (and thus for every  $z$ ). The first and second terms of equation (4.27) describe the evolution of a light beam due to the expansion of the universe, therefore  $\Omega$  appears; the third term describes the convergence of a light beam due to the local homogeneous matter density  $\tilde{\alpha} \rho_F$  in the empty cone (no clumps); for this reason, a term  $\tilde{\alpha} \Omega$  occurs. Consider a solution  $D(z_1, z)$  of (4.27) which is zero at redshift  $z_1$  and whose derivative with respect to redshift obeys the local Hubble law, or equivalently, the infinitesimal quantity  $(dD/dz) dz$  equals the infinitesimal proper length  $dD_{proper}(z_1)$  at redshift  $z_1$ . Then  $D(z_1, z_2)$  is the empty cone angular diameter distance from redshift  $z_1$  to  $z_2$ ; it can be described by a function  $r(z_1, z)$ , solving (4.27) with boundary conditions

$$\frac{d}{dz} r(z_1, z)|_{z=z_1} = \frac{1}{(1 + z_1)^2 \sqrt{\Omega z_1 + 1}} \quad r(z_1, z)|_{z=z_1} = 0 \quad (4.28)$$

in the following form:

$$D(z_1, z_2) = \frac{c}{H_0} |r(z_1, z_2)|. \quad (4.29)$$

The general solution of this initial value problem is provided in Seitz and Schneider (1994). If there is no inhomogeneity in the beam between its vertex and redshift  $z$ , the Jacobi matrix  $\mathcal{D}(z)$  is given by  $\mathcal{D}(z) = (c/H_0)r(0, z)\mathcal{I}$ ; in particular, at the first inhomogeneity at  $z_1$ ,  $\mathcal{D}(z_1) = D(0, z_1)\mathcal{I}$ . To describe the solution of equation (2.4a) between the  $(n-1)$ th and  $n$ th and between the  $n$ th and  $(n+1)$ th inhomogeneity, we put:

$$\mathcal{D}(z) = X_1 B_1(z) + Y B_2(z) \quad z \in [z_{n-1}, z_n] \quad (4.30)$$

$$\mathcal{D}(z) = X_2 B_1(z) + Z B_2(z) \quad z \in [z_n, z_{n+1}]. \quad (4.31)$$

Here,  $B_1$  and  $B_2$  are linearly independent solutions of the Dyer–Roeder differential equation; we choose them as  $B_1(z) := D(0, z) =: D(z)$  and  $B_2(z) = D(z_n, z)$ .  $X_1$ ,  $X_2$ ,  $Y$  and  $Z$  are real  $2 \times 2$  matrices, determined by the boundary conditions. Evaluating (4.30) and (4.31) at  $z_n$  immediately yields  $X := X_1 = X_2 = (1/D_n)\mathcal{D}_n =: A_n$ . Then, we calculate the derivatives of (4.30) and (4.31) with respect to  $\lambda$ , evaluate these at  $\lambda_n$  and obtain with (4.3) and (4.28) the difference:

$$\dot{\mathcal{D}}_n^+ - \dot{\mathcal{D}}_n^- = Z \frac{d}{d\lambda} D(z_n, z)|_{z=z_n} - Y \frac{d}{d\lambda} D(z_n, z)|_{z=z_n} = (1+z_n)[Z+Y]. \quad (4.32)$$

The matrices  $Y$  and  $Z$  can be calculated by evaluating (4.30) and (4.31) at  $z_{n-1}$  and  $z_{n+1}$ , respectively. With the abbreviations  $D(z_i, z_j) =: D_{ij}$  and  $D(z_i) =: D_i$  this yields:

$$Y = \frac{1}{D_{n,n-1}} \left\{ \mathcal{D}_{n-1} - \frac{D_{n-1}}{D_n} \mathcal{D}_n \right\} \quad Z = \frac{1}{D_{n,n+1}} \left\{ \mathcal{D}_{n+1} - \frac{D_{n+1}}{D_n} \mathcal{D}_n \right\}. \quad (4.33)$$

We insert (4.33) and (4.32) into (4.26), use the Etherington (1933) reciprocity relation

$$\frac{D(z_1, z)|_{z=z_2}}{1+z_1} = \frac{D(z_2, z)|_{z=z_1}}{1+z_2}$$

and obtain for  $\mathcal{D}_{n+1}$ :

$$\begin{aligned} \mathcal{D}_{n+1} = & -D_{n,n+1} \tilde{U}_n \mathcal{D}_n - \frac{1+z_{n-1}}{1+z_n} \frac{D_{n,n+1}}{D_{n-1,n}} \mathcal{D}_{n-1} \\ & + \left\{ \frac{(1+z_{n-1})}{1+z_n} \frac{D_{n,n+1} D_{n-1}}{D_{n-1,n} D_n} + \frac{D_{n+1}}{D_n} \right\} \mathcal{D}_n. \end{aligned} \quad (4.34)$$

