Canonical transformations to warped surfaces: Correction of aberrated optical images

Canonical transformations to warped surfaces: Correction of aberrated optical images

Eisso J Atzema†, Guillermo Krötzsch‡ and Kurt Bernardo Wolf‡

† Research Institute for Applications of Computer Algebra, Stichting CAN, Kruislaan 419, 1099 SJ Amsterdam, the Netherlands
‡ Instituto de Investigaciones en Matemáticas Aplicadas y en Sistemas–Cuernavaca, Universidad Nacional Autónoma de México, Apartado Postal 48-3, 62251 Cuernavaca, Morelos, México

Abstract. The projection of optical images on warped screens is a canonical transformation of phase space between flat and warped evolution-parameter surfaces. In mechanics, the evolution parameter is time; in geometric optics it is the optical axis of coordinate space. We consider the specific problem of bringing to focus an axis-symmetric aberrating optical system by warping the output screen. The solution for the surface curvature coefficients is given in terms of the Lie aberration coefficients of the system; a linear optimization strategy applies.

1. Introduction

The dynamical evolution of a Hamiltonian system is a canonical transformation of its phase space generated by its Hamiltonian function. The evolution parameter in classical mechanics is normally time t, while in geometrical optics its counterpart is the coordinate along the optical z-axis. The 'final' phase space is a deformed image of the 'initial' phase space, where volume is preserved as in the flow of an incompressible fluid, and the Poisson bracket structure conserved. The evolution of the initial phase-space points can be pictured by snapshots for every time instant $t_i \leqslant t \leqslant t_f$ in classical mechanics. In geometric optics, 'final' and 'initial' refer to ideal plane screens placed perpendicular to the optical z-axis, and $z_i \leqslant z \leqslant z_f$, where position and momentum (local index of refraction times ray direction) can be measured at any plane z [1].

Symplectic geometry has been recognized as a very appropriate language to describe both mechanics and optics [2]. Geometrical and wave optics, the latter particularly in the paraxial regime, have been addressed through symplectic techniques because they allow for coherency and other fundamental properties of light in a very natural way, as in the programs of Mukunda, Simon and Sudarshan (see [3]). In classical mechanics, the prevailing paradigm of Newtonian time has discouraged applied studies of canonical transformations to non-flat space-time surfaces. Geometrical optics on the other hand, fundamentally considers transformations due to non-flat surfaces of discontinuity between two different optical media, and the projection of images on screens that may be warped. In section 2 we show that the *root* transformation [4] is the canonical map between rays from a reference to a warped screen. Two root transformations, composed to and from a warped mirror or interface, compose to the reflection and refraction maps of optical rays in phase space.

We apply the root map to examine the general problem of minimizing the focusing error of an optical apparatus by warping the output screen. To this end we use Lie theoretical

methods, fully analytic in all the parameters of the system. In section 3 we recall the linear (paraxial) and Lie aberration coefficients of such systems in the metaxial régime. To third-order aberration, we confirm in section 4 the known result that only third-order curvature of field can be corrected on a paraboloidal screen [5].

Section 5 distinguishes non-correctable aberrations from the three controllable ones at each aberration order, and the two (distortion and *p*-unfocus or pocus) which are immaterial to image formation, in fifth and higher orders. The induced selection rules were verified to hold to ninth-order aberration. The generic and optimized solutions to problems of image correction can thus be handled by symbolic computer programs [6]. In section 6 we highlight the pattern of selection rules that arise in warping the screen in the context of the geometric optical realization of canonical maps.

2. Optical phase space and root transformation

The phase space of geometric optics is the manifold of directed lines (rays) in three-dimensional optical space. It is a four-dimensional symplectic manifold conveniently (but not globally) parametrized by a 2-vector of *position* $q \in \Re^2$ (the intersection of the line with the *standard screen* S_0: the $z = 0$ plane), and a canonically conjugate 2-vector of *momentum* p [1]. When the optical medium is homogeneous, the lines are straight. Momentum is the projection on S_0 of a 3-vector along the ray, whose length is the local refractive index n; if θ is the angle between the ray and the optical axis, $|p| = n|\sin\theta| \leqslant n$. (There are actually two disks $|p| \leqslant n$ distinguished by a sign σ indicating that the ray heads in the $\pm z$-direction.) However, in the *metaxial régime*, and for the purpose of aberration (Taylor) expansions, one considers $p \in \Re^2$, as in two-dimensional mechanics (and only the chart $\sigma = +$ of 'forward' rays). Evolution thus occurs in the $+z$-direction, corresponding to time in mechanics. In figure 1 we show the deformation of a phase-space patch due to free flight in two-dimensional optical space; it is easier to plot since phase space is $\Re^2$.

Consider now a surface S_ζ given by $z = \zeta(q)$ in three-dimensional optical space (q, z). One must assume ζ is at least twice differentiable and single-valued [4]; in this paper we work with polynomial revolution surfaces, the primary purpose being algorithmic. The ray $(p, q) \in \Re^4$ will intersect S_ζ, as shown in the two-dimensional figure 2, at the point $\bar{q}$, namely

$$\mathcal{R}_{n;\zeta}; \, q \mapsto \bar{q}(p, q) = q + \zeta(\bar{q})p/\sqrt{n^2 - p^2}. \tag{2.1a}$$

