Hyperbolic reflections as fundamental building blocks for multilayer optics

Alberto G. Barriuso, Juan J. Monzón, and Luis L. Sánchez-Soto
Departamento de Óptica, Facultad de Física, Universidad Complutense, 28040 Madrid, Spain
José F. Cariñena
Departamento de Física Teórica, Facultad de Ciencias, Universidad de Zaragoza, 50009 Zaragoza, Spain

We reelaborate on the basic properties of lossless multilayers by using bilinear transformations. We study some interesting properties of the multilayer transfer function in the unit disk, showing that hyperbolic geometry turns out to be an essential tool for understanding multilayer action. We use a simple trace criterion to classify multilayers into three classes that represent rotations, translations, or parallel displacements. Moreover, we show that these three actions can be decomposed as a product of two reflections in hyperbolic lines. Therefore, we conclude that hyperbolic reflections can be considered as the basic pieces for a deeper understanding of multilayer optics.

I. INTRODUCTION

Although special relativity is perhaps the first theory that comes to mind when speaking about the interplay between physics and geometry, one cannot ignore that geometrical ideas are essential tools in the development of many branches of modern physics. The optics of layered media is not an exception: in recent years many concepts of geometrical nature have been introduced to gain further insights into the behavior of multilayers. The algebraic basis for these developments is the fact that the transfer matrix associated with a lossless multilayer is an element of the group SU(1, 1), which is locally isomorphic to the (2 + 1)-dimensional Lorentz group SO(2, 1). This leads to a natural and complete identification between reflection and transmission coefficients and the parameters of the corresponding Lorentz transformation.

As soon as one realizes that SU(1, 1) is also the basic group of the hyperbolic geometry, it is tempting to look for an enriching geometrical interpretation of the multilayer optics. Accordingly, we have recently proposed to view the action of any lossless multilayer as a bilinear (or Möbius) transformation on the unit disk, obtained by stereographic projection of the unit hyperboloid of SO(2, 1). This kind of representation has been previously discussed for, e.g., the Poincaré sphere in polarization optics, for Gaussian beam propagation, in laser mode-locking and optical pulse transmission, and also in modelling visual processing.

The point we wish to emphasize is that these bilinear transformations preserve hyperbolic distance between points on the unit disk. In Euclidean geometry any transformation of the plane that preserves distance can be written as a composition of reflections, which can be then considered as the most basic transformations. In fact, the composition of two reflections in straight lines is a rotation, or a translation, according these lines are intersecting or parallel.

In hyperbolic geometry, each circle orthogonal to the boundary of the unit disk is a hyperbolic line and reflections appear as inversions. However, we have an essential difference with the Euclidean case because there are three different kind of lines: intersecting, parallel, and ultra-parallel (which are neither intersecting nor parallel). In consequence, the composition of two reflections in hyperbolic lines is now a rotation, a parallel displacement, or a translation: these are precisely the transformations of the unit disk that preserve distance.

A powerful way of characterizing transformations is through the study of the points that they leave invariant. For example, in Euclidean geometry a rotation can be characterized by having only one fixed point, while a translation has no invariant point. For a reflection the fixed points consist of all the points of a line (the reflection axis).

In this paper we shall consider the fixed points of the bilinear transformation induced by the multilayer, showing that they can be classified according to the trace of the multilayer matrix. From this viewpoint, the three transformations mentioned above; namely, rotations, parallel displacements, and translations appear linked to the fact that the trace of the multilayer transfer matrix has a magnitude lesser, equal or greater than 2.

Since reflections appear as the basic building blocks of these geometric motions, we show that any multilayer action can be decomposed in terms of two inversions whose meaning is investigated. Such a decomposition is worked out for practical examples. This shows the power of the method and, at the same time, allows for a deeper understanding of layered media.

