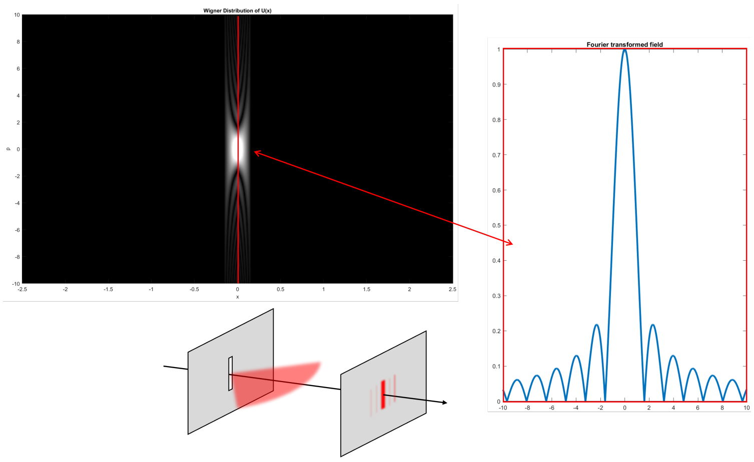
Time-Frequency Methods and Optics: A Ray and Wave Analysis

Wen Perng

Department of Electrical Engineering, National Taiwan University, Taipei, Taiwan
Email: b10901042@ntu.edu.tw



Abstract – Optics is often divided into separate fields of ray and wave optics, seemingly mutually exclusive. This text provides insights and linkages between the two viewpoints of optics by utilizing the methods of time-frequency analysis. The methods of linear canonical transformation, Fourier transform, fractional Fourier transform and Wigner distribution function were introduced, linking up relations between disparate viewpoints of matrix optics, diffraction formulas and phase space representation of light.

28 December 2023

1. INTRODUCTION

From Fermat's principle of least action and Newton's corpuscular theory of light, to Maxwell's equations and Hertz' experiment, optics has always been a wonder of discoveries in the course of human scientific advancements. Many different theories and techniques were developed to handle optical problems at different scales and from different point of views. To name a few, the geometric optics viewed light as rays, traveling in straight lines through space, and are suitable to be described by matrices in the paraxial regions. The wave optics considers the polarization, amplitude and phase of light waves, discussing the diffraction and interference effects of light. The corpuscular theory of light, on the other hand, view light as particles following the Newton's law of motion. Though this point of view might seem wrong in the modern notion of light, it is highly correlated with the Fermat's principle and Hamiltonian optics. All these point of views are seemingly disparate and having few intersections, but we shall see that they are in fact linked via methods of time-frequency analysis.

The time-frequency analysis as a family of integral transforms builds upon the foundation of Fourier transform, providing deeper insights into signal processing problems. Methods including linear canonical transformation fractional Fourier transform, Wigner distribution function and etc. Amazingly, the development of time-frequency analysis has always been entangled with that of optics. In fact, transformations of light by optical systems are equivalent to the integral transforms above. By treating optical fields as signals, these transformations by optical systems are equivalent to the time-frequency analyses.

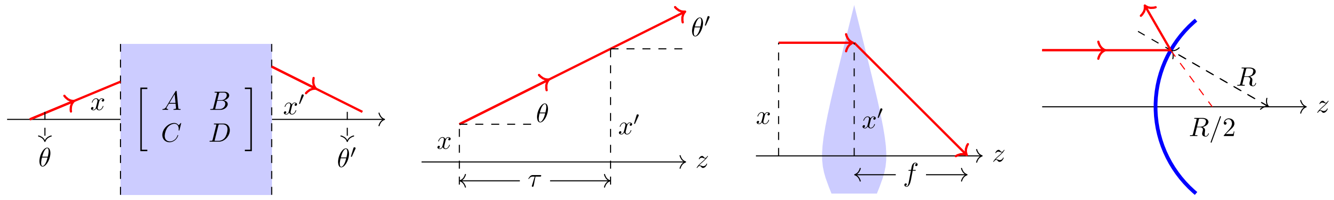
This text provides an overview of the different linkages between optics and time-frequency analysis. And in turn, using the ideas in time-frequency analysis to unite the many different perspectives on optics.

1.1. Sections Overview

- To ensure the readers are acquainted with the following text, section 2 goes over some basic knowledge of optics including matrix optics, light fields, diffraction formulas, and phase space representation of light. Only the essentials are provided in this section, and many of the derivations are glossed over.
- section 3 discusses the applications of the linear canonical transformation in optics. This section relates the transformation of fields in wave optics to the ray transfer matrices of ray optics by utilizing the linear canonical transformation. Some important aspects on the notation and units of the transform is also discussed. The derivation of the linear canonical transform is provided in Appendix A.
- section 4 introduces Fourier transform and the fractional Fourier transform as special cases of the linear canonical transformation. They are closely related to diffraction after refraction by a thin lens, and is useful in the study of optical signal processing called Fourier optics.
- section 5 ties up all the loose ends introduced in the previous sections by introducing the Wigner distribution function. It combines the ray and wave point of view of light by the phase space representation. Using this method, a generalization to the usual linear canonical transformation is also given, with its motives and derivation provided in detail in Appendix B. By utilizing its time averaging definition, it can handle both coherent and incoherent light sources, while the projection property allows a new perspective on interference of light.

1.2. Notations

- $\boldsymbol{x}$: bold fonts for vectors.
- $|\boldsymbol{x}|$: vector 2-norm.
- $(\cdot)^{\mathsf{T}}$: transposition; $(\cdot)^{-\mathsf{T}} = \left((\cdot)^{-1}\right)^{\mathsf{T}} = \left((\cdot)^{\mathsf{T}}\right)^{-1}$.
- $(\cdot)^*$: complex conjugation.
- $\mathbb{R}$: the reals; $\mathbb{R}_{\geq 0}$: non-negative reals; $\mathbb{C}$: the complex numbers.
- M: serif fonts especially for matrices.
- $\mathcal{O}$, $\mathcal{Q}$, $\mathcal{P}$, $\mathcal{R}$, $\mathcal{W}$, $\mathscr{F}$, $\mathscr{L}$: calligraphic and script fonts for operators.
- $\mathcal{O}^n$: applying operator $\mathcal{O}$ on the same variable n times.
- $\mathcal{O}^{\otimes n}$: applying the operator $\mathcal{O}$ on n different variables, i.e. the tensor product of operators.
- $\delta(\cdot)$: the delta function/distribution.



(a) A General ABCD System (b) Transmission (c) Refraction by Lens (d) Reflection by Mirror

FIGURE 1: Paraxial Optics

2. OPTICS OVERVIEW

When learning optics, it is more often than not divided into two main categories: ray optics and wave optics, the former views light as individual rays traveling through space and the latter views light as a wave propagating with amplitude and phase, interfering with each other. This section provides an overview on the two point of views on optics, discussing topics regarding matrix optics, light fields, diffraction, phase space and coherence.