However, this relation is equivalent to the recurrence relation for the Jacobi matrices in lens theory. This becomes clear if one rewrites this equation, as is common in lens theory, in dimensionless form. One has to insert the dimensionless deflection matrix  $U(x)$  related to  $\tilde{U}(\xi)$  via

$$U_j(x_j) = \frac{D_{j,N+1} D_j}{D_{N+1}} \tilde{U}_j(\xi_j) \quad \xi_j =: x_j D_j$$

and the definitions of the dimensionless Jacobi matrices  $A_i(x_i) := (1/D_i)\mathcal{D}_i(D_i x_i)$ . Defining the geometrical quantities

$$\vartheta_i := \frac{(1+z_i)}{c} \frac{D_i D_{i+1}}{D_{i,i+1}} \quad 0 \leq i \leq N \quad v_i := -\sqrt{\frac{\vartheta_{i-1}}{\vartheta_i}} \quad 1 \leq i \leq N$$

and

$$\beta_{ij} := \frac{D_{ij} D_{N+1}}{D_j D_{i,N+1}} \quad 1 \leq i < j \leq N+1$$

as in paper I, this yields

$$A_{n+1} = -\beta_{n,n+1} U_n A_n - v_n^2 A_{n-1} + (1+v_n^2) A_n = T_n A_n - v_n^2 A_{n-1} \quad (4.35)$$

where the  $2 \times 2$  matrices  $T_n$  are defined as  $T_n := (1+v_n^2)\mathcal{I} - \beta_{n,n+1} U_n$ ,  $1 \leq n \leq N$  and the starting condition is  $A_1 = \mathcal{I}$ . This is the same recurrence relation as that in gravitational lens theory, see e.g. equation (2.21) of paper I. Hence we have shown that the recurrence relation for the mapping of the Jacobi matrices in lens theory can be derived as a direct approximation from geometrical optics.

#### 4.4. The deflection angle, the lens equation

We have seen that light propagation for infinitesimal light beams can be derived from geometrical optics. Can one also derive the lens equation and the deflection angle from geometrical optics? Yes, provided that the matter outside the clumps is homogeneous and the source of shear due to the clumps is assumed to vanish outside of the clumps. Therefore, the mapping between two consecutive lens planes can be considered to be linear on a large scale, i.e. not just for infinitesimal beams, but also for 'fat beams' (which of course have to be smaller than the typical separation between clumps).

Consider two rays  $\gamma_0$  and  $\gamma_1$  including an angle  $\theta$  at their intersection point at the observer, where this angle is small enough to ensure that these rays are approximately parallel, but not necessarily infinitesimally small.

We treat one of them ( $\gamma_0$ ) as a reference ray, adapt a screen to it (as defined in section 2) and denote the screen position of  $\gamma_1$  at redshift  $z$  by  $\xi^I(z)$ . We calculate the evolution of this separation vector from the observer ( $z=0$ ) to the source at  $z_s = z_{N+1}$  in two steps:

(1) Due to the remark above, the separation vector has to satisfy the Jacobi deviation equation (2.2a) with the source term  $\mathcal{T} = \mathcal{R}_{bg}\mathcal{I}$  outside inhomogeneities. Hence, each component of this separation vector has to satisfy the Dyer-Roeder differential equation (4.27). Thus, if we indicate the screen position of  $\gamma_1$  (relative to  $\gamma_0$ ) at the  $j$ th inhomogeneity by  $\xi_j^I$  we can describe this separation vector between the  $(n-1)$ th and the  $n$ th lens plane by

$$\xi^I(z) = \frac{r(z_n, z)}{r(z_n, z_{n-1})} \xi_{n-1}^I + \frac{r(z_{n-1}, z)}{r(z_{n-1}, z_n)} \xi_n^I \quad z_{n-1} \leq z \leq z_n. \quad (4.36)$$

Note, that  $r(z_n, z)$  and  $r(z_{n-1}, z)$  form a pair of linearly independent solutions of the Dyer-Roeder equation and that inserting  $z_n$  and  $z_{n-1}$  yields the correct boundary conditions.

(2) If there was no inhomogeneity at redshift  $z_n$ , (4.36) would stay valid also for  $z \geq z_n$ . But since there is an inhomogeneity, we have to correct for this and we have to take into

account that for  $z > z_n$  the optical tidal matrix again becomes  $T = \mathcal{R}_{bg} \mathcal{I}$ . The correction function has to be a solution  $B(z)$  of the Dyer–Roeder equation. Thus we obtain

$$\xi^I(z) = \frac{r(z_n, z)}{r(z_n, z_{n-1})} \xi_{n-1}^I + \frac{r(z_{n-1}, z)}{r(z_{n-1}, z_n)} \xi_n^I - B(z) c_n(\xi_n^I) \quad z_n \leq z \leq z_{n+1}. \quad (4.37)$$

$c_n$  is a non-zero vector quantity, therefore  $B$  must vanish at  $z_n$ . We can choose the derivative of  $B$  at  $z_n$  such that

$$\left. \frac{dB(\lambda)}{d\lambda} \right|_{\lambda=z_n} = (1 + z_n) \quad (4.38)$$

holds, and thus  $B(z) = D(z_n, z)$ .