II. MULTILAYERS AND THE UNIT DISK

We first briefly summarize the essential ingredients of multilayer optics we shall need for our purposes. We deal with a stratified structure, illustrated in Fig. 1, that consists of a stack of \(1, \ldots, j, \ldots, m\), plane-parallel layers sandwiched between two semi-infinite ambient \((a)\) and substrate \((s)\) media, which we shall assume to be iden-
The matrix $M_{as}$ can be shown to be

$$M_{as} = \begin{bmatrix} 1/T_{as} & R_{as}/T_{as} \\ R_{as}/T_{as} & 1/T_{as} \end{bmatrix} \equiv \begin{bmatrix} \alpha & \beta \\ \beta^* & \alpha^* \end{bmatrix}, \quad (3)$$

where the complex numbers $R_{as}$ and $T_{as}$ are, respectively, the overall reflection and transmission coefficients for a wave incident from the ambient. Because $|R_{as}|^2 + |T_{as}|^2 = 1$, we have the additional condition $|\alpha|^2 - |\beta|^2 = 1$ or, equivalently, $\det M_{as} = +1$ and then the set of lossless multilayer matrices reduces to the group SU(1, 1), whose elements depend on three independent real parameters.

The identity matrix corresponds to $T_{as} = 1$ and $R_{as} = 0$, so it represents an antireflection system. The matrix that describes the overall system obtained by putting two multilayers together is the product of the matrices representing each one of them, taken in the appropriate order. So, two multilayers, which are inverse, when composed give an antireflection system [14].

In Refs. [2] and [3] we have proposed to view the multilayer action in a relativisticlike framework. Without going into details, it is convenient to characterize the state of the fields at each side of the multilayer by means of the “space-time” coordinates

$$e^0 = \frac{1}{2} |E^{(+)}|^2 + |E^{(-)}|^2,$$

$$e^1 = \text{Re}[E^{(+)*}E^{(-)}],$$

$$e^2 = \text{Im}[E^{(+)*}E^{(-)}],$$

$$e^3 = \frac{1}{2} |E^{(+)}|^2 - |E^{(-)}|^2, \quad (4)$$

for both ambient and substrate media. The coordinate $e^3$ is the semi-difference of the fluxes (note that this number can take any real value) and, therefore, is constant because the multilayer is lossless. In consequence, we have that

$$(e^0)^2 - (e^1)^2 - (e^2)^2 = (e^3)^2 = \text{constant}. \quad \text{(5)}$$

Equation (5) defines a two-sheeted hyperboloid of radius $e^3$, which without loss of generality will be taken henceforth as unity [13].

A simple calculation shows that if one uses stereographic projection taking the south pole ($-1, 0, 0$) as projection center (see Fig. 2), the projection of the point $(e^0, e^1, e^2)$ becomes in the complex plane

$$z = \frac{e^1 + ie^2}{1 + e^0} = \frac{E^{(-)}}{E^{(+)}}. \quad \text{(6)}$$

The upper sheet of the unit hyperboloid is projected into the unit disk, the lower sheet into the external region, while the infinity goes to the boundary of the unit disk.

The geodesics in the hyperboloid are intersections with the hyperboloid of planes passing through the origin. Consequently, hyperbolic lines are obtained from these

\begin{figure}[h]
\centering
\includegraphics[width=0.5\textwidth]{fig1.png}
\caption{Wave vectors of the input $[E_a^{(+)}]$ and output $[E_a^{(-)}]$ fields in a multilayer sandwiched between two identical semi-infinite ambient and substrate media.}
\end{figure}
by stereographic projection and they correspond to circle arcs that orthogonally cut the boundary of the unit disk.

It seems natural to consider the complex variables in Eq. (6) for both ambient and substrate. In consequence, Eq. (6) defines a transformation on the complex plane $\mathbb{C}$, mapping the point $z_s$ into the point $z_a$ according to

$$z_a = \Phi[M_{as}, z_s] = \frac{\beta^* + \alpha^* z_s}{\alpha + \beta z_s}, \quad (7)$$

which is a bilinear (or Möbius) transformation. The action of the multilayer can be seen as a function $z_a = f(z_s)$ that can be appropriately called the multilayer transfer function $\Phi$. The action of the inverse matrix $M_{as}^{-1}$ is $z_s = \Phi[M_{as}^{-1}, z_a]$. One can show that the unit disk, the external region and the boundary remain invariant under the multilayer action.

For later purposes, we need the concept of distance in the unit disk. To this end, it is customary to define the cross ratio of four distinct points $z_A, z_B, z_C,$ and $z_D$ as the number

$$(z_A, z_B; z_C, z_D) = \frac{(z_A - z_C)(z_B - z_d)}{(z_A - z_d)(z_B - z_C)}, \quad (8)$$

which is real only when the four points lie on a circle or a straight line. In fact, bilinear transformations preserve this cross ratio.