2.1. Matrix Optics

From what we've learnt from high school, we know that light, when having its energy culminated in a small region of space, can be viewed as a ray traveling in a straight line, changing its direction under reflection or refraction. This is the simplest notion of what light, or to be precise, light ray, is. A light ray locates at a point $\boldsymbol{x} \in \mathbb{R}^3$ in space with a vector direction[1] $\boldsymbol{\theta} \in S^2$ pointing forwards. This inspires a suitable description of ray optics using the formalism called *matrix optics* or *ABCD optics*, where rays of lights are depicted using a vector and an optical system is described by a square matrix acting on the rays.

The matrix optics is the easiest to understand and is crucial for many of the analysis below. We shall first discuss the two-dimensional case, then extend our discussion to the three-dimensional case. Throughout our discussions, we shall consider the whole optical system be immersed in empty space with refractive index of $n = 1$ unless mentioned.

2.1.1. Two-Dimensional Ray Optics

For a two-dimensional paraxial optical system, we consider light rays that are close to the horizontal optical axis. As shown in subfigure 1b, a paraxial light ray with an inclination of $\tan\theta$ away from the optical axis (z axis) can be approximated by

$$\theta \approx \sin\theta \approx \tan\theta.$$

This is the paraxial approximation. We represent a light ray at a point by its height x of the point and the inclination θ of the ray with respect to the optical axis times the refractive index n, often written as a column vector:

$$\begin{bmatrix} x \\ n\theta \end{bmatrix} \overset{n=1}{\equiv} \begin{bmatrix} x \\ \theta \end{bmatrix}. \tag{1}$$

When the light ray passes through a system consisting of combinations of transmissions, reflections and refractions, the effect of the system on the ray can be approximated to the first order, giving us a linear transformation, this is the so-called *first-order optics*. Under the first-order optics, an optical system can be represented by a 2×2 matrix with real entries A, B, C and D, written as below:

$$M = \begin{bmatrix} A & B \\ C & D \end{bmatrix}. \tag{2}$$

This is called the ABCD matrix of the system or the system matrix, requiring that $AD - BC = 1$. With an input ray of $[x, \theta]^\mathsf{T}$, this system results in a transformation with output ray $[x', \theta']^\mathsf{T}$ following the transformation law below (see subfigure 1a):

$$\begin{bmatrix} y' \\ \theta' \end{bmatrix} = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} y \\ \theta \end{bmatrix}. \tag{3}$$

Not that A and D are unitless, while B has units of length, and C has units of 1 over length. For multiple consecutive optical systems with system matrices $\{M_i\}_{i=1}^n$ linking up in series, they form a new system with the system matrix being:

$$M = M_n M_{n-1} \cdots M_2 M_1. \tag{4}$$

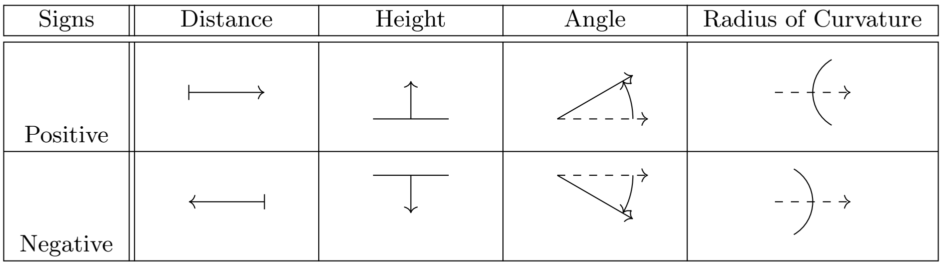


TABLE 1: Signature of Parameters in Paraxial Optics: assume ray propagates from left to right.

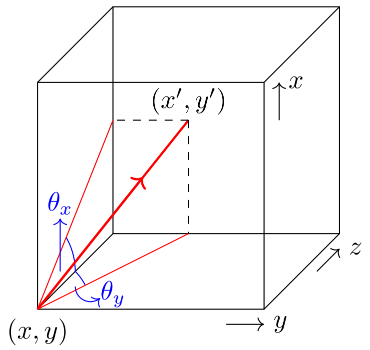


FIGURE 2: Ray Vectors in 3D

Consider the parameter convention as listed in Table 1, three basic optical systems to discuss include the transmission of light ray by a distance τ (subfigure 1b), refraction by a thin lens of focal length f (subfigure 1c), and reflection by a spherical lens of radius R (subfigure 1d). The first has a system matrix of the form:

$$\begin{bmatrix} 1 & \tau \\ 0 & 1 \end{bmatrix}; \tag{5}$$

the second has a system matrix of the form:

$$\begin{bmatrix} 1 & 0 \\ -\phi & 1 \end{bmatrix}, \tag{6}$$

where $\phi = 1/f$ is the *(optical) power* of the lens. The third considers reflection of rays by a concave spherical mirror of radius $R > 0$ as in subfigure 1d (the case for convex mirrors $R < 0$ follow similar notions), this inverts the propagation direction of the light. The reflection can be described by the matrix:

$$\begin{bmatrix} 1 & 0 \\ -\frac{2}{R} & 1 \end{bmatrix}. \tag{7}$$

Note that by comparing eqns.(6) and (7), we can see that a spherical mirror has focal length of $R/2$.

2.1.2. Three-Dimensional Ray Optics

The same idea applies for three-dimensional paraxial optics. Consider the optical axis as the z axis, then an arbitrary ray at a constant z can be represented by a 4×1 column vector

$$\begin{bmatrix} \boldsymbol{x} \\ n\boldsymbol{\theta} \end{bmatrix} \overset{n=1}{\equiv} \begin{bmatrix} \boldsymbol{x} \\ \boldsymbol{\theta} \end{bmatrix} = \begin{bmatrix} x \\ y \\ \theta_x \\ \theta_y \end{bmatrix}, \tag{8}$$

[1] S^2 is the two-sphere embedded in $\mathbb{R}^3$, coordinates other than $\boldsymbol{\theta}$ can also be used as long as a one-to-one correspondence exists.

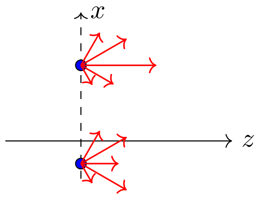


FIGURE 3: Light Field

where the variables are defined as in Figure 2. A paraxial optical system in three-dimensional space can hence be described by a 4×4 real matrix of the form

$$\left[\begin{array}{cc:cc} A_1 & A_2 & B_1 & B_2 \\ A_3 & A_4 & B_3 & B_4 \\ \hdashline C_1 & C_2 & D_1 & D_2 \\ C_3 & C_4 & D_3 & D_4 \end{array}\right] \equiv \left[\begin{array}{c:c} \mathsf{A} & \mathsf{B} \\ \hdashline \mathsf{C} & \mathsf{D} \end{array}\right] \tag{9}$$

acting on the ray vectors, requiring that $AD^\mathsf{T} - BC^\mathsf{T} = \mathbb{1}$, the 2×2 identity matrix. A *symmetric* optical system is one where under rotations about the z axis, the ABCD matrix doesn't change its form, viz., the blocks in eqn.(9) are all diagonal matrices: $\mathsf{A} = A\mathbb{1}$, $\mathsf{B} = B\mathbb{1}$, $\mathsf{C} = C\mathbb{1}$, $\mathsf{D} = D\mathbb{1}$ (A, B, C, $D \in \mathbb{R}$).