This is only half of the phase-space transformation. The optical momentum $\bar{p}$, canonically
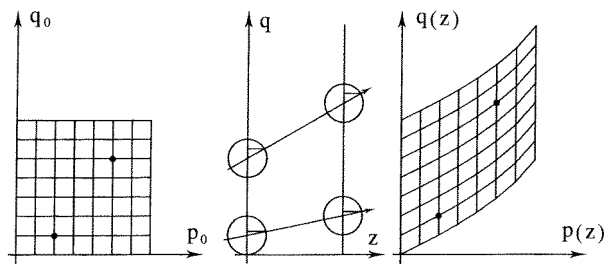
Figure 1. Optical phase space parametrizes rays by their *position* q and *momentum* p, with respect to the optical z-axis and at the standard screen. Free flight is described by a flat screen advancing in the direction of the optical axis; it deforms phase space preserving its volume elements.
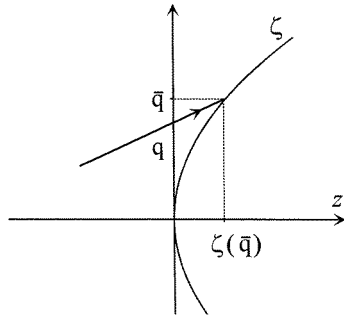
Figure 2. The root transformation maps the optical phase space of rays at the standard screen onto a phase space on the surface S_ζ given by $z = \zeta(q)$. The point of impact of the ray (p, q) on S_ζ is $\bar{q}$. The canonically conjugate map yields momentum $\bar{p}$ (not shown) on the warped surface S_ζ.

conjugated to the intersection point $\bar{q}$ on the surface S_ζ, can be found from geometric considerations [4]. The result is

$$\mathcal{R}_{n;\zeta} : p \mapsto \bar{p}(p, q) = p + \nabla\zeta(\bar{q})\sqrt{n^2 - p^2} \tag{2.1b}$$

where $\nabla\zeta(\bar{q})$ is the normal to S_ζ at the point of intersection. Together, equations (2.1) define $\mathcal{R}_{n;\zeta}$ as the *root transformation* of the chart of phase space that contains the ray of interest, due to the surface S_ζ in the medium n. The boundary of the chart is the set of rays tangent to the surface at any point. Again, since our approach here is that of aberration expansions, we do not treat the global properties of the transformation in this paper.

The root transformation is the factor to reflection and refraction; this was its first interpretation. Indeed, reflection by the surface S_ζ is $\mathcal{R}_{n;\zeta}\mathcal{R}_{n;-\zeta}^{-1}$. (Changing the sign of the z-component of $\vec{p}$ is equivalent to changing the sign of ζ, or the sign of the square roots in (2.1).) Refraction between media n and n' separated by the same surface S_ζ is $\mathcal{R}_{n;\zeta}\mathcal{R}_{n';\zeta}^{-1}$. For flat $\zeta(q) = \zeta_0$, $\mathcal{R}_{n;\zeta_0}$ is free flight by a distance ζ_0, i.e. $\bar{p} = p$, $\bar{q} = q + \zeta_0 \tan\theta$ (see figure 1). An important property is that free flight between planes can be factored to the left of the root transformation, because for $\zeta(q) = \zeta_0 + \zeta_1(q)$, $\mathcal{R}_{n;\zeta} = \mathcal{R}_{n;\zeta_0}\mathcal{R}_{n;\zeta_1}$. It is important to note that in equation (2.1b), $|\nabla\zeta|$ is unconstrained, and therefore $|\bar{p}|$ is not bound by n or any number; hence $(\bar{p}, \bar{q})$ is not *optical* phase space. Nevertheless, because we shall work in the metaxial régime, where polynomial expansions are of interest in the local neighbourhood around the chosen optical axis, here we can gloss over the important issues of the global domain and range of the root transformation.

From the computational point of view, the root transformation (2.1a) provides an implicit equation of partially reduced form $\bar{q} = q + F(p, q, \bar{q})$ to find the impact point $\bar{q}(p, q)$; this is amenable to algorithmic recursion and solution by aberration series. Once $\bar{q}$ is known, the solution $\bar{p}(p, q)$ of (2.1b) is explicit. If the surface S_ζ is tangent to the screen S_0 and is of revolution, with the polynomial expansion

$$z = \zeta(q^2) = \zeta_2 q^2 + \zeta_4(q^2)^2 + \zeta_6(q^2)^3 + \zeta_8(q^2)^4 + \cdots \tag{2.2}$$

then, to third-order aberration (i.e. keeping terms of degree up to three in the phase-space variables), the root map in vector/matrix form is

$$\mathcal{R}_{n;\zeta} : \begin{pmatrix} p \\ q \end{pmatrix} \mapsto \begin{pmatrix} \bar{p}(p, q) \\ q(p, q) \end{pmatrix} = \begin{pmatrix} 1 & 2n\zeta_2 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} p \\ q \end{pmatrix}$$

$$+ \begin{pmatrix} 2\zeta_2^2 q^2 \boldsymbol{p} - \zeta_2 \frac{1}{n} p^2 \boldsymbol{q} + 4n\zeta_4 q^2 \boldsymbol{q} \\ \zeta_2 \frac{1}{n} q^2 \boldsymbol{p} \end{pmatrix} + O_5(\boldsymbol{p}, \boldsymbol{q}). \tag{2.3}$$

The root transformation (2.1) is canonical. This can be proven directly for the aberration series approximants: for (2.3) the Poisson brackets are $\{\bar{q}(\boldsymbol{p}, \boldsymbol{q}), \bar{p}(\boldsymbol{p}, \boldsymbol{q})\} = 1 + O_5(\boldsymbol{p}, \boldsymbol{q})$. The straightforward proof for the closed forms (2.1) is surprisingly difficult because of the implicit equation for $\bar{q}$. In [4, appendix A] we used differential-form techniques to prove that the two-form is preserved: $\mathrm{d}\boldsymbol{p}\,\mathrm{d}\boldsymbol{q} = \mathrm{d}\bar{\boldsymbol{p}}\,\mathrm{d}\bar{\boldsymbol{q}}$ (cf Arnold [2], p 239). A constructive proof due to Delgado [7] writes (2.1) as

$$\boldsymbol{q} = \frac{\partial D_{n;\zeta}(\bar{\boldsymbol{q}}, \boldsymbol{p})}{\partial \boldsymbol{p}} \qquad \bar{\boldsymbol{p}} = \frac{\partial D_{n;\zeta}(\bar{\boldsymbol{q}}, \boldsymbol{p})}{\partial \bar{\boldsymbol{q}}}. \tag{2.4}$$