Let now $z$ and $z'$ be two points that are joined by the hyperbolic line whose endpoints on the unit circle are $E$ and $E'$. The hyperbolic distance between $z$ and $z'$ is defined as

$$d_H(z, z') = \frac{1}{2} |\ln(E, E'| z, z')|, \quad (9)$$

This can be seen as arising from the usual Minkowski distance in the unit hyperboloid (obtained through geodesics) by stereographic projection. The essential point for our purposes here is that bilinear transformations are isometries; i.e., they preserve this distance.

### III. Trace Criterion for the Classification of Multilayers

Bilinear transformations constitute an important tool in many branches of physics. For example, in polarization optics they have been employed for a simple classification of polarizing devices by means of the concept of eigenpolarizations of the transfer function $\Phi$. The equivalent concept in multilayer optics can be stated as the field configurations such that $z_a = z_s \equiv z_f$ in Eq. (7), that is

$$z_f = \Phi[M_{as}, z_f], \quad (10)$$

whose solutions are

$$z_f = \frac{1}{2\beta} \left\{ -2i \Im(\alpha) \pm \sqrt{\text{Tr}(M_{as})^2 - 4} \right\}. \quad (11)$$

These values $z_f$ are known as fixed points of the transformation $\Phi$. The trace of $M_{as}$ provides then a suitable tool for the classification of multilayers.

When $|\text{Tr}(M_{as})|^2 < 4$ the multilayer action is elliptic and it has only one fixed point inside the unit disk, while the other lies outside. Since in the Euclidean geometry a rotation is characterized for having only one invariant point, this multilayer action can be appropriately called a hyperbolic rotation.

When $|\text{Tr}(M_{as})|^2 > 4$ the multilayer action is hyperbolic and it has two fixed points both on the boundary of the unit disk. The hyperbolic line joining these two fixed points remains invariant and thus, by analogy with the Euclidean case, this action will be called a hyperbolic translation.

Finally, when $|\text{Tr}(M_{as})|^2 = 4$ the multilayer action is parabolic and it has only one (double) fixed point on the boundary of the unit disk. This action will be called a parallel displacement.

To proceed further let us note that by taking the conjugate of $M_{as}$ with any matrix $C \in SU(1, 1)$; i.e.,

$$\hat{M}_{as} = C M_{as} C^{-1}, \quad (12)$$

we obtain another matrix of the same type, since $\text{Tr}(\hat{M}_{as}) = \text{Tr}(M_{as})$. Conversely, if two multilayer matrices have the same trace, one can always find a matrix $C$ satisfying Eq. (12).

The fixed points of $\hat{M}_{as}$ are then the image by $C$ of the fixed points of $M_{as}$. In consequence, given any multilayer...
matrix $M_{as}$ one can always reduce it to a $\tilde{M}_{as}$ with one of the following canonical forms:

$$
\tilde{K}(\varphi) = \begin{bmatrix} \exp(i\varphi/2) & 0 \\ 0 & \exp(-i\varphi/2) \end{bmatrix},
$$

$$
\tilde{A}(\chi) = \begin{bmatrix} \cosh(\chi/2) & i \sinh(\chi/2) \\ -i \sinh(\chi/2) & \cosh(\chi/2) \end{bmatrix},
$$

$$
\tilde{N}(\eta) = \begin{bmatrix} 1 - i \eta/2 & \eta/2 \\ \eta/2 & 1 + i \eta/2 \end{bmatrix},
$$

that have as fixed points the origin (elliptic), $+i$ and $-i$ (hyperbolic) and $+i$ (parabolic), and whose physical significance has been studied before $[10]$. The explicit construction of the family of matrices $\tilde{C}$ is easy: it suffices to impose that $\tilde{C}$ transforms the fixed points of $M_{as}$ into the ones of $\tilde{K}(\varphi)$, $\tilde{A}(\chi)$, or $\tilde{N}(\eta)$.