2.1.3. Light Field

With matrix optics, we have seen how we can associate at each point a light ray by a vector containing the position and inclination of the ray. In real world scenarios, however, we would like to consider a bundle of rays attached to every point in space, each with a radiance associated with the inclination of the respective ray at a point. This is called the *light field*.

The light field L is described by a five-dimensional *plenoptic function* that tells us the amplitude of light emitted from a point $\boldsymbol{x} \in \mathbb{R}^3$ at a given orientation $\boldsymbol{\theta} \in S^2$. I.e., $L = L(\boldsymbol{x}, \boldsymbol{\theta}) : \mathbb{R}^3 \times S^2 \to [0, \infty)$. When we restrict our discussion to a plane of constant z coordinate in space, then we can rewrite the light field as:

$$L = L(x, y, \theta_1, \theta_2),$$

the two angles θ_1 and θ_2 are as defined in Figure 2. Also see Figure 3 for a two-dimensional diagram of a light field. Record the 4 parameters by a vector $\tilde{\boldsymbol{x}}^\mathsf{T} = [x, y, \theta_1, \theta_2]^\mathsf{T}$, then the transformation of light field under a system with ABCD matrix M will be

$$L'(\tilde{\boldsymbol{x}}') = L(M^{-1}\tilde{\boldsymbol{x}}').$$

2.2. Diffraction

From the advancements made by Thomas Young, James Clerk Maxwell, Heinrich Rudolf Hertz and many others, we know that lights are far from being one-dimensional line segments in space. Lights are in fact electromagnetic waves, with both its electric and magnetic fields satisfying the wave equation:

$$\nabla^2 U = \frac{1}{c^2}\partial_t^2 U, \tag{10}$$

where the field profile U is proportional to the electric (or the magnetic) field, c is the speed of light in empty space, ∇^2 is the Laplacian operator, and ∂_t the time derivative. Note that $U : \mathbb{R}^3 \to \mathbb{C}$ is a complex scalar[2] function, with its magnitude and phase being the amplitude and phase of the field, respectively. The light intensity I is proportional to $|U|^2$. The Green's function / impulse response to the wave equation is the spherical wave emanating from the origin of the form

$$h(\boldsymbol{x}) = \frac{1}{i\lambda|\boldsymbol{x}|}\mathrm{e}^{i2\pi\boldsymbol{u}\cdot\boldsymbol{x}}, \tag{11}$$

where $\boldsymbol{x} = [x, y, z]^\mathsf{T}$. The *wave number* $2\pi\boldsymbol{u} \in \mathbb{R}^3$ is related to the light source characteristic, it denotes the spatial frequency and direction of the light emitted, i.e. $|\boldsymbol{u}| = 1/\lambda$ and it points radially outwards. Hence, given a light

[2] In general, $U : \mathbb{R}^3 \to \mathbb{C}^3$ is a vector field in space. But for our discussion, we shall consider the *scalar theory of light*, where polarization of light is omitted.

source distribution of $U_0(\boldsymbol{x})$, the electric field at $\boldsymbol{x}' = [x', y', z']^\mathsf{T}$ is proportional to

$$U(\boldsymbol{x}') = (U_0 * h)(\boldsymbol{x}') = \int_{\mathbb{R}^3} U_0(\boldsymbol{x}) \frac{\mathrm{e}^{i2\pi \boldsymbol{u}\cdot(\boldsymbol{x}'-\boldsymbol{x})}}{i\lambda|\boldsymbol{x}'-\boldsymbol{x}|}\,\mathrm{d}\boldsymbol{x}. \tag{12}$$

This is related to the *Huygen's principle* and is a result from the *scalar theory of light* where we assume that the light can be described by a scalar complex profile.

If we consider the paraxial case where the light wave travels along the z axis and $U_0 = U_0(x, y)$ with

$$z' \gg (x'-x)^2 + (y'-y)^2$$

is satisfied, then eqn.(12) can be approximated by $\boldsymbol{u}\cdot(\boldsymbol{x}'-\boldsymbol{x}) \approx u|\boldsymbol{x}'-\boldsymbol{x}|$, with $u = |\boldsymbol{u}|$, yielding

$$\begin{aligned} U(\boldsymbol{x}') &\approx \int_{\mathbb{R}^2} U_0(x,y)\frac{\mathrm{e}^{i2\pi u\sqrt{z'^2+(x'-x)^2+(y'-y)^2}}}{i\lambda z'}\,\mathrm{d}x\,\mathrm{d}y \\ &\approx \frac{\mathrm{e}^{i2\pi u z'}}{i\lambda z'}\int_{\mathbb{R}^2} U_0(x,y)\mathrm{e}^{i\frac{\pi u}{z'}\left[(x'-x)^2+(y'-y)^2\right]}\,\mathrm{d}x\,\mathrm{d}y. \end{aligned}$$

Eqn.(13) is called the *Fresnel diffraction formula*. If the region of consideration is far away and the *far-field approximation*

$$\frac{x^2+y^2}{z'} \ll \pi$$

can be used, the Fresnel diffraction formula can be further reduced to the *Fraunhofer diffraction formula*:

$$U(\boldsymbol{x}') \approx \frac{\mathrm{e}^{i2\pi u z'}}{i\lambda z'}\mathrm{e}^{i\frac{\pi u}{z'}(x'^2+y'^2)}\int_{\mathbb{R}^2} U_0(x,y)\mathrm{e}^{-i\frac{2\pi u}{z'}(x'x+y'y)}\,\mathrm{d}x\,\mathrm{d}y. \tag{13}$$

The Fresnel diffraction formula is hence also coined the near-field diffraction, while the Fraunhofer diffraction formula is called the far-field diffraction.

The above equations can be summarized as: when the region under discussion is paraxial, i.e., close to the optical axis of z, then the Fresnel diffraction formula can be used; furthermore, if the region is also far away from the light sources, the Fraunhofer diffraction formula can be used. Readers with keen eyes might notice that the Fraunhofer's diffraction formula looks oddly like the two-dimensional Fourier transform of $U_0(x, y)$! This idea will be further developed in subsection 4.2.

Moreover, the above discussions on diffraction are limited to monochromatic light sources, viz., the light is of a single wavelength λ. If the field under discussion is generated by multiple light sources with different wavelengths, each wavelength component can be computed separately by one of the diffraction formulas above, and superimposed at the end to form the final field.

2.3. Hamiltonian Optics

From the discussion above, we see how we can describe lights as rays or waves. The two views seems not associated or even conflicted! However, we shall introduce the idea of Hamiltonian optics that sheds some light on the possible linkages between the two seemingly unrelated viewpoints of light.

We have seen how a light ray in three-dimensional space can be represented by a vector in $\mathbb{R}^4$, changing as the ray enters different optical systems. If we plot the vector in a *phase space*, the changing of the vector through systems traces out an evolving curve in the phase space. To make the following arguments and diagrams easily drawable, we shall first consider the two-dimensional case of eqn.(1).