The Hamiltonian characteristic function $D_{n;\zeta}(\bar{\boldsymbol{q}}, \boldsymbol{p})$ which generates this (therefore *canonical*) transformation [8] is obtained from the identity characteristic function in three dimensions, $I(\vec{\bar{q}}, \vec{p}) = \vec{\bar{q}}\vec{p} = \bar{\boldsymbol{q}} \cdot \boldsymbol{p} + \bar{q}_z p_z$, constrained to the optical momentum sphere $|\vec{p}|^2 = p_z^2 + |\boldsymbol{p}|^2 = n^2$, and to the surface $\bar{q}_z = \zeta(\boldsymbol{q})$. This is

$$D_{n;\zeta}(\bar{\boldsymbol{q}}, \boldsymbol{p}) = \bar{\boldsymbol{q}} \cdot \boldsymbol{p} + \zeta(\bar{\boldsymbol{q}})\sqrt{n^2 - |\boldsymbol{p}|^2}. \tag{2.5}$$

Placed in (2.4), this reproduces (2.1).

It follows thus that both reflection and refraction are canonical transformations: the presence of a surface S_ζ provides a map between two phase spaces, unprimed or 'initial' $(\boldsymbol{p}, \boldsymbol{q})$, and primed or 'final' $(\boldsymbol{p}', \boldsymbol{q}')$, such that

$$\mathcal{R}_{n;\zeta}\begin{pmatrix} \boldsymbol{p} \\ \boldsymbol{q} \end{pmatrix} = \begin{pmatrix} \bar{\boldsymbol{p}} \\ \bar{\boldsymbol{q}} \end{pmatrix} = \mathcal{R}_{n';\zeta}\begin{pmatrix} \boldsymbol{p}' \\ \boldsymbol{q}' \end{pmatrix}. \tag{2.6}$$

The laws of refraction are contained in the statement that $(\bar{\boldsymbol{p}}, \bar{\boldsymbol{q}})$ are *conserved*. The position $\bar{\boldsymbol{q}}$ is the point of intersection: the initial and final rays meet at S_ζ. Its canonically conjugate momentum $\bar{\boldsymbol{p}}$ at the point of impact, in cross product with the surface normal $\nabla\zeta$ leads to the law of sines and coplanarity of refraction [4, 7]. For reflection we set $(n', \zeta) \mapsto (n, -\zeta) \sim (-n, \zeta)$.

We take this opportunity to stress that the root transformation and its composition to reflection and refraction have not been as exhaustively studied as their purported importance, both physical and mathematical, would require. In particular, it would be desirable to understand fully its global aspects and caustic problems which do not appear in the present aberration–expansion treatment.

3. Imaging and aberrating systems

We pose the concrete problem of this article in the following way: assume an optical instrument with axial and reflection symmetry is built to focus in first order, $\boldsymbol{q}' = \mu\boldsymbol{q} + O_3(\boldsymbol{p}, \boldsymbol{q})$, $\mu \neq 0$, with third- and higher-order aberrations; can we correct the focusing ability of the instrument by warping the screen? To articulate the answer, we recall the definitions of the paraxial and Lie aberration parts of a general (optical) system.

An *imaging system* $\mathcal{A}$ is a canonical map of phase space written in the generic form

$$\mathcal{A} = \mathcal{G}\{\mathbf{A}, \mathbf{M}\} = \mathcal{G}\{\mathbf{A}, \mathbf{1}\}\mathcal{G}\{\mathbf{0}, \mathbf{M}\} \qquad \text{where } \mathbf{M} = \begin{pmatrix} \mu & \lambda \\ 0 & \mu^{-1} \end{pmatrix} \tag{3.1}$$

is the *paraxial part* of the transformation, whose action on phase space is linear,

$$\mathcal{G}\{\mathbf{0}, \mathbf{M}\}\begin{pmatrix} \boldsymbol{p} \\ \boldsymbol{q} \end{pmatrix} = \mathbf{M}^{-1}\begin{pmatrix} \boldsymbol{p} \\ \boldsymbol{q} \end{pmatrix} = \begin{pmatrix} \mu^{-1} & -\lambda \\ 0 & \mu \end{pmatrix}\begin{pmatrix} \boldsymbol{p} \\ \boldsymbol{q} \end{pmatrix} = \begin{pmatrix} \mu^{-1}\boldsymbol{p} - \lambda\boldsymbol{q} \\ \mu\boldsymbol{q} \end{pmatrix} = \begin{pmatrix} \boldsymbol{p}'(\boldsymbol{p}, \boldsymbol{q}) \\ \boldsymbol{q}'(\boldsymbol{q}) \end{pmatrix}$$

$$\tag{3.2}$$

and **A** labels the *aberration part*. The unit system is $\mathcal{G}\{\mathbf{0},\mathbf{1}\}$ and the composition rule between two systems holds as $\mathcal{G}\{\mathbf{A},\mathbf{M}\}\mathcal{G}\{\mathbf{B},\mathbf{N}\} = \mathcal{G}\{\mathbf{A}\sharp\mathbf{D}(\mathbf{M})\mathbf{B},\mathbf{MN}\}$, with $\sharp$ a Baker–Campbell–Hausdorff-type product and $\mathbf{D}(\mathbf{M})$ a representation of the 2×2 matrix $\mathbf{M}$ [9]. (The specific formulae are not needed here.) The aberration part is further factorized into a product of Lie–Poisson exponential operators corresponding to aberrations of increasing order [9, 10]

$$\mathcal{G}\{\mathbf{A},\mathbf{1}\} = \cdots \times \exp\{A_4,\circ\}\exp\{A_3\circ\}\exp\{A_2,\circ\} \tag{3.3a}$$