*The deflection angle.* We define the derivatives of the separation vector of the two rays with respect to the affine parameter, before and after the  $n$ th inhomogeneity:

$$\xi_{n+}^I := \lim_{\Delta\lambda \searrow 0} \frac{d}{d\lambda} \xi^I(\lambda_n + \Delta\lambda) \quad \xi_{n-}^I := \lim_{\Delta\lambda \searrow 0} \frac{d}{d\lambda} \xi^I(\lambda_n - \Delta\lambda). \quad (4.39)$$

Since  $dD_{\text{proper}} = (1 + z_n)d\lambda$  for an observer at  $z_n$ ,

$$e_{\text{out}} = (1 + z_n)^{-1} \xi_{n+}^I \quad \text{and} \quad e_{\text{in}} = (1 + z_n)^{-1} \xi_{n-}^I \quad (4.40)$$

are the angular directions of  $\gamma_1$  relative to  $\gamma_0$  before ( $e_{\text{in}}$ ), and after traversing the inhomogeneity ( $e_{\text{out}}$ ), respectively. We use (4.36), (4.37) and (4.38) to obtain

$$\xi_{n+}^I - \xi_{n-}^I = - \lim_{\Delta\lambda \searrow 0} \frac{d}{d\lambda} B(\lambda_n + \Delta\lambda) c_n(\xi_n^I) = -(1 + z_n) c_n(\xi_n^I) \quad (4.41)$$

with (4.40) this becomes

$$(e_{\text{out}} - e_{\text{in}}) = -c_n(\xi_n^I). \quad (4.42)$$

Hence,  $c_n(\xi_n^I)$  is the difference of the deflection angles at the screen position  $\xi_n^I$  and the reference ray position ( $\xi_n = 0$ ). We now calculate the value of the vector  $c_n(\xi_n^I)$  as a function of the surface mass density  $\Sigma$  of the inhomogeneity and show that it is equal to the difference of the deflection angles  $\hat{\alpha}_n(\xi_n^I) - \hat{\alpha}_n(0)$  used in lens theory.

Consider a family of rays forming an infinitesimal beam with central ray  $\gamma_1$ ; we denote their screen vectors in the  $n$ th lens plane by  $\xi_n = \xi_n^I + \Delta\xi_n$  and their angular positions relative to  $\gamma_1$  at the observer by  $\Delta\theta$ . Discussing the Jacobian map of this infinitesimal beam  $\mathcal{D}_n(\xi_n^I) = \partial\Delta\xi_n/\partial\Delta\theta$  and its derivatives

$$\dot{\mathcal{D}}_n^+(\xi_n^I) = \frac{\partial\Delta\xi_{n+}}{\partial\Delta\theta} \quad \dot{\mathcal{D}}_n^-(\xi_n^I) = \frac{\partial\Delta\xi_{n-}}{\partial\Delta\theta}$$

at the inhomogeneity, we obtain with (4.41) for the difference of these matrices

$$\begin{aligned} \dot{\mathcal{D}}_n^+(\xi_n^I) - \dot{\mathcal{D}}_n^-(\xi_n^I) &= -(1 + z_n) \left[ \frac{\partial c_n(\xi_n)}{\partial\Delta\theta} \right]_{\xi_n=\xi_n^I} \\ &= -(1 + z_n) \left[ \left( \frac{\partial c_n(\xi_n)}{\partial\xi_n} \right) \left( \frac{\partial\xi_n}{\partial\Delta\theta} \right) \right]_{\xi_n=\xi_n^I} \\ &= -(1 + z_n) \left[ \left( \frac{\partial c_n}{\partial\xi_n} \right) \mathcal{D}_n(\xi_n) \right]_{\xi_n=\xi_n^I}. \end{aligned} \quad (4.43)$$

On the other hand, we have from (4.26)

$$\dot{D}_n^+(\xi_n^1) - \dot{D}_n^-(\xi_n^1) = -(1+z_n) [\tilde{U}_n(\xi_n) \mathcal{D}_n(\xi_n)]|_{\xi_n=\xi_n^1}.$$

This implies

$$\left[ \frac{\partial c_n(\xi_n)}{\partial \xi_n} \right] \Big|_{\xi_n=\xi_n^1} = [\tilde{U}_n(\xi_n)]|_{\xi_n=\xi_n^1} = \mathcal{H} [\tilde{\Psi}_n(\xi_n)]|_{\xi_n=\xi_n^1} \quad (4.44)$$

for every  $\xi_n^1$  and therefore,

$$\frac{\partial c_n(\xi_n)}{\partial \xi_n} = \mathcal{H} [\tilde{\Psi}_n(\xi_n)] \Rightarrow c_n(\xi_n) = \nabla_{\xi_n} \tilde{\Psi}_n(\xi_n) + \text{const.} \quad (4.45)$$