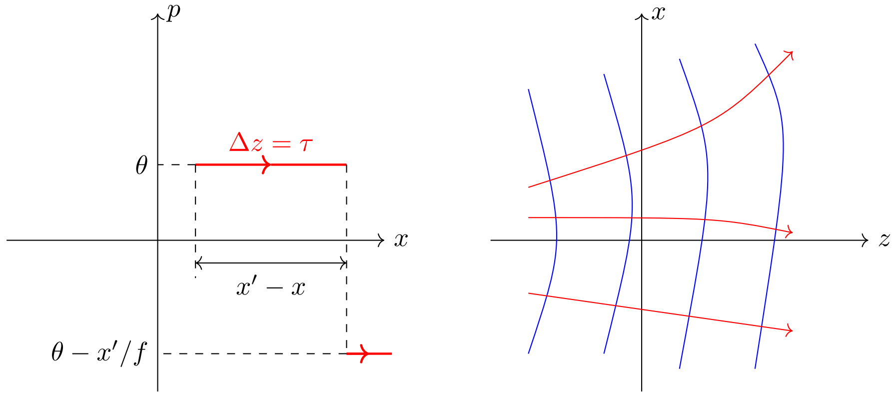
2.3.1. Hamiltonian Mechanics

From *Fermat's principle of least action*, we know that a light ray traveling in the x-z plane traverses the path $\boldsymbol{\gamma} = (x(t), z(t))$ (parameterized by arc length s) between two fixed points $\boldsymbol{x}_i$ and $\boldsymbol{x}_f$ that extremizes the time T taken:

$$T[\boldsymbol{\gamma}] = \int_{\boldsymbol{\gamma}} \mathrm{d}t = \int_{\boldsymbol{\gamma}} \frac{\mathrm{d}t}{\mathrm{d}s}\frac{\mathrm{d}s}{\mathrm{d}z}\,\mathrm{d}z = \int_{\boldsymbol{\gamma}} \frac{n}{c}\sqrt{1+\dot{x}^2}\,\mathrm{d}z. \tag{14}$$

$\frac{\mathrm{d}s}{\mathrm{d}t} = c/n$ is the speed of light in a medium with refractive index n and $\dot{q} := \frac{\mathrm{d}q}{\mathrm{d}z}$ for any variable q. We can hence define the *Lagrangian* of the system as

$$\mathcal{L}(x, \dot{x}, z) = n\sqrt{1+\dot{x}^2}. \tag{15}$$



(a) Phase Space of Light Ray: transmission as in subfigure 1b followed by lens refraction as in subfigure 1c.

(b) Wavefronts (blue) and Optical Momentum (red)

FIGURE 4: Hamiltonian Optics

The extremization of T is equivalent to $\boldsymbol{\gamma}$ satisfying the *Euler-Lagrange equation*: by setting $x_1 = x$, $x_2 = y$, $x_3 = z$,

$$\frac{\partial \mathcal{L}}{\partial x_i} - \frac{\mathrm{d}}{\mathrm{d}z}\frac{\partial \mathcal{L}}{\partial \dot{x}_i} = 0 \qquad (i = 1, 2, 3). \tag{16}$$

Next, we apply a coordinate transform from the *tangent bundle* coordinates of $(x, \dot{x}, z)$ to that of the *cotangent bundle* (x, p, z), where p, the conjugate coordinate to x, is defined as:

$$p \equiv p_x := \frac{\partial \mathcal{L}}{\partial \dot{x}} = n\frac{\dot{x}}{\sqrt{1+\dot{x}^2}} = n \sin\theta, \tag{17}$$

θ is as defined in subfigure 1b. We hence obtain the phase space representation of a light ray (see subfigure 4a) as

$$\begin{bmatrix} x \\ p \end{bmatrix} = \begin{bmatrix} x \\ n\sin\theta \end{bmatrix}. \tag{18}$$

Note that the above representation reduces to eqn.(1) under paraxial approximation, and the whole phase space is a representation of the light field.

The vector and its two coordinates evolves through z under the equations of:

$$\begin{aligned} \dot{x} &= \frac{\partial \mathcal{H}}{\partial p}, \\ \dot{p} &= -\frac{\partial \mathcal{H}}{\partial x}, \end{aligned} \tag{19}$$

where $\mathcal{H}$ is the *Hamiltonian* of the system, defined as

$$\mathcal{H} := \dot{x}p - \mathcal{L} = -\frac{n}{\sqrt{1+\dot{x}^2}} = -\sqrt{n^2 - p^2}. \tag{20}$$

The *optical path length* of a given time-extremizing path $\boldsymbol{\gamma}$ is given by

$$S[\gamma] = \int_{\boldsymbol{\gamma}} n\sqrt{1+\dot{x}^2}\,\mathrm{d}z = \int_{\boldsymbol{\gamma}} \mathcal{L}\,\mathrm{d}z. \tag{21}$$

For every light source $\boldsymbol{x}_i$ and point in space $\boldsymbol{x}_f = \boldsymbol{x}$, a time extremizing path $\boldsymbol{\gamma}$ can be given, it is a function of both $\boldsymbol{x}_i$ and $\boldsymbol{x}$. For fixed $\boldsymbol{x}_i$, the optical path lengths are functions of $\boldsymbol{x}$. The surfaces with constant S's corresponding to different time-extremizing $\boldsymbol{\gamma}$'s are known as *wavefronts*, their normals are the direction of the light rays: see subfigure 4b, we have

$$\nabla S = \sum_i \hat{x}_i \int \frac{\partial \mathcal{L}}{\partial x_i}\,\mathrm{d}z = \sum_i \hat{x}_i \int \frac{\mathrm{d}p_i}{\mathrm{d}z}\,\mathrm{d}z = (p_x, p_z) =: \boldsymbol{p}. \tag{22}$$

One should be careful that eqn.(15) and (22) are not the same. Even though both time-extremizing paths satisfy the Euler-Lagrange equation, the Lagrangian in the former is defined with fixed $\boldsymbol{x}_f$, while the latter has varying end point $\boldsymbol{x}_f = \boldsymbol{x}$ (i.e. $\boldsymbol{\gamma} = \boldsymbol{\gamma}(\boldsymbol{x}_f)$). The *optical momentum* $\boldsymbol{p}$ satisfies $|\boldsymbol{p}| = n$.

Note that all the derivations above can be easily extended to the three-dimensional case of $\boldsymbol{\gamma} = (x(t), y(t), z(t))$ and are suitable even in non-paraxial regions. The phase space representation of a light ray will, in this case, be

$$\begin{bmatrix} x \\ y \\ p_x \\ p_y \end{bmatrix} = \begin{bmatrix} x \\ y \\ n\sin\theta_x \\ n\sin\theta_y \end{bmatrix} = \begin{bmatrix} x \\ y \\ \partial_{\dot{x}}\mathcal{L} \\ \partial_{\dot{y}}\mathcal{L} \end{bmatrix}, \tag{23}$$

with the Lagrangian being defined as

$$\mathcal{L} = n\sqrt{1 + \dot{x}^2 + \dot{y}^2}. \tag{24}$$

Similarly, we have

$$\nabla S = \boldsymbol{p} = (p_x, p_y, p_z). \tag{25}$$

The Hamiltonian formulation of optics gives a link between the ray and the wave aspects of light. This final equation effectively links up the relation between ray optics and wave optics. Ray representation of light is the right hand side of eqn.(25); while the wave representation of light is related to the wavefront S. Remember that the field profile at a constant z is $U(\boldsymbol{x}') \in \mathbb{C}$, wavefronts are points $\boldsymbol{x}' \in \mathbb{R}^3$ where U have the same phase $\mathrm{Arg}(U)$.