$$\exp\{A_k,\circ\} = 1 + \{A_k,\circ\} + \frac{1}{2!}\{A_k,\{A_k,\circ\}\} + \cdots \tag{3.3b}$$

$$\{A_k,\circ\} = \frac{\partial A_k(\boldsymbol{p},\boldsymbol{q})}{\partial \boldsymbol{q}} \cdot \frac{\partial}{\partial \boldsymbol{p}} - \frac{\partial A_k(\boldsymbol{p},\boldsymbol{q})}{\partial \boldsymbol{p}} \cdot \frac{\partial}{\partial \boldsymbol{q}} \tag{3.3c}$$

where $\{f,\circ\}g(\boldsymbol{p},\boldsymbol{q}) = \{f,g\}(\boldsymbol{p},\boldsymbol{q})$ is the Poisson Lie operator and bracket, and $A_k(\boldsymbol{p},\boldsymbol{q})$ are *aberration polynomials* of homogeneous degree $2k$ in the phase-space coordinates (p_i, q_j). We call $k \geqslant 2$ the *rank* of the aberration.

The aberration polynomials of axis-symmetric systems can only contain the axis-symmetric coordinates of phase space: the three quadratic variables $p^2 = \sum_i p_i^2$, $\boldsymbol{p} \cdot \boldsymbol{q} = \sum_i p_i q_i$, and $q^2 = \sum_i q_i^2$. Their generic form is

$$A_k(p^2, \boldsymbol{p} \cdot \boldsymbol{q}, q^2) = \sum_{k_+ + k_0 + k_- = k} A_{k_+,k_0,k_-}(p^2)^{k_+}(\boldsymbol{p} \cdot \boldsymbol{q})^{k_0}(q^2)^{k_-} \tag{3.4}$$

where $\{A_{k_+,k_0,k_-}\}_{k_+ + k_0 + k_- = k}$ are the $\frac{1}{2}(k+1)(k+2)$ *aberration coefficients* of that rank in the *monomial basis* that we label by the triplet (k_+, k_0, k_-). Below, we represent them by points in a hexagonal array in a triangle. (The *symplectic* basis, which is block diagonal with respect to the action of the paraxial $\mathrm{Sp}(2,\Re)$ group, is preferred for symbolic computation because composition is faster [6, 11].)

If the aberration part **A** is not zero—as it generally is not in actual optical devices—the system will produce images on the flat screen that are not independent of the object ray momentum $\boldsymbol{p}$ but, from (3.2),

$$\mathcal{A}\begin{pmatrix}\boldsymbol{p}\\\boldsymbol{q}\end{pmatrix} = \mathcal{G}\{\mathbf{A},\mathbf{1}\}\mathcal{G}\{\mathbf{0},\mathbf{M}\}\begin{pmatrix}\boldsymbol{p}\\\boldsymbol{q}\end{pmatrix} = \mathcal{G}\{\mathbf{A},\mathbf{1}\}\begin{pmatrix}\boldsymbol{p}'(\boldsymbol{p},\boldsymbol{q})\\\boldsymbol{q}'(\boldsymbol{q})\end{pmatrix} = \begin{pmatrix}\boldsymbol{p}'(\boldsymbol{p}_A(\boldsymbol{p},\boldsymbol{q}),\boldsymbol{q}_A(\boldsymbol{p},\boldsymbol{q}))\\\boldsymbol{q}'(\boldsymbol{q}_A(\boldsymbol{p},\boldsymbol{q}))\end{pmatrix}$$
$$\tag{3.5}$$

where the generic aberrated coordinates are, by (3.3),

$$\mathcal{G}\{\mathbf{A},\mathbf{1}\}\begin{pmatrix}\boldsymbol{p}\\\boldsymbol{q}\end{pmatrix} = \cdots \times (1 + \{A_4,\circ\} + \cdots)(1 + \{A_3,\circ\} + \cdots) \times (1 + \{A_2,\circ\}$$

$$+ \frac{1}{2!}\{A_2,\{A_2,\circ\}\} + \frac{1}{3!}\{A_2,\{A_2,\{A_2,\circ\}\}\} + \cdots)\begin{pmatrix}\boldsymbol{p}\\\boldsymbol{q}\end{pmatrix}$$

$$= (1 + \{A_2,\circ\} + \{A_3,\circ\} + \frac{1}{2!}\{A_2,\{A_2,\circ\}\} + \{A_4,\circ\} + \{A_3,\{A_2,\circ\}\}$$

$$+ \frac{1}{3!}\{A_2,\{A_2,\{A_2,\circ\}\}\} + \cdots)\begin{pmatrix}\boldsymbol{p}\\\boldsymbol{q}\end{pmatrix} = \begin{pmatrix}\boldsymbol{p}_A(\boldsymbol{p},\boldsymbol{q})\\\boldsymbol{q}_A(\boldsymbol{p},\boldsymbol{q})\end{pmatrix}. \tag{3.6}$$

Since the Poisson bracket $\{A_k, A_{k'}\}$ of two homogeneous polynomials of degrees $k, k' \geqslant 2$ (in the quadratic variables), is of degree $k + k' - 1 > k, k'$, terms beyond the ellipses can be disregarded consistently. To third order explicitly,

$$\mathcal{G}\{\mathbf{A},\mathbf{1}\}|_{k=2} = 1 + \{A_2,\circ\} \tag{3.7a}$$

$$A_2(\boldsymbol{p},\boldsymbol{q}) = A_{2,0,0}(p^2)^2 + A_{1,1,0}p^2\boldsymbol{p}\cdot\boldsymbol{q} + A_{1,0,1}p^2q^2 + A_{0,2,0}(\boldsymbol{p}\cdot\boldsymbol{q})^2$$
$$+ A_{0,1,1}\boldsymbol{p}\cdot\boldsymbol{q}q^2 + A_{0,0,2}(q^2)^2. \tag{3.7b}$$

There are six independent *Lie* aberrations to third order, $\{A_{k_+,k_0,k_-}\}_{k_++k_0+k_-=2}$ (cf [5, 10]): 200, spherical aberration; 110, coma; 101, curvature of field; 020, astigmatism; 011, distortion; 002, pocus (defocus) [9].