The additive constant has to be  $-\nabla_{\xi_n} \tilde{\Psi}_n(0)$ ; this can be obtained from the limit  $\theta \rightarrow 0$ , i.e. the case where the ray considered coincides with the reference ray: for this ray  $\xi(z) \equiv 0$ . Therefore, we finally obtain with (4.18)

$$c_n(\xi_n) = \frac{4G}{c^2} \int_{\mathbb{R}^2} \Sigma(\xi') \left[ \frac{(\xi_n - \xi')}{|\xi_n - \xi'|^2} - \frac{(0 - \xi')}{|0 - \xi'|^2} \right] d^2 \xi' \equiv \hat{\alpha}_n(\xi_n) - \hat{\alpha}_n(0) \quad (4.46)$$

as claimed before this is the difference of the deflection angles of the ray considered and the reference ray.

*The lens equation.* Evaluating (4.37) at redshift  $z_{n+1}$ , inserting  $B(z_{n+1}) = D(z_n, z_{n+1})$ , (4.46), (4.29) with  $r(z_n, z_{n-1}) = -|r(z_n, z_{n-1})|$ , Etherington's reciprocity relation, and dropping the indices 'I' yields:

$$\xi_{n+1} = -\frac{(1+z_{n-1})D_{n,n+1}}{(1+z_n)D_{n-1,n}} \xi_{n-1} + \frac{D_{n-1,n+1}}{D_{n-1,n}} \xi_n - D_{n,n+1} [\hat{\alpha}_n(\xi_n) - \hat{\alpha}_n(0)]. \quad (4.47)$$

Using the quantities  $v_n^2$  and  $\beta_{n,n+1}$  and the dimensionless impact vectors  $\mathbf{x}_j = \xi_j/D_j$  shows that the first term on the rhs of (4.47) can be rewritten as

$$-\frac{(1+z_{n-1})D_{n,n+1}}{(1+z_n)D_{n-1,n}} \xi_{n-1} = -v_n^2 D_{n+1} \mathbf{x}_{n-1} \quad (4.48a)$$

for the second one, using the equations (C2) and (C5) of paper I, we obtain:

$$\frac{D_{n-1,n+1}}{D_{n-1,n}} \xi_n = D_{n+1}(1+v_n^2) \mathbf{x}_n. \quad (4.48b)$$

With the definition of the scaled deflection angle  $\alpha := (D_j/D_s)\hat{\alpha}$ , we find

$$D_{n,n+1} [\hat{\alpha}_n(\mathbf{x}_n) - \hat{\alpha}_n(0)] = D_{n+1} \beta_{n,n+1} [\alpha_n(\mathbf{x}_n) - \alpha_n(0)] \quad (4.48c)$$

inserting the equations (4.48) in (4.47) yields the dimensionless recurrence relation for the impact vectors  $\mathbf{x}_j$  in the lens planes

$$\mathbf{x}_{n+1} = (1+v_n^2) \mathbf{x}_n - v_n^2 \mathbf{x}_{n-1} - \beta_{n,n+1} [\alpha_n(\mathbf{x}_n) - \alpha_n(0)] \quad 1 \leq n \leq N. \quad (4.49)$$

We transform the centre of the coordinate system in each lens plane such that

$$x'_j := x_j - \sum_{i=1}^{j-1} \beta_{ij} \alpha_i(0) \quad (4.50a)$$

define

$$\alpha'_j(x'_j) := \alpha_j(x_j) \quad (4.50b)$$

and obtain with (C8) of paper I and the comment below this equation in paper I, the recurrence relation one uses in lens theory (see paper I, equation (2.19)):

$$x'_{n+1} = (1 + v_n^2) x'_n - v_n^2 x'_{n-1} - \beta_{n,n+1} \alpha'_n(x'_n) \quad 1 \leq n \leq N. \quad (4.51)$$

Whereas (4.49) describes the mapping of a ray relative to a reference ray, which is also deflected at every inhomogeneity, (4.51) describes the mapping of a ray relative to the 'optical axis'. This optical axis can be constructed by piecewise smooth null geodesics (of the empty cone metric) connecting the (new) centres of the coordinate systems on consecutive lens planes with each other; thus this optical axis represents a kinematically possible ray, but not necessarily an actual light ray (see Fermat's principle in SEF, e.g. chapter 9.2). It has been shown already in SEF that the formulation (4.51) of the multiple lens plane equation is equivalent to the more familiar one (now we drop the primes),

$$x_j = x_1 - \sum_{i=1}^{j-1} \beta_{ij} \alpha_i(x_i) \quad 1 \leq j \leq N+1$$

for the special case  $j = N+1$ , we obtain with  $\beta_{i,N+1} = 1$  for the source position that

$$y := x_{N+1} = x_1 - \sum_{i=1}^N \alpha_i(x_i).$$

Therefore we have shown in this chapter that the equations describing the mapping of a light ray and that of a light beam in gravitational lens theory can be derived, with a series of well defined approximations, from the description of light propagation in geometrical optics. In essence, the multiple deflection gravitational lens equation can be viewed as a discretization of the exact propagation equation (2.4), applied to the case of weak gravitational fields (but not necessarily weak matter inhomogeneities).