The concept of orbit is especially appropriate for obtaining an intuitive picture of these actions. We recall that given a point $z$, its orbit is the set of points $z'$ obtained from $z$ by the action of all the elements of the group. In Fig. 3a we have plotted typical orbits for each one of the canonical forms $\tilde{K}(\varphi)$, $\tilde{A}(\chi)$, and $\tilde{N}(\eta)$. For matrices $\tilde{K}(\varphi)$ the orbits are circumferences centered at the origin and there are no invariant hyperbolic lines. For $\tilde{A}(\chi)$, they are arcs of circumference going from the point $+i$ to the point $-i$ through $z$ and they are known as hypercycles. Every hypercycle is equidistant [in the sense of the distance $[11]$] from the imaginary axis, which remains invariant (in the Euclidean plane the locus of a point at a constant distance from a fixed line is a pair of parallel lines). Finally, for $\tilde{N}(\eta)$ the orbits are circumferences passing through the point $+i$ and joining the points $z$ and $-z^*$ and they are known as horocycles: they can be viewed as the locus of a point that is derived from the point $+i$ by a continuous parallel displacement $[4]$.

For a general matrix $M_{as}$ the corresponding orbits can be obtained by transforming with the appropriate matrix $\tilde{C}$ the orbits described before. In Fig. 3b we have plotted typical examples of such orbits for elliptic, hyperbolic, and parabolic actions. We stress that once the fixed points of the multilayer matrix are known, one can ensure that $z_a$ will lie in the orbit associated to $z_a$.

In the Euclidean plane any isometry is either a rotation, a translation, or a reflection. In any case, reflections are the ultimate building blocks, since any isometry can be expressed as the composition of reflections. In this Euclidean plane two distinct lines are either intersecting or parallel. Accordingly, the composition of two reflections in two intersecting lines forming an angle $\varphi$ is a rotation of angle $2\varphi$ while the composition of two reflections in two parallel lines separated a distance $d$ is a translation of value $2d$.

However, in the hyperbolic geometry induced in the unit disk, any two distinct lines are either intersecting (they cross in a point inside the unit disk), parallel (they meet at infinity; i.e., at a point on the boundary of the unit disk), or ultraparallel (they have no common points). A natural question arises: what is the composition of reflections in these three different kind of lines? To some extent, the answer could be expected: the composition is a rotation, a parallel displacement, or a translation, respectively. However, to gain further insights one needs to know how to deal with reflections in the unit disk. This is precisely the goal of next Section.

### IV. Reflections in the Unit Disk

In the Euclidean plane given any straight line and a point $P$ which does not lie on the line, its reflected image $P'$ is such that the line is equidistant from $P$ and $P'$. In other words, a reflection is a special kind of isometry in which the invariant points consist of all the points on the line.

The concept of hyperbolic reflection is completely analogous: given the hyperbolic line $\Gamma$ and a point $P$, to obtain its reflected image $P'$ in $\Gamma$ we must drop a hyperbolic line $\Gamma'$ from $P$ perpendicular to $\Gamma$ (such a hyperbolic line exists and it is unique) and extending an equal hyperbolic distance [according to $[12]$] on the opposite side of $\Gamma$ from $P$. In the unit disk, this corresponds precisely to an inversion.

To maintain this paper as self-contained as possible, let us first recall some facts about the concept of inversion. Let $C$ be a circle with center $w$ and radius $R$. An inversion on the circle $C$ maps the point $z$ into the point $z'$ along the same radius in such a way that the product of distances from the center $w$ satisfies

$$
|z' - w| = |z - w| = R^2,
$$

and hence one immediately gets

$$
z' = w + \frac{R^2}{z^* - w^*} = \frac{R^2 + wz^* - w^*w}{z^* - w^*}. \tag{15}
$$

If the circle $C$ is a hyperbolic line, it is orthogonal to the boundary of the unit disk and fulfills $ww^* = R^2 + 1$. In consequence

$$
z' = \frac{wz^* - 1}{z^* - w^*}. \tag{16}
$$

![Figure 3: Plot of typical orbits in the unit disk for: (a) canonical transfer matrices as given in Eq. (13) and (b) arbitrary transfer matrices.](image-url)
One can check that inversion maps circles and lines into circles and lines, and transforms angles into equal angles (although reversing the orientation). If a circle \( C' \) passes through the points \( P \) and \( P' \), inverse of \( P \) in the circle \( C \), then \( C \) and \( C' \) are perpendicular. Moreover, the hyperbolic distance is invariant under inversions. This confirms that inversions are indeed reflections and so they appear as the most basic isometries of the unit disk.