The third order, the aberrated phase-space coordinates (3.6) are then

$$p_A^{[3]}(p, q) = \mathcal{G}\{\mathbf{A}, \mathbf{1}\}p = p + (A_{1,1,0}p^2 + 2A_{0,2,0}p \cdot q + A_{0,1,1}q^2)p + (2A_{1,0,1}p^2$$
$$+ 2A_{0,1,1}p \cdot q + 4A_{0,0,2}q^2)q \tag{3.8a}$$

$$q_A^{[3]}(p, q) = \mathcal{G}\{\mathbf{A}, \mathbf{1}\}q = q - (4A_{2,0,0}p^2 + 2A_{1,1,0}p \cdot q + 2A_{1,0,1}q^2)p - (A_{1,1,0}p^2$$
$$+ 2A_{0,2,0}p \cdot q + A_{0,1,1}q^2)q. \tag{3.8b}$$

In particular, for $q_A^{[3]}(p, q)$ to be independent of p to *third*-order aberration, it is sufficient that the first four aberration coefficients be zero: $A_{2,0,0} = A_{1,1,0} = A_{1,0,1} = A_{0,2,0} = 0$. It is not necessary that the *distortion* coefficient $A_{0,1,1}$ vanish, because such a system *will* produce a (third-order) image $q_A^{[3]}(q) = q - A_{0,1,1}q^2q$ that *is* in focus, albeit distorted. Finally, the *pocus* coefficient $A_{0,0,2}$ does not affect image formation at all; it only changes the p-focus of the rays forming the image at the screen, $p_A^{[3]}(p, q)$. (Pocus acts as a spherical aberration in the conjugate momentum variable.)

4. Third-order correction by warped screens

We first characterize the root transformation (2.1) as a paraxial imaging system with the Lie aberrations defined in the last section. Using the matrix in the paraxial part of (2.3), equations (3.5) and (3.6) are written to third-order aberration as

$$\mathcal{R}_{n,\zeta} : \begin{pmatrix} p \\ q \end{pmatrix} = \mathcal{G}\left\{\mathbf{R}, \begin{pmatrix} 1 & -2n\zeta_2 \\ 0 & 1 \end{pmatrix}\right\} \begin{pmatrix} p \\ q \end{pmatrix}$$
$$= \mathcal{G}\{\mathbf{R}, \mathbf{1}\} \begin{pmatrix} p + 2n\zeta_2 q \\ q \end{pmatrix} = \begin{pmatrix} (1 + \{R, \circ\})(p + 2n\zeta_2 q) \\ (1 + \{R, \circ\})q \end{pmatrix}. \tag{4.1a}$$

Comparison with the aberration part of (2.3) finds the third-order aberration polynomial of the root transformation:

$$R_2(p, q) = -\frac{\zeta_2}{2n}p^2q^2 + n\zeta_4(q^2)^2. \tag{4.1b}$$

The third-order aberration coefficients of the root transformation are therefore

$$R_{1,0,1} = -\zeta_2/2n \qquad R_{0,0,2} = n\zeta_4 \tag{4.1c}$$

and the four others vanish: $R_{2,0,0} = R_{1,1,0} = R_{0,2,0} = R_{0,1,1} = 0$.

We can now restate the concrete problem to which we address this paper. Assume a general imaging system $\mathcal{A} : (p, q) \mapsto (p', q')$ in paraxial focus, but with aberrations $\mathbf{A}$. We use the root transformation $\mathcal{R}_{n,\zeta} : (p', q') \mapsto (\bar{p}, \bar{q})$ as a *bona fide* optical element performing free flight to the warped surface S_ζ. The purpose of the correction is to sharpen the point images $\bar{q}(p, q)$ on S_ζ by eliminating some aberrations. The compound optical system that projects images on the warped screen is shown in figure 3. It is $\mathcal{C} = \mathcal{A}\mathcal{R}_{n;\zeta}$, and its action on phase space $w = (p, q)$ is

$$\mathcal{C} : w \mapsto \bar{w} = \mathcal{A}\mathcal{R} : w = \mathcal{A} : \bar{w}(w) = \bar{w}(\mathcal{A} : w) = \bar{w}(w'(w)). \tag{4.2}$$

The third-order approximation can be found from (2.3) and (3.5)–(3.8) kept to the third degree. For the position coordinate we find

$$\bar{q}^{[3]}(p, q) = q'(p, q) + \frac{\zeta_2}{n}q'^2(p, q)p'(p, q)$$
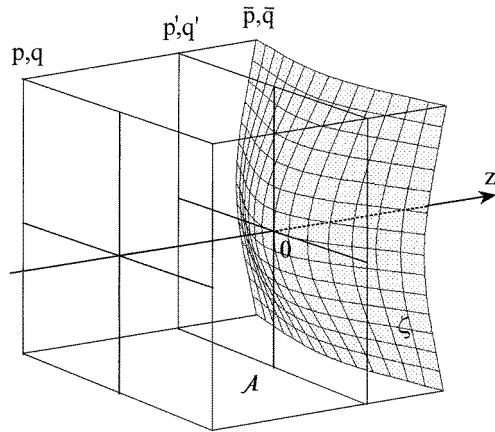
Figure 3. The imaging, aberrating system $\mathcal{A}$ maps the object phase space (p, q) on the flat image space (p', q'). The screen is warped to a surface $z = \zeta(q^2)$ in a homogeneous medium n; the root transformation maps (p', q') to $(\bar{q}, \bar{p})$.