#### 4.5. Remark on Fermat's principle

In SEF, section 4.6, the derivation of the lens equation was based on a relativistic version of Fermat's principle. The argument leading to the geometric contribution to the time delay, equation (4.65), p 145 in SEF, suffers from an apparent inconsistency. On p 143, it is first stated that light rays from the source to the neighbourhood of the deflector and, after deflection, those from that neighbourhood to the observer, form 'shear-free beams ... subject only to the focusing of the smooth part of matter', i.e. to  $\tilde{\alpha} \rho_F$ ; but the subsequent calculations are said to be based on the large scale RW metric which is related to the average density,  $\rho_F$ . This, however, presents an apparent difficulty only. In the 'empty' region, outside clumps, the shear of light beams is assumed negligible there. Now, it is known that the

only conformally flat non-static dust spacetimes are Friedmann ones (see Kramer *et al* 1980, sections 22.2, 32.42, 32.5). Therefore, it seems reasonable to approximate the universe in 'empty' cone regions by a Friedmann model whose mean motion equals that of the large scale background model, but whose density is  $\tilde{\alpha}\rho_F$ . This implies that the metric,  $\overline{ds}^2$ , is related to the large scale metric,  $ds^2$ , by a constant conformal factor,

$$\overline{ds}^2 = \tilde{\alpha}^{-1} ds^2 = \tilde{\alpha}^{-1} R^2(\eta) \{d\eta^2 - d\sigma_k^2\}.$$

Therefore, the spatial paths of light rays in empty regions are the same for  $\overline{ds}^2$  as for  $ds^2$ , namely geodesics 'of  $d\sigma_k^2$ ', and the reasoning on pp 144–5 leading to equation (4.65) applies without change, since that equation is invariant under a constant rescaling of the RW metric. (Angles and the redshift  $z_d$  remain unchanged, and the distances  $c\Delta_{\text{geom.}}$ ,  $D_d$ ,  $D_s$ ,  $D_{ds}$  are rescaled by the same factor.)

## 5. The magnification of the flux of light beams

### 5.1. The flux of a radiation field, magnification factor

The monochromatic flux  $S_\omega$  of a radiation field, measured by an observer at frequency  $\omega$ , is given by the product of its specific intensity  $I_\omega$  and the solid angle  $d\Omega$  the source subtends on the observer's sky:  $S_\omega = I_\omega d\Omega$ . The specific intensity at the observer is related to that at the source by the conservation of the phase space density of photons. This implies, according to SEF, section 3.6, that for any non-interacting radiation field the scalar  $I_\omega/\omega^3$  is observer-independent, i.e. independent of his four velocity, and constant on a light beam:

$$\frac{I_{\omega(\lambda)}(\lambda)}{\omega^3(\lambda)} = \frac{I_{\omega_s}(\lambda_s)}{\omega_s^3} = \frac{I_{\omega_0}(0)}{\omega_0^3} \quad (5.1)$$

where  $\lambda$  is the affine parameter of the central light ray of the beam,  $\omega(0) =: \omega_0$  and  $\omega(\lambda_s) =: \omega_s$ .

Consider an infinitesimal monochromatic source radiating with frequency  $\omega_s$ , and observed with frequency  $\omega$ ; its observed flux  $S_\omega$  depends on the source of shear and convergence along the beam connecting source and observer. Changing these source terms such that the frequency at the observer and the affine parameter redshift relation stays the same, then, for the same observer, the observed flux of the source changes according to (5.1) to  $S_\omega = S_\omega^0 (d\Omega/d\Omega^0)$ , with  $S_\omega^0$  being the flux before changing the source terms. In an on-average Friedmann universe the frequency of the light is not changed by the deflection and, by definition, the affine parameter redshift relation is not affected by the clumps. Hence, we can compare the flux  $S_\omega$  of the source with the case where there are no intervening clumps between source and observer, and obtain for the ratio

$$\mu := \frac{S_\omega}{S_\omega^0} = \frac{d\Omega}{d\Omega^0} \quad 0 \leq \mu. \quad (5.2)$$