2.4. Coherence

Spatial coherence and *temporal coherence* are two important properties of wave optics.

Previously when we discussed the diffraction formulas, we mentioned that the light sources are monochromatic, that is, they have the same wavelength. This can be seen as the temporal coherence of the light sources, where when observing the fluctuation of field at a point in space, the observed wavelength or frequency stays constant through time. Temporal coherence is critical when discussing interference of light. If the different sources are temporally incoherent, then the light emitted from these sources will not interfere with each other when averaged through time.

Observe subfigure 4b, where the wavefront has a varying profile through space. It can be observed that for different z's, the wavefront profiles are different, this is spatial incoherence. However, if we zoom in on a small interval of x (say, $x < 0$) and see the change of wavefront profile under a change of z, it seems like the profile is unchanged. This is spatial coherence. The larger the interval of x with unchanging wavefront profile, the larger the *coherence length* (the size of the interval of x) the light source has. For example, a plane wave through space has a coherence length of infinity.

3. LINEAR CANONICAL TRANSFORMATION

The geometrical ray aspect of light is often regarded as the polar opposite of the wave aspect. However, we shall show in this subsection that they are in fact related through the integral transform coined the *linear canonical transformation (LCT)*, we discuss the relationship between the two viewpoints of optics via the LCT.

The linear canonical transform is an integral transform described by four coefficients (a, b, c, d), it is of the form:

$$\mathcal{O}_{\begin{bmatrix} a & b \\ c & d \end{bmatrix}} \{f(x)\}(x') = \begin{cases} \sqrt{\frac{1}{ib}} \mathrm{e}^{i\pi \frac{d}{b} x'^2} \int_{-\infty}^{\infty} \mathrm{e}^{-i2\pi \frac{1}{b} x' x} \mathrm{e}^{i\pi \frac{a}{b} x^2} f(x) \,\mathrm{d}x & , b \neq 0; \\ \\ \sqrt{d} \mathrm{e}^{i\pi c d x'^2} f(d \cdot x') & , b = 0. \end{cases} \tag{26}$$

The coefficients (a, b, c, d) should satisfy $ad - bc = 1$, viz., it is an element of the *special linear group*:

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} \in \mathrm{SL}(2, \mathbb{R}) := \left\{ M \in \mathbb{R}^{2\times 2} \mid \det(M) = 1 \right\}.$$

3.1. Additivity of LCT

The LCT satisfies the following additivity property: the LCT transform and its coefficients isomorphic to the group SL(2,$\mathbb{R}$), i.e., for matrices M_1, $M_2 \in \mathrm{SL}(2, \mathbb{R})$,

$$\mathcal{O}_{M_2} \circ \mathcal{O}_{M_1} \{\cdot\} = \mathcal{O}_{M_2 \cdot M_1} \{\cdot\}. \tag{27}$$

The proof of additivity is often omitted in discussions of LCT, but precarious details are hidden within the formulations. Hence, a proof is given in Appendix A. This property turns out to be super useful in dealing with optical problems.

3.2. From Ray Optics to Wave Optics

The ray and wave point of view of light is related via the LCT. Consider a symmetric optical system described by the ABCD matrix

$$\begin{bmatrix} A\mathbb{1} & B\mathbb{1} \\ C\mathbb{1} & D\mathbb{1} \end{bmatrix}.$$

If a monochromatic field $U_0(x, y)$ with wavelength λ enters the system, the output field $U(x', y')$ can be described by the LCT as:

$$U(x', y') = \mathcal{O}^{\otimes 2}_{\begin{bmatrix} A & B/\lambda \\ C\lambda & D \end{bmatrix}} \{U_0(\lambda x, \lambda y)\} \left(\frac{x'}{\lambda}, \frac{y'}{\lambda} \right). \tag{28}$$

Note that (x, y) are unitless, and (x', y') have units of length. The proof will be provided after the examples below. This result effectively connects the two different aspect of optics. The same parameters that described the ray optics can in fact be used to describe the integral transformation of wave optics as well. Moreover, matrix multiplications are far easier to calculate than integrals. Using eqn.(28), we can first calculate the ABCD matrix of an optical system before doing any integrals, effectively reducing the computational complexities of wave profile transformation through composite optical systems.

3.3. Examples

Two special optical systems are to be discussed here: a free space diffraction system and a single focusing lens system. The latter will also be used in discussions of subsection 4.2.

3.3.1. Diffraction

In the overview of subsection 2.2, we introduced the Fresnel and Fraunhofer diffraction formula. They can be easily represented using the LCT. The ABCD matrix for transmission by a distance z is

$$\begin{bmatrix} 1 & z \\ 0 & 1 \end{bmatrix}.$$

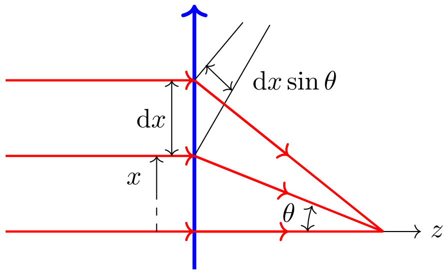


FIGURE 5: Lens Refraction

Hence, the diffracted field is

$$\begin{aligned}U_{\mathrm{diff}}(x',y') &= \mathcal{O}^{\otimes 2}_{\begin{bmatrix}1 & z/\lambda\\ 0 & 1\end{bmatrix}}\left\{U_0\left(\lambda x,\lambda y\right)\right\}\left(\frac{x'}{\lambda},\frac{y'}{\lambda}\right)\\ &= \sqrt{\frac{\lambda}{iz}}\mathrm{e}^{i\pi\frac{\lambda}{z}(y'/\lambda)^2}\int_{-\infty}^{\infty}\mathrm{e}^{-i2\pi\frac{\lambda}{z}(y'/\lambda)y}\mathrm{e}^{i\pi\frac{\lambda}{z}y^2}\sqrt{\frac{\lambda}{iz}}\mathrm{e}^{i\pi\frac{\lambda}{z}(x'/\lambda)^2}\int_{-\infty}^{\infty}\mathrm{e}^{-i2\pi\frac{\lambda}{z}(x'/\lambda)x}\mathrm{e}^{i\pi\frac{\lambda}{z}x^2}U_0(\lambda x,\lambda y)\,\mathrm{d}x\,\mathrm{d}y\\ &= \frac{1}{i\lambda z}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\mathrm{e}^{i\frac{\pi}{\lambda z}\left[(x'-x)^2+(y'-y)^2\right]}U_0(x,y)\,\mathrm{d}x\,\mathrm{d}y.\end{aligned} \tag{29}$$

Compare the above with the Fresnel diffraction formula of eqn.(13), we can immediately see that it is identical under a global phase difference of $\mathrm{e}^{i\frac{2\pi}{\lambda}z}$, with $u = 1/\lambda$.