$$= \mu \left(q - \left(4A_{2,0,0}p^2 + 2A_{1,1,0}p \cdot q + 2\left[A_{1,0,1} - \frac{\zeta_2}{2n} \right] q^2 \right) \right.$$
$$\left. \times p - \left(A_{1,1,0}p^2 + 2A_{0,2,0}p \cdot q + \left[A_{0,1,1} + \frac{\zeta_2}{n}\lambda\mu \right] q^2 \right) q \right). \tag{4.3}$$

The 'corrected system' (4.2) is thus characterized as the canonical map $\mathcal{C} = \mathcal{G}\{\mathbf{C}, \mathbf{N}\}$ where, from (2.3) and (3.1), the paraxial part is

$$\mathbf{N} = \mathbf{MP} = \begin{pmatrix} \mu & \lambda - 2n\zeta_2\mu \\ 0 & \mu^{-1} \end{pmatrix} \tag{4.4}$$

and the third-order aberration coefficients $\mathbf{C} = \{C_{k_+, k_0, k_-}\}_{k_+ + k_0 + k_- = 2}$ of $\mathcal{C}$ are found to be

$$C_{0,2,0} = A_{0,2,0} \qquad C_{1,1,0} = A_{1,1,0} \qquad C_{2,0,0} = A_{2,0,0} \tag{4.5a}$$

$$C_{0,1,1} = A_{0,1,1} + \frac{\zeta_2}{n}\lambda\mu \qquad C_{1,0,1} = A_{1,0,1} - \frac{\zeta_2}{2n} \tag{4.5b}$$

$$C_{0,0,2} \text{ undetermined.} \tag{4.5c}$$

Let us examine the possible aberration corrections. The first row (4.5a) informs us (from left to right) that the astigmatism (020), coma (110), and spherical aberration (200) of the paraxial-imaging-but-aberrating system $\mathcal{A}$ *cannot* be corrected by warping the screen because they do not contain any of its parameters ζ_{2k} in (2.2). The second row (4.5b) identifies two aberrations that *can* be controlled: third-order distortion (011) and curvature of field (101). As we argued before, the correction of distortion is unimportant for imaging. Thus, the curvature of field aberration coefficient $C_{1,0,1}$ can be set to zero on a paraboloidal screen with the value

$$\zeta_2 = 2n A_{1,0,1}. \tag{4.6}$$

The coefficient of distortion on the warped screen will then be $C_{0,1,1} = A_{0,1,1} + 2\lambda\mu A_{1,0,1}$, but this does not affect focusing. The last coefficient, $C_{0,0,2}$, is irrelevant for image formation and cannot be found from (4.3). To determine it we must also expand $\bar{p}^{[3]}$ in (p, q) and compare it with (3.8a). It is

$$C_{0,0,2} = A_{0,0,2} - \zeta_2 \frac{\lambda^2 \mu^2}{2n} + \zeta_4 n \mu^4. \tag{4.7}$$

On the paraboloidal warped screen (4.6), this aberration is $A_{0,0,2} - \lambda^2\mu^2 A_{1,0,1} + \zeta_4 n\mu^4$; it could be made to vanish by fixing the quartic parameter ζ_4 of the surface (2.2). Alternatively, since the value of $C_{0,0,2}$ is unimportant for imaging, the quartic shape of the surface may be used to eliminate some *fifth*-order aberration, as we proceed to analyse below.

5. Fifth- and higher-order corrections

Higher-order aberration expansions follow (2.7). The aberration polynomials A_k, $k = 2, 3, 4, \ldots$ (for aberration orders $3, 5, 7, \ldots, 2k - 1$; cf (2.8b) for $k = 2$) have $6, 10, 15, \ldots, \frac{1}{2}(k + 1)(k + 2)$, independent aberration coefficients. In fact, in [12] we showed that for surfaces tangent to the standard screen at the optical axis, in every rank k there hold $k + 2$ selection rules on the coefficients of the root transformation: $R_{k-\kappa,\kappa,0} = 0$ for $\kappa = 0, 1, \ldots, k$, and $R_{0,k-1,1} = 0$. The root transformation (2.1) for a surface (2.2) with warp parameters $\zeta_2, \zeta_4, \ldots, \zeta_{2k}$, composed to the same order, will yield results corresponding to (4.5)–(4.7) for all the aberration coefficients. We find that only *three* coefficients at each order (up to nine) can be set to zero individually by warping the screen, as detailed below.

Performing the steps indicated for rank $k = 3$, the aberration polynomial of the root transformation is found to be

$$R_3(\boldsymbol{p}, \boldsymbol{q}) = -\frac{\zeta_2}{8n^3}(p^2)^2 q^2 - \frac{\zeta_2^2}{2n^2} p^2 \boldsymbol{p} \cdot \boldsymbol{q} q^2 - \frac{\zeta_4}{2n} p^2 (q^2)^2 + 2\zeta_2\zeta_4 \boldsymbol{p} \cdot \boldsymbol{q}(q^2)^2 + \zeta_6 n(q^2)^3.$$
(5.1)

Among the 10 fifth-order aberration coefficients of $\mathcal{C} = \mathcal{A}\mathcal{R}$, we find that $C_{3,0,0}$, $C_{2,1,0}$, $C_{1,2,0}$ and $C_{0,3,0}$ are independent of the quartic warp coefficient ζ_4. They cannot be adjusted by quartic warping. Next, there are three coefficients which depend linearly on ζ_4; they are