$\mu$  is the so called *magnification factor*; if  $\mu > 1$ , the light beam is said to be *magnified* relative to the empty beam.  $d\Omega$  and  $d\Omega^0$  are the solid angles which the source subtends on the sky for the cases with and without clumps in the beam. If we use (2.7), we obtain that the magnification  $\mu(\lambda)$  of a source at the affine parameter  $\lambda$  compared with the case where the source is observed through the empty beam, can be described as

$$\mu(\lambda) = \left| \frac{\det \mathcal{D}^0(\lambda)}{\det \mathcal{D}(\lambda)} \right| = \left| \frac{D^2(\lambda)}{\det \mathcal{D}(\lambda)} \right| = \left| \frac{1}{\det A(\lambda)} \right|. \quad (5.3)$$



For the second equality we have used that, for the empty beam, the Jacobi matrix is given by  $\mathcal{D}^0(\lambda) = D(\lambda)\mathcal{I}$ , with  $D(\lambda)$  being the angular diameter distance of the empty beam, i.e. the solution of the Dyer–Roeder equation (4.27) with boundary conditions (4.28). The third equality follows from the definition of the dimensionless Jacobian matrix  $A(\lambda) = (1/D(\lambda))\mathcal{D}(\lambda)$ . Hence, the discussion of the matrix  $A$  or the magnification factor  $\mu$  in gravitational lens theory always implies the discussion of light propagation relative to the empty beam case. This point of view is reasonable.

(1) As long as there are only a few clumps, i.e. if  $1 - \tilde{\alpha}$  is small, most light beams are empty cone beams. Therefore, the magnification factor in (5.3) describes the observed flux of a source whose beam is distorted between source and observer, relative to the most typical case, where the beam is not distorted.

(2) The other extreme is the case where  $1 - \tilde{\alpha}$  becomes approximately one: the source of convergence becomes extremely small, and for the description of the very few light beams that do not traverse a matter inhomogeneity, one cannot neglect the source of shear, which is different along every individual beam. Hence, there no longer exists a typical light beam, and the definition of the magnification factor as in (5.2) and (5.3) has no illustrative meaning: it compares the flux of the considered light beam with that of a fictitious beam.

(3) As mentioned at the beginning of section 4, for a weakly inhomogeneous universe (e.g. if one considers spatial scales on which the matter inhomogeneities are (quasi-)linear), the magnification is defined relative to that of the smooth Friedmann–Lemaître universe. In this case, the angular diameter distances  $D(\lambda)$  are those obtained from (4.27) with  $\tilde{\alpha} = 1$ , and  $\rho_{\text{cl}} = \delta\rho$  (the density fluctuations) can have either sign—therefore,  $\mathcal{R}_{\text{cl}}$  no longer is non-positive.

## 5.2. The relative focusing equation

The focusing equation (2.8) describes the evolution of the angular diameter distance of a light beam due to the Ricci focusing and the shear rate of the beam. In the case of an on-average Friedmann universe all light beams have the empty cone background density as a common contribution to their focusing, and different additional source terms due to the clumps. Therefore we want to derive a differential equation which describes the evolution of the beam relative to the empty beam; the source terms of this relative focusing equation are then produced by the clumps only.

Consider the differential equation

$$\frac{d^2}{d\lambda^2} w(\lambda) = [h(\lambda) + c(\lambda)] w(\lambda) \quad (5.4)$$

and let  $w(\lambda)$  be the (unknown) solution of (5.4) with boundary conditions  $w(0) = 0$  and  $\dot{w}(0) = 1$ . Assume that  $v(\lambda)$  is the well known solution of (5.4) for the case  $c(\lambda) \equiv 0$ ,

$$\frac{d^2}{d\lambda^2} v(\lambda) = h(\lambda)v(\lambda) \quad (5.5)$$

with the same boundary conditions:  $v(0) = 0$ ,  $\dot{v}(0) = 1$ . We define a strictly monotonically decreasing function  $X(\lambda)$  by

$$X(\lambda) := \int_{\lambda}^{\lambda_{\text{max}}} \frac{d\lambda}{v^2(\lambda)} \quad (5.6)$$

so that  $X(\lambda_{\max}) = 0$ ; the value of  $\lambda_{\max}$  will be specified below. Then, inserting the equations (5.5) and (5.6) in (5.4), we obtain for the ratio  $a := w/v$  the differential equation:

$$\frac{d^2}{dX^2} a(X) = v^4(X) c(X) a(X). \quad (5.7)$$

Using  $\ddot{w}(\lambda)|_{\lambda=0} = 0 = \ddot{v}(\lambda)|_{\lambda=0}$ , the boundary conditions for  $a$  become, as a function of  $\lambda$ ,

$$a(\lambda)|_{\lambda=0} = 1 \quad \frac{d}{d\lambda} a(\lambda)|_{\lambda=0} = 0. \quad (5.8)$$