3.3.2. Single Lens System

The ABCD matrix for reflection by thin lens of focus f is

$$\begin{bmatrix}1 & 0\\ -1/f & 1\end{bmatrix}.$$

Hence, the refracted field is

$$U_{\mathrm{refr}}(x',y') = \mathcal{O}^{\otimes 2}_{\begin{bmatrix}1 & 0\\ -\lambda/f & 1\end{bmatrix}}\left\{U_0\left(\lambda x,\lambda y\right)\right\}\left(\frac{x'}{\lambda},\frac{y'}{\lambda}\right) = \mathrm{e}^{-i\pi\frac{\lambda}{f}(x'/\lambda)^2}\mathrm{e}^{-i\pi\frac{\lambda}{f}(y'/\lambda)^2}U_0(x',y') = \mathrm{e}^{-i\frac{\pi}{f\lambda}(x'^2+y'^2)}U_0(x',y'). \tag{30}$$

The total effect is a quadratic phase modulation. How does this additional phase account for the refraction effects? See Figure 5, consider a plane wave propagating in space with rays parallel to the optical axis. When refracted by a lens, the ray at height x is refracted to an inclination angle of

$$-\theta \approx -\tan\theta = -\frac{x}{f}.$$

Suppose the lens adds a phase profile of $\exp(i\phi)$ to the incoming wave, with $\phi = \phi(x)$. The phase difference that corresponds to this *tilting* of rays and wavefronts is equal to

$$-\frac{2\pi}{\lambda}\cdot \mathrm{d}x\sin\theta = \frac{\partial\phi}{\partial x}\mathrm{d}x.$$

Hence,

$$\frac{\partial\phi}{\partial x} \approx -\frac{2\pi x}{f\lambda} \to \phi = -\frac{\pi x^2}{f\lambda}. \tag{31}$$

This is exactly equivalent to the phase modulation we obtained in eqn.(30).

3.4. Proof of LCT and ABCD Matrix Relation

First consider the case where $C = 0$: if $B = 0$, this is the trivial case of an identity transform; if $B \neq 0$, this is the first example of a transmission system. Next, the case where $B = 0$ and $C \neq 0$, this is the thin lens case discussed above. Hence, the relation (28) holds.

As for the case of B, $C \neq 0$, first notice the fact that an SL(2,$\mathbb{R}$) matrix can be decomposed as two refractions sandwiching a transmission for $B \neq 0$:

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ \frac{D-1}{B} & 1 \end{bmatrix} \begin{bmatrix} 1 & B \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ \frac{A-1}{B} & 1 \end{bmatrix}. \tag{32}$$

Then the effect of the whole optical system is, by the additivity of LCT,

$$\mathcal{O}_{\begin{bmatrix} 1 & 0 \\ \frac{D-1}{B}\lambda & 1 \end{bmatrix}} \mathcal{O}_{\begin{bmatrix} 1 & B/\lambda \\ 0 & 1 \end{bmatrix}} \mathcal{O}_{\begin{bmatrix} 1 & 0 \\ \frac{A-1}{B}\lambda & 1 \end{bmatrix}} = \mathcal{O}_{\begin{bmatrix} 1 & 0 \\ \frac{D-1}{B}\lambda & 1 \end{bmatrix} \begin{bmatrix} 1 & B/\lambda \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ \frac{A-1}{B}\lambda & 1 \end{bmatrix}} = \mathcal{O}_{\begin{bmatrix} A & B/\lambda \\ C\lambda & D \end{bmatrix}}. \tag{33}$$

Tensoring the transform to $\mathcal{O}^{\otimes 2}$ for symmetric systems doesn't change the result. Thus it is proven. ∎

The relationship between the 4×4 optical matrix and the integral transform for am asymmetric optical systems, i.e. general paraxial optical system in three dimensions, will be discussed in subsection 5.3.

3.5. On Units of LCT

The LCT as defined in eqn.(26) have the coordinate x, the conjugate coordinate x', and the coefficients (a, b, c, d) be unitless. Some sources may abuse the notation of the linear canonical transformation to allow for the above parameters to have units. Allowing such abuse of notation, eqn.(28) can be rewritten as

$$U(x', y') = \mathscr{O}^{\otimes 2}_{\begin{bmatrix} A & B\lambda \\ C/\lambda & D \end{bmatrix}} \left\{ U_0\left(x, y\right) \right\} \left(x', y'\right). \tag{34}$$

The script font LCT $\mathscr{O}$ for parameters with unit is used to differentiate from the unitless LCT $\mathcal{O}$. The two transforms coincide when all parameters are unitless. As can be seen from comparison between eqn.(28) and (34), the following relation holds for the transformation of fields:

$$\begin{aligned} \mathscr{O}_{\begin{bmatrix} A & B\lambda \\ C/\lambda & D \end{bmatrix}} &= \mathscr{O}_{\begin{bmatrix} \lambda & 0 \\ 0 & 1/\lambda \end{bmatrix}} \mathscr{O}_{\begin{bmatrix} A & B/\lambda \\ C\lambda & D \end{bmatrix}} \mathscr{O}_{\begin{bmatrix} 1/\lambda & 0 \\ 0 & \lambda \end{bmatrix}} \\ &= \mathscr{O}_{\begin{bmatrix} \lambda & 0 \\ 0 & 1/\lambda \end{bmatrix}} \mathcal{O}_{\begin{bmatrix} A & B/\lambda \\ C\lambda & D \end{bmatrix}} \mathscr{O}_{\begin{bmatrix} 1/\lambda & 0 \\ 0 & \lambda \end{bmatrix}}. \end{aligned} \tag{35}$$

Where the sandwiching transformation is just uniform scaling: let σ be the scaling factor, then

$$\mathscr{O}_{\begin{bmatrix} \sigma^{-1} & 0 \\ 0 & \sigma \end{bmatrix}} \{f(x)\} = f(\sigma x).$$

In fact, the above decomposition of LCT has two degrees of freedom of scaling:

$$\begin{aligned} \mathscr{O}_{\begin{bmatrix} a \cdot \sigma_2/\sigma_1 & b \cdot \sigma_1\sigma_2 \\ c/\sigma_1\sigma_2 & d \cdot \sigma_1/\sigma_2 \end{bmatrix}} &= \mathscr{O}_{\begin{bmatrix} \sigma_2 & 0 \\ 0 & 1/\sigma_2 \end{bmatrix}} \mathscr{O}_{\begin{bmatrix} a & b \\ c & d \end{bmatrix}} \mathscr{O}_{\begin{bmatrix} 1/\sigma_1 & 0 \\ 0 & \sigma_1 \end{bmatrix}} \\ &= \mathscr{O}_{\begin{bmatrix} \sigma_2 & 0 \\ 0 & 1/\sigma_2 \end{bmatrix}} \mathcal{O}_{\begin{bmatrix} a & b \\ c & d \end{bmatrix}} \mathscr{O}_{\begin{bmatrix} 1/\sigma_1 & 0 \\ 0 & \sigma_1 \end{bmatrix}}, \end{aligned} \tag{36}$$

as long as both σ_1 and σ_2 have the same units, eqn.(36) further requires the units of σ_1 and σ_2 be length.