$$C_{1,1,1} = [-2\lambda\mu A_{1,1,0}/n + \lambda/2\mu n^3 + 4\lambda^2\mu^2 A_{2,0,0}/n]\zeta_2 - \zeta_2^2/2n^2 - 8\mu^4 n\zeta_4 A_{2,0,0} + A_{1,1,1}$$
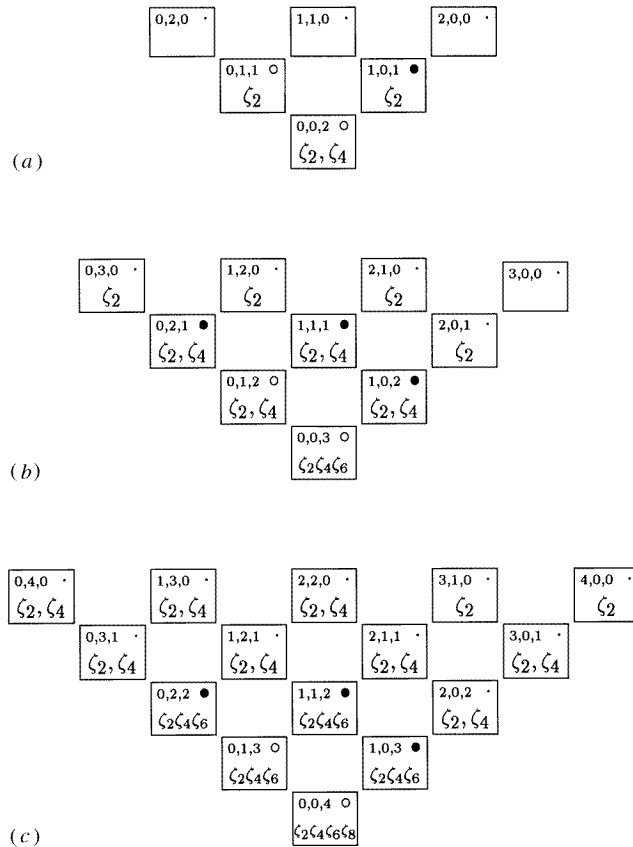(5.2a)

$$C_{1,0,2} = [\lambda^2\mu^2 A_{1,1,0}/n - \lambda^2/4n^3]\zeta_2 - [\mu^2/2n + 2\mu^4 n A_{1,1,0}]\zeta_4 + \lambda\mu\zeta_2^2/2n^2 + A_{1,0,2}$$
(5.2b)

$$C_{0,2,1} = [-2\lambda\mu A_{0,2,0}/n - 2\lambda\mu A_{1,0,1}/n + 2\lambda^2\mu^2 A_{1,1,0}/n - \lambda^2/2n^3 - A_{0,1,1}/n]\zeta_2$$
$$+ \lambda\mu\zeta_2^2/n^2 - 4\mu^4 n\zeta_4 A_{1,1,0} + A_{0,2,1}.$$
(5.2c)

As argued before, distortion $C_{0,1,2}$ and pocus $C_{0,0,3}$ contain ζ_4 (and the latter also contains the sextic surface parameter ζ_6), but are of no importance for image formation.

In figure 4 we show the pattern of corrections for third-, fifth- and seventh-order aberrations, displaying them in triangular arrays. For rank k (aberration order $2k - 1$) we indicate by a dot every aberration which *cannot* be corrected by surface warping at that order (because ζ_{2k-2} is absent in its expression); we indicate by a full circle those aberrations which *can* be corrected to that order because ζ_{2k-2} is present (as a *linear* factor in one summand); and by an open circle those aberrations whose correction is immaterial for focusing: distortion $(0, 1, k - 1)$ and pocus $(00k)$. The last is the only equation that contains the highest warping parameter ζ_{2k}; this can be used for correction at the next higher order.

Figure 4(a) organizes equations (4.5). The top row (astigmatism, coma, spherical aberration) cannot be corrected. The full circle marks the curvature of the field, which fixes ζ_2. For the fifth-order aberration, shown in figure 4(b), the top row of aberrations cannot be corrected, and neither can the aberration 201. The three coefficients (5.2) that *do* depend on ζ_4 can be individually put to zero: the 021, 111, or 102 boxes marked by full

(a)

0,2,0 ·	1,1,0 ·	2,0,0 ·

0,1,1 ○ ζ_2	1,0,1 ● ζ_2

0,0,2 ○ ζ_2, ζ_4

(b)

0,3,0 · ζ_2	1,2,0 · ζ_2	2,1,0 · ζ_2	3,0,0 ·

0,2,1 ● ζ_2, ζ_4	1,1,1 ● ζ_2, ζ_4	2,0,1 · ζ_2

0,1,2 ○ ζ_2, ζ_4	1,0,2 ● ζ_2, ζ_4

0,0,3 ○ $\zeta_2 \zeta_4 \zeta_6$

(c)

0,4,0 · ζ_2, ζ_4	1,3,0 · ζ_2, ζ_4	2,2,0 · ζ_2, ζ_4	3,1,0 · ζ_2	4,0,0 · ζ_2

0,3,1 · ζ_2, ζ_4	1,2,1 · ζ_2, ζ_4	2,1,1 · ζ_2, ζ_4	3,0,1 · ζ_2, ζ_4

0,2,2 ● $\zeta_2 \zeta_4 \zeta_6$	1,1,2 ● $\zeta_2 \zeta_4 \zeta_6$	2,0,2 · ζ_2, ζ_4

0,1,3 ○ $\zeta_2 \zeta_4 \zeta_6$	1,0,3 ● $\zeta_2 \zeta_4 \zeta_6$

0,0,4 ○ $\zeta_2 \zeta_4 \zeta_6 \zeta_8$

Figure 4. Aberrations that are non-correctable (·), controllable (●), and immaterial (○) for focusing on warped screens. Data for C_{k_+, k_0, k_-} are arranged on the three-axis plane $k_+ + k_0 + k_- = k$. (a) Third-order aberration (rank $k = 2$), (b) fifth-order aberration ($k = 3$), and (c) seventh-order aberration ($k = 4$). In each box we show the warp parameters ζ_{2k} on which the coefficient depends.

circles. The latter is the fifth-order curvature of the field. Finally, 012 and 003 (fifth-order distortion and pocus) do not unfocus the image.