It will prove useful to introduce the conjugate bilinear transformation associated with a matrix \( M_{\alpha s} \) as [compare with Eq. (1)]

\[
z_a = \Phi^*[M_{\alpha s}, z_{s}] = \frac{\beta^* + \alpha^* z_{s}^*}{\alpha + \beta z_{s}^*}.
\]

With this notation we can recast Eq. (10) as

\[
z' = \Phi^*[l_w, z],
\]

where the matrix \( l_w \in SU(1, 1) \) associated to the inversion is

\[
l_w = \begin{bmatrix} -i w^*/R & i/R \\ -i/R & i w/R \end{bmatrix}.
\]

The composition law for inversions can be stated as follows: if \( z' = \Phi^*[l_w, z] \) and \( z'' = \Phi^*[l_w', z'] \) then

\[
z'' = \Phi[l_w l_w', z].
\]

To shed light on the physical meaning of the inversion, assume that incoming and outgoing fields are interchanged in the basic configuration shown in Fig. 1. In our case, this is tantamount to reversing the time arrow. It is well known that given a forward-traveling field \( E^+(+) \), the conjugate field \( E^+(+)^* \) represents a backward phase-conjugate wave of the original field [21]. In other words, the time-reversal operation can be viewed in this context as the transformation

\[
z \mapsto \frac{1}{z^*},
\]

for both ambient and substrate variables; that is, it can be represented by an inversion in the unit circle. The transformed points lie outside the unit circle because, according to Eq. (4), this time reversal transforms the upper sheet into the lower sheet of the hyperboloid.

Moreover, by direct inspection it is easy to convince oneself that the matrix relating these time-reversed fields is precisely \( M_{\alpha s}^* \) and so the action can be put as

\[
(1/z_a)^* = \frac{\beta^* + \alpha^* (1/z_a)^*}{\alpha + \beta (1/z_a)^*},
\]

which expresses a general property of the time-reversal invariance in our model.
FIG. 5: Decomposition of the multilayer action in terms of two reflections in two intersecting lines for the same multilayer as in Fig. 4 with $\delta_2 = 3$ rad (elliptic case).

FIG. 6: Decomposition of the multilayer action in terms of two reflections in two ultraparallel lines for the same multilayer as in Fig. 4 with $\delta_2 = 1$ rad (hyperbolic case).

There are no invariant points in the unit disk, but the hyperbolic line joining $z_{f1}$ and $z_{f2}$ is the axis of the hyperbolic translation. In Fig. 6, we have plotted the horocycle passing through $z_a$ and $z_s$. The multilayer action can be now interpreted as the composition of two reflections in two ultraparallel hyperbolic lines $\Gamma_1$ and $\Gamma_2$ orthogonal to the translation axis. If $\Gamma_1$ and $\Gamma_2$ intersect the horocycle at the points $z_1$ and $z_2$, they must fulfill

$$d_H(z_a, z_s) = 2d_H(z_1, z_2), \quad (23)$$

in complete analogy with what happens in the Euclidean plane. Once again, there are infinity pairs of lines fulfilling this condition.

Finally, we take $\delta_2 = 0.4328$ rad, so we are in the parabolic case ($|\text{Tr}(M_{m_2})|^2 = 4$), and $z_a = -0.1615 + 0.6220i$. The (double) fixed point is $z_f = -1$. In Fig. 7 we have plotted the horocycle connecting $z_a$ and $z_s$ and the fixed point. Now, we have the composition of two reflections in two parallel lines $\Gamma_1$ and $\Gamma_2$ that intersect at the fixed point $z_f$ and with the same constraint as before.

VI. CONCLUDING REMARKS

In this paper, we have proved a geometric scenario to deal with multilayer optics. More specifically, we have reduced the action any lossless multilayer (no matter how complicated it might be) to a rotation, a parallel displacement or a translation, according to the magnitude of its trace. These are the basic isometries of the unit disk and we have expressed them as the composition of two reflections in intersecting, ultraparallel, or parallel lines. There is no subsequent factorization in simpler terms so, reflections are the most basic motions one can find in the unit disk.