4. FOURIER TRANSFORM

Fourier transform is the classic signal processing integral transformation that inspired many of the developments in time-frequency analysis. As we have seen, through the LCT, many integral transforms can be represented by a simple symmetric optical system, the Fourier transform is, without a doubt, one of them. In this section, we shall show how the Fourier transform arise naturally from a transmission-refraction setup, leading to in inspiring field of optical analysis called *Fourier optics*.

Firstly, we shall define the *Fourier transform* of a function as

$$\mathscr{F}\{f(x)\}(x') = \int_{-\infty}^{\infty} f(x) \mathrm{e}^{-i2\pi x' x} \, \mathrm{d}x, \tag{37}$$

and the *fractional Fourier transform* of *angle* α is

$$\mathscr{F}_\alpha \left\{ f(x) \right\} (x') = \sqrt{1 - i\cot(\alpha)} \, \mathrm{e}^{i\pi \cot(\alpha) x'^2} \int_{-\infty}^{\infty} \mathrm{e}^{-i2\pi \csc(\alpha) x' x} \mathrm{e}^{i\pi \cot(\alpha) x^2} f(x) \, \mathrm{d}x. \tag{38}$$

The following properties regarding the two transforms hold:

1. $\mathscr{F} = \sqrt{i}\,\mathcal{O}_{\begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}}$.
2. $\mathscr{F}_\alpha = \mathrm{e}^{i\frac{\alpha}{2}}\,\mathcal{O}_{\begin{bmatrix} \cos\alpha & \sin\alpha \\ -\sin\alpha & \cos\alpha \end{bmatrix}}$.
3. $\mathscr{F} = \mathscr{F}_{\pi/2}$.
4. $\mathscr{F}_\alpha \mathscr{F}_\beta = \mathscr{F}_{\alpha+\beta}$.
5. $\mathscr{F}^n = \mathscr{F}_{n\pi/2}$.
6. Allowing abuse of notation for x and x' to contain units as long as the integrand has unitless exponents, then for a scaling of σ with units: $\mathscr{F}\{f(\sigma x)\}(x') = \frac{1}{\sigma}\mathscr{F}\{f(x)\}\left(\frac{x'}{\sigma}\right)$.

4.1. Diffraction Across a Lens

First consider the field transformation between the front and rear focal plane of a lens with focal distance f. The ABCD matrix for such a system is

$$\begin{bmatrix} 1 & f \\ 0 & 1 \end{bmatrix}\begin{bmatrix} 1 & 0 \\ -1/f & 1 \end{bmatrix}\begin{bmatrix} 1 & f \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 0 & f \\ -1/f & 0 \end{bmatrix}.$$

By utilizing the LCT, given an input field of $U_0(x, y)$ on the front focal plane, the output field at the rear focal plane will be transformed by the transformation of

$$\mathscr{O}^{\otimes 2}_{\begin{bmatrix} 0 & f\lambda \\ -1/f\lambda & 0 \end{bmatrix}} = \mathscr{O}^{\otimes 2}_{\begin{bmatrix} \lambda & 0 \\ 0 & 1/\lambda \end{bmatrix}} \mathcal{O}^{\otimes 2}_{\begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}} \mathscr{O}^{\otimes 2}_{\begin{bmatrix} 1/f & 0 \\ 0 & f \end{bmatrix}}. \tag{39}$$

We can clearly see that the two fields are related via the Fourier transform (along with some scalings)! The associated ideas will be further developed in subsection 4.2.

Next consider the field transformation between the a plane of distance z in front of a lens with focal distance f and another plane also with a distance z behind the lens. The ABCD matrix for such a system between the two planes is

$$\begin{bmatrix} 1 & z \\ 0 & 1 \end{bmatrix}\begin{bmatrix} 1 & 0 \\ -1/f & 1 \end{bmatrix}\begin{bmatrix} 1 & z \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 - z/f & 2z - z^2/f \\ -1/f & 1 - z/f \end{bmatrix}.$$

The transformation of fields follows

$$\mathscr{O}^{\otimes 2}_{\begin{bmatrix} 1 - z/f & (2z - z^2/f)\lambda \\ -1/f\lambda & 1 - z/f \end{bmatrix}} = \mathscr{O}^{\otimes 2}_{\begin{bmatrix} \sigma & 0 \\ 0 & 1/\sigma \end{bmatrix}} \mathcal{O}^{\otimes 2}_{\begin{bmatrix} \cos\alpha & \sin\alpha \\ -\sin\alpha & \cos\alpha \end{bmatrix}} \mathscr{O}^{\otimes 2}_{\begin{bmatrix} 1/\sigma & 0 \\ 0 & \sigma \end{bmatrix}}, \tag{40}$$

where

$$\cos\alpha = 1 - \frac{z}{f}, \qquad \sin\alpha = \sqrt{2\frac{z}{f} - \frac{z^2}{f^2}}, \qquad \sigma = \left(\lambda^2(2zf - z^2)\right)^{1/4}.$$

Notice that a restriction of $f \geq z$ exists. We can see that such a system follows the fractional Fourier transform of angle $\alpha = \arccos(1 - z/f)$.

4.2. Fourier Optics

We've seen that

$$\begin{aligned} U(x', y') &= \mathscr{O}^{\otimes 2}_{\begin{bmatrix} 0 & f\lambda \\ -1/f\lambda & 0 \end{bmatrix}}\{U_0(x, y)\}(x, y) = \mathscr{O}^{\otimes 2}_{\begin{bmatrix} \lambda & 0 \\ 0 & 1/\lambda \end{bmatrix}} \mathcal{O}^{\otimes 2}_{\begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}} \mathscr{O}^{\otimes 2}_{\begin{bmatrix} 1/f & 0 \\ 0 & f \end{bmatrix}}\{U_0(x, y)\}(x', y') \\ &= \left(\sqrt{\frac{1}{i}}\right)^2 \left(\mathscr{O}^{\otimes 2}_{\begin{bmatrix} \lambda & 0 \\ 0 & 1/\lambda \end{bmatrix}} \mathscr{F}^{\otimes 2} \mathscr{O}^{\otimes 2}_{\begin{bmatrix} 1/f & 0 \\ 0 & f \end{bmatrix}}\right)\{U_0(x, y)\}(x', y') \\ &= \frac{1}{i}\,\mathscr{F}^{\otimes 2}\{U_0(fx, fy)\}\left(\frac{x'}{\lambda}, \frac{y'}{\lambda}\right). \end{aligned} \tag{41}$$

That is, the field transformation between the front and rear focal plane of a lens is the Fourier transform. This fact forms the foundation of Fourier optics, a branch of optical system analysis, useful in constructing optical image processing systems. The above setup allows the spatial frequencies $(k_x, k_y) = 2\pi(1/\lambda_x, 1/\lambda_y)$ to be representable by spatial (x, y) coordinates. This allows us to modulate a field in its frequency domain spatially.