A corresponding procedure for seventh-order aberrations requires symbolic computer algorithms such as have been developed by the authors [6]. We find a set of equations for the coefficients whose content is summarized in figure 4(c). We see there are again only three aberrations that can be controlled by fixing the surface polynomial coefficient ζ_6: 022, 112, and 103. All aberrations 'above' these cannot be altered, and those 'below' do not matter. The computation work has been pushed to ninth order by EJA, confirming the triangular pattern of controllable aberrations. We surmise that for generic aberration order $2k - 1$, the surface polynomial coefficient ζ_{2k-2} will appear linearly in the curvature of the field $(1, 0, k - 1)$, in $(1, 1, k - 2)$ and $(0, 2, k - 2)$ aberrations. Three coefficients depending linearly on a warp parameter allow a simple optimization algorithm if one provides a cost function (which can be the size and/or shape of the spot diagram for each aberration).

6. Conclusions

The study of images on warped screens can be made in phase space through the root transformation of reflection and refraction. It is a canonical transformation where the image space is warped in the direction of the evolution parameter. Such transformations have a clear meaning in geometric optics; insight into its properties may be important for classical mechanics, where the parameter is time.

The root map was previously used only as a computationally convenient intermediate step to find the refraction map. It now acquires reality as a *bona fide* optical element representing propagation to a warped screen. The problem addressed here is a prototype of a fully parametric design. The computational task is in fact simplified when we work only with the aberration coefficients, composing the generic aberrating system in paraxial focus with the root transformation whose coefficients are known. Our new results are that there are three correctable aberrations at orders 5, 7 and 9. There are more aberrations of order up to $2k - 1$ than available surface coefficients $\zeta_2, \zeta_4, \ldots, \zeta_{2k-2}$, but we are lead to a well defined linear optimization strategy for the three correctable aberrations at each order.

The correcting element here was propagation to a warped screen in a homogeneous medium. Other media may be of interest, such as axially symmetric graded-index elements with polynomial index profile. The Taylor-expanded refractive index $n(q^2)$ will take the place of the constant n in equations (3.2)–(3.4) thus providing more parameters for possible aberration correction. The case of elliptic index-profile 'fibres' was studied in [13] and the aberration polynomials of the root map in such media were given explicitly to seventh-order aberration.

References

[1] Dragt A J, Forest E and Wolf K B 1986 Foundations of a Lie algebraic theory of geometrical optics *Lie Methods in Optics (Lecture Notes in Physics 250)* (Heidelberg: Springer) ch 4, pp 105–58
[2] Arnold V I 1978 *Mathematical Methods of Classical Mechanics (Springer Graduate Texts in Mathematics, 60)* (New York: Springer) translated K Vogtmann and A Weinstein
     Guillemin V and Sternberg S 1984 *Symplectic Techniques in Physics* (Cambridge: Cambridge University Press)
     Libermann P and Marle C-M 1987 *Symplectic Geometry and Analytical Mechanics* (Dordrecht: Reidel)
     Sławianowski J 1991 *Geometry of Phase Space* (Warsaw: Wiley)
     Hofer H and Zehnder E 1994 *Symplectic Invariants and Hamiltonian Dynamics* (Basel: Birkhäuser)
[3] Simon R, Sudarshan E C G and Mukunda N 1984 Generalized rays in first-order optics: transformation properties of Gaussian Schell-model fields *Phys. Rev.* A **29** 3272–9
     Simon R, Sudarshan E C G and Mukunda N 1985 Anisotropic Gaussian Schell-model beams: passage through optical systems and associated invariants *Phys. Rev.* A **31** 2419–34
     Mukunda N, Simon R and Sudarshan E C G 1985 Fourier optics for the Maxwell field: formalism and applications *J. Opt. Soc. Am.* A **2** 416–26
     Simon R, Sudarshan E C G and Mukunda N 1987 Gaussian–Wigner distributions in quantum mechanics and optics *Phys. Rev.* A **36** 3868–80

Mukunda N, Sudarshan E C G and Simon R 1988 Families of Bose rays in quantum optics *Found. Physics* **18** 277–306

Simon R, Sudarshan E C G and Mukunda N 1988 Gaussian pure states in quantum mechanics and the symplectic group *Phys. Rev.* A **37** 3028–38

[4] Navarro-Saad M and Wolf K B 1986 Factorization of the phase-space transformation produced by an arbitrary refracting surface *J. Opt. Soc. Am.* A **3** 340–6

Wolf K B and Krötzsch G 1995 Geometry and dynamics in refracting systems *Eur. J. Phys.* **16** 14–20

[5] Buchdahl H A 1970 *An Introduction to Hamiltonian Optics* (Cambridge: Cambridge University Press)

Born M and Wolf E 1984 *Principles of Optics* 6th edn (Oxford: Pergamon)

[6] Wolf K B and Krötzsch G 1995 mexLIE 2. A set of symbolic computation functions for geometric aberration optics *Manual No 10* IIMAS–UNAM

[7] Delgado J (unpublished); see Wolf K B and Krötzsch G 1991 La transformación raíz de superficies refractantes y espejos *Revista Mexicana de Física* **37** 540–54

[8] Goldstein H 1950 *Classical Mechanics* (Reading, MA: Addison-Wesley) section 9.3

[9] Wolf K B 1988 Symmetry-adapted classification of aberrations *J. Opt. Soc. Am.* A **5** 1226–32

[10] Dragt A J 1982 Lie-algebraic theory of geometrical optics and optical aberrations *J. Opt. Soc. Am.* **72** 372–9

[11] Wolf K B and Krötzsch G 1991 Group-classified polynomials of phase space in higher-order aberration expansions *J. Symbolic Comput.* **12** 673–95

[12] Wolf K B 1986 Symmetry in Lie optics *Ann. Phys., NY* **172** 1–25

[13] Wolf K B 1991 Refracting surfaces between fibers *J. Opt. Soc. Am.* A **8** 1389–98