We interpret† (5.4) by inserting  $h = \mathcal{R}_{bg}(\lambda)$  and  $c = \mathcal{R}_{cl}(\lambda) - |\sigma(\lambda)|^2$ ; then,  $w$  and  $v$  denote the angular diameter distances of the ‘actual’ beam considered and that in an empty cone, respectively. Therefore, (5.7) describes how the considered light beam is focused relative to the empty beam and is therefore called the *relative focusing equation*; the solution of (5.7) can be described, with

$$v(\lambda) = D(\lambda) = \frac{c}{H_0} r(z(\lambda))$$

as

$$a(\lambda) = \frac{1}{D(\lambda)} \mathcal{SQ}[\det \mathcal{D}](\lambda) = \mathcal{SQ}[\det A(\lambda)]. \quad (5.9)$$

The inverse of (5.9) yields the magnification of the beam at a position  $\lambda: |a(\lambda)|^2 = \mu^{-1}(\lambda)$ . We can identify  $(c/H_0)X$  with the cosmological  $\chi$ -function, defined in equation (4.68) of SEF, since

$$\frac{dX}{d\lambda} = - \left( \frac{c}{H_0} \right)^{-2} \frac{1}{r^2(\lambda)}$$

yields, if we put  $\lambda_{\max} = \lim_{z \rightarrow \infty} \lambda(z)$  and use equation (4.3),

$$X(z) = \left( \frac{c}{H_0} \right)^{-1} \int_z^\infty \frac{dz'}{r^2(z')(1+z')^3 \sqrt{\Omega_2 z' + 1}} = \left( \frac{c}{H_0} \right)^{-1} \chi(z). \quad (5.10)$$

Inserting (5.10) and (5.9) in (5.7), the relative focusing equation can be rewritten as

$$\frac{d^2}{d\chi^2} a(\chi) = \left( \frac{c}{H_0} \right)^2 r^4(\chi) [\mathcal{R}_{cl} - |\sigma(\chi)|^2] a(\chi) \quad a(\chi) = \mathcal{SQ}[\det A](\chi). \quad (5.11)$$

Note that due to the strictly monotonic behaviour of  $\chi$  and  $\lambda$  as functions of  $z$ , we can consider any variable on a light ray as a function of  $z$ ,  $\lambda$  or  $\chi$ .

† One can calculate the relative magnification of two light beams with (5.7) even in a case of a non-Friedmann universe, if the affine parameters of these light beams are the same (e.g. as a function of redshift).

### 5.3. The focusing theorem

The non-positiveness of the source term  $\mathcal{R}_{cl} - |\sigma|^2$  due to the clumps in the focusing equation shows that a beam propagating through clumps is always more focused than the empty beam (in the absence of conjugate points between source and observer). Hence, as long as the beam has not formed its first conjugate point, the angular diameter distance must not be greater than that of an empty comparison beam at the same redshift. This so called *focusing theorem* can be restated with the use of the relative focusing equation: as long as the light beam has not formed its first conjugate point, the function  $a(\lambda)$  is always between one and zero,

$$1 \geq a(\lambda) \geq 0 \iff \mu(\lambda) \geq 1 \quad 0 \leq \lambda \leq \lambda_c \quad (5.12)$$

or, the light beam is not de-magnified relative to one in an empty cone. This can be proved immediately. Using the boundary conditions of  $a(\lambda)$  described in equation (5.8) and that, due to the non-positiveness of  $[\mathcal{R}_{cl} - |\sigma|^2]$ , the second derivative of  $a$  in (5.11) is always non-positive, one obtains that the value of  $a$  is always between one and zero in the interval between the vertex and the first conjugate point of the beam. One can also prove a stronger statement: *as long as the beam has not passed a conjugate point, the function  $a(\lambda)$  is monotonically decreasing.*

*Proof.* Since  $\chi$  tends to plus infinity at the vertex, and  $\lim_{\chi \rightarrow \infty} (da/d\chi)(\chi) = 0$ , one can write

$$\frac{da}{d\chi}(\chi) = \int_{\infty}^{\chi} \left( \frac{c}{H_0} \right)^2 r^4(\chi') [\mathcal{R}_{cl} - |\sigma|^2] a(\chi') d\chi'$$

and therefore,  $da/d\lambda$  can be rewritten as

$$\frac{da}{d\lambda}(\lambda) = \left( \frac{c}{H_0} \right) \frac{1}{r^2(\lambda)} \int_{\chi(\lambda)}^{\infty} r^4(\chi') [\mathcal{R}_{cl}(\chi') - |\sigma|^2(\chi')] a(\chi') d\chi'.$$

Since the integrand is non-positive,  $da/d\lambda \leq 0$  follows.  $\square$

## 6. Summary and conclusions

We have investigated the propagation of infinitesimally small light beams in arbitrary spacetimes, and have derived a Jacobi type of differential equation for the matrix providing the linear mapping from the inclination angle of a light ray of the beam to the separation vector at arbitrary values of the affine parameter. This matrix carries full information about the size, shape, orientation and twist of the beam. We have then concentrated on the investigation of the behaviour of light beams near a vertex and near conjugate points; in particular, we have derived asymptotic representations of the optical scalars near such points. It has been pointed out that near a vertex and a focus, the twist of a beam is a higher order contribution to the Jacobi mapping than are expansion and shear.