We hope that this approach will complement the more standard algebraic method in terms of transfer matrices, and together they will aid to obtain a better physical and geometrical feeling for the properties of multilayers.

Finally, we stress that the benefit of this formulation lies not in any inherent advantage in terms of efficiency in solving problems in layered structures. Rather, we expect that the formalism presented here could provide a general and unifying tool to analyze multilayer performance in a way closely related to other fields of physics, which seems to be more than a curiosity.
Acknowledgments

We thank J. Zoido for his help in computing some of the figures of this paper.

[1] B. F. Schutz, *Geometrical methods of Mathematical Physics* (Cambridge University Press, Cambridge, 1997).
[2] J. J. Monzón and L. L. Sánchez-Soto, “Lossless multilayers and Lorentz transformations: more than an analogy,” Opt. Commun. 162, 1-6 (1999).
[3] J. J. Monzón and L. L. Sánchez-Soto, “Fully relativistic formulation of multilayer optics,” J. Opt. Soc. Am. A 16, 2013-2018 (1999).
[4] H. S. M. Coxeter, *Introduction to Geometry* (Wiley, New York, 1969).
[5] T. Yonte, J. J. Monzón, L. L. Sánchez-Soto, J. F. Cariñena, and C. López-Lacasta, “Understanding multilayers from a geometrical viewpoint,” J. Opt. Soc. Am. A 19, 603-609 (2002).
[6] J. J. Monzón, T. Yonte, L. L. Sánchez-Soto, and J. F. Cariñena, “Geometrical setting for the classification of multilayers,” J. Opt. Soc. Am. A 19, 985-991 (2002).
[7] R. M. A. Azzam and N. M. Bashara, *Ellipsometry and Polarized Light* (North-Holland, Amsterdam, 1987).
[8] D. Han, Y. S. Kim, and M. E. Noz, “Polarization optics and bilinear representations of the Lorentz group,” Phys. Lett. A 219, 26-32 (1996).
[9] H. Kogelnik, “Imaging of optical modes – resonators with internal lenses,” Bell Syst. Techn. J. 44, 455-494 (1965).
[10] M. Nakazawa, J. H. Kubota, A. Sahara, and K. Tamura, “Time-domain ABCD matrix formalism for laser mode-locking and optical pulse transmission,” IEEE J. Quant. Electron. QE34, 1075-1081 (1998).
[11] R. Melter, A. Rosenfeld, and P. Bhattacharya, *Vision Geometry* (American Math. Soc., Providence, 1991).
[12] K. A. Dunn, “Poincaré group as reflections in straight lines,” Am. J. Phys. 49, 52-55 (1981).
[13] When ambient (0) and substrate (m + 1) media are different, the angles $\theta_0$ and $\theta_{m+1}$ are connected by Snell law $n_0 \sin \theta_0 = n_{m+1} \sin \theta_{m+1}$, where $n_j$ denotes the refractive index of the jth medium.
[14] J. J. Monzón and L. L. Sánchez-Soto, “Origin of the Thomas rotation that arises in lossless multilayers,” J. Opt. Soc. Am. A 16, 2786-2792 (1999).
[15] J. J. Monzón and L. L. Sánchez-Soto, “A simple optical demonstration of geometric phases from multilayer stacks: the Wigner angle as an anholonomy,” J. Mod. Opt. 48, 21-34 (2001).
[16] D. Pedoe, *A course of Geometry* (Cambridge University Press, Cambridge, 1970).
[17] A. Mischenko and A. Fomenko, *A Course of Differential Geometry and Topology* (MIR, Moscow, 1988), Sec. 1.4.
[18] L. L. Sánchez-Soto, J. J. Monzón, T. Yonte, and J. F. Cariñena, “Simple trace criterion for classification of multilayers,” Opt. Lett. 26, 1400-1402 (2001).
[19] J. J. Monzón, T. Yonte, and L. L. Sánchez-Soto, “Basic factorization for multilayers,” Opt. Lett. 26, 370-372 (2001).
[20] B. Ya. Zel’dovich, N. F. Pilipetsky, and V. V. Shkunov, *Principles of Phase Conjugation* (Springer-Verlag, Berlin, 1985).