We then turned to the special case where the metric is that of a perturbed Friedmann universe, i.e. where the overall geometry of the universe is described by a Friedmann metric, which, however, is locally modified to allow for matter inhomogeneities. If the matter inhomogeneities are considered to be weak, so that they can be described locally by

a post-Minkowskian metric, and geometrically thin and isolated, so that typical light beams propagate most of the time through the background Friedmann metric, the influence of the matter inhomogeneities on the light beam can be described by a sum of delta distributional contributions to the source term of the Jacobi equation for the linear mapping mentioned above. In this way we have derived the equations of gravitational lens theory, which represents an approximation to the exact propagation equations which is particularly useful for, and applies to, most astrophysically relevant situations of light propagation in the universe. We want to point out that in contrast to earlier treatments of the lens equations (e.g. SEF, section 4.6), we have made no use of the existence of an optical axis relative to which the impact vectors are defined; instead, our reference ray is a physical, i.e. deflected, light ray. To relate our formulation to the earlier treatment, a redefinition of the coordinate frames in the lens planes was performed which yielded the lens equation in the standard form.

We remind the reader that a derivation of the gravitational lens equation can also start from Fermat's principle (see Blandford and Narayan 1986, SEF section 4.6 and references therein); however, the derivation presented here appears to be more direct, in that one does not make use of geometrical constructions for the calculation of the 'geometrical time delay', which are less easy to justify in an 'on-average Friedmann universe' than the approximations used here. The advantage of our derivation of the lens equations lies in its explicit listing of approximations which have to be made. All but two are not critical and well satisfied in astrophysically relevant situations. The two which are as yet not very well understood are: (1) the source of shear has been assumed to vanish between two consecutive lens planes; (2) it has been assumed that the metric of a clumpy universe can be written locally as a post-Minkowskian modification of the standard Friedmann metric. Note that a number of investigations have suggested the validity of this latter approximation (e.g. Futamase and Sasaki 1989, Jacobs *et al* 1993). The former assumption certainly has to break down if the universe is highly clumpy, i.e. for  $\bar{\alpha}\Omega$  of order unity. However, since it seems that the clumpiness of our universe is much smaller than unity, we conclude that the (multiple deflection) gravitational lens equations provide a useful and fairly accurate approximation in most relevant cases.

Finally, we have derived an equation for the size of a light beam in a clumpy universe, relative to the size of a beam which is unaffected by the matter inhomogeneities. If we require that this second-order differential equation contains only the contribution by matter clumps as source term, the independent variable is uniquely defined and agrees with the  $\chi$ -function previously introduced (see SEF, equation (4.68)) for other reasons. This relative focusing equation immediately yields the result that a light beam cannot be less focused than a reference beam which is unaffected by matter inhomogeneities, prior to the propagation through its first conjugate point. In other words, no source can appear fainter to the observer than in the case that there are no matter inhomogeneities close to the line of sight to this source, a result previously demonstrated for the case of one (Schneider 1984) and several (paper I, Seitz and Schneider 1994) lens planes.

## References

- Blandford R D and Narayan R 1986 *Astrophys. J.* **310** 568
- 1992 *Ann. Rev. Astron. Astrophys.* **30** 311
- Blandford R D, Saust A B, Brainerd T G and Villumsen J V 1991 *Mon. Not. R. Astron. Soc.* **251** 600
- Dyer C C and Roeder R C 1972 *Astrophys. J.* **174** L115
- 1973 *Astrophys. J.* **180** L31

- Etherington I M H 1933 *Phil. Mag.* **15** 761  
 Futamase T and Sasaki M 1989 *Phys. Rev. D* **40** 2502  
 Gunn J E 1967 *Astrophys. J.* **150** 737  
 Jacobs M W, Linder E V and Wagoner R V 1993 *Phys. Rev. D* **48** 4623  
 Kaiser N 1992 *Astrophys. J.* **388** 272  
 Kasai M 1993 *Phys. Rev. D* **47** 3214  
 Kramer D, Stephani H, MacCallum M and Herlt E 1980 *Exact solutions of Einstein field equation* (Berlin: VEB)  
 Peebles P J E 1993 *Principles of Physical Cosmology* (Princeton, NJ: Princeton University Press)  
 Sachs R K 1961 *Proc. R. Soc. A* **264** 309  
 Schneider P 1984 *Astron. Astrophys.* **140** 119  
 Schneider P, Ehlers J and Falco E E 1992 *Gravitational Lenses* (Berlin: Springer)  
 Seitz S and Schneider P 1992† *Astron. Astrophys.* **265** 1  
 ——— 1994 *Astron. Astrophys.* **287** 349  
 Wald R M 1984 *General Relativity* (Chicago: University of Chicago Press)  
 Zel'dovich Ya B 1964 *Sov. Astron.* **8** 13