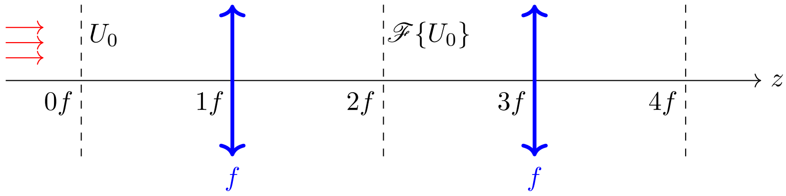


FIGURE 6: $4f$-Correlator: the blue arrows denote lenses.

4.2.1. Relation with Far Field Diffraction

In eqn.(13), we have seen that far field diffraction is equivalent to the Fourier transform of the input field, too. How is it related to the front-to-rear focal plane system above? Conceptually, parallel rays shall travel to and intercept at the point at infinity, i.e. the far-field. By placing a lens in front of the parallel light rays, we effectively force the intercepts of the light rays to be on the rear focal plane. Since a lens moves the plane at infinity to the rear focal plane, the diffraction pattern for these two cases shall be the same, both are described by the Fourier transform. Mathematically, note that

$$(13) = \mathrm{e}^{i\frac{2\pi}{\lambda}z} \cdot \mathscr{O}^{\otimes 2}_{\begin{bmatrix} 0 & \lambda z \\ -1/\lambda z & 1 \end{bmatrix}} = \mathrm{e}^{i\frac{2\pi}{\lambda}z} \mathscr{O}^{\otimes 2}_{\begin{bmatrix} \lambda & 0 \\ 0 & 1/\lambda \end{bmatrix}} \mathcal{O}^{\otimes 2}_{\begin{bmatrix} 1 & 0 \\ \lambda/z & 1 \end{bmatrix}} \mathcal{O}^{\otimes 2}_{\begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}} \mathscr{O}^{\otimes 2}_{\begin{bmatrix} 1/z & 0 \\ 0 & z \end{bmatrix}}.$$

The field is first normalized by its traversing distance z, then Fourier transformed, then phase modulated proportional to the optical length traversed, the scaled back with respect to its wavelength.

On the other hand, in experiments that want to know the far-field diffraction pattern of field through an aperture, instead of placing a screen at infinity, one can simply use a lens to conduct the experiment in finite space.

4.2.2. 4f-Correlator

Consider the optical setup as in Figure 6, this is called a *4f-correlator*. Placing an object at $0f$ and illuminate from behind by a monochromatic light source of wavelength λ, creating an input field of U_{0f}. Its image at $2f$ after passing through the lens will be its Fourier transform. Spatial modulation at $2f$ will result in the $4f$ location an image of the frequency-modulated original object (with coordinates inverted). That is, given the input field at $0f$ be $U_{0f}(x, y)$ and a spatial filter of $A(x', y')$ (A stands for aperture) is placed at $2f$, the field at $4f$ will be proportional to

$$\begin{aligned} U_{4f}(x'', y'') &= \frac{1}{i}\mathscr{F}^{\otimes 2}\{A(fx', fy')U_{2f}(fx', fy')\}\left(\frac{x''}{\lambda}, \frac{y''}{\lambda}\right) \\ &= -\mathscr{F}^{\otimes 2}\left\{A(fx', fy') \cdot \mathscr{F}^{\otimes 2}\{U_{0f}(fx, fy)\}\left(\frac{fx'}{\lambda}, \frac{fy'}{\lambda}\right)\right\}\left(\frac{x''}{\lambda}, \frac{y''}{\lambda}\right) \\ &= -\frac{\lambda^2}{f^2}\,\mathscr{F}^{\otimes 2}\left\{A(x', y') \cdot \mathscr{F}^{\otimes 2}\{U_0(x, y)\}(x', y')\right\}(x'', y''). \end{aligned} \tag{42}$$

The last line abuses the notation of Fourier transform to allow for parameters with units, making the whole equation cleaner.

Some special cases of A is to be considered:

1. Consider the ideal case where $A(x'y') = 1$, i.e. there is no filtering, then $U_{4f}(x'', y'') \propto U_{0f}(-x, -y)$.
2. In reality, optical elements such as lenses have finite sizes, hence creating a finite aperture stop within the system. From eqn.(42), a finite aperture stop corresponds to $A(x', y') = 0$ for x and y big enough, creating a *low-pass filtering* effect on the final image of U_{4f}, making it blurry.

4.2.3. Maximum Spatial Frequency

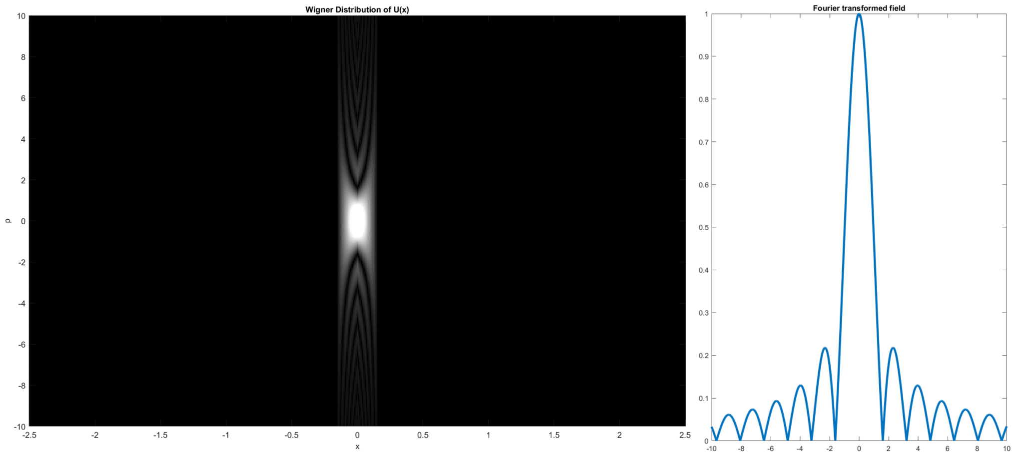
The above analysis is in essence flawed. In fact, there exists a low-pass filter on U_{0f} already. U_{0f} cannot perfectly represent the details of the original object. Consider Figure 5, the light source of wavelength λ is diffuse after passing through the original object. If it is diffused with inclination of θ, then the effective wavelength it projects on the $0f$ plane will be

$$\lambda / \sin\theta,$$

corresponding to a spatial frequency of

$$\sin\theta/\lambda.$$

The maximum spatial frequency in U_{0f} allowed is hence $1/\lambda$. When conducting the $4f$-correlator experiment, when a red light is used to illuminate the object, the image at $4f$ will be blurrier than that when using a blue light.



(a) WDF of a rectangular function. (b) Integration of WDF along x.

FIGURE 7: Wigner Distribution Function

5. WIGNER DISTRIBUTION FUNCTION

The Wigner distribution function (WDF) of a function $f = f(x, t)$ that depends on both space x and time t is defined as

$$W_f(x,p) := \int_{-\infty}^{\infty} \left\langle f\left(x+\frac{\xi}{2},t\right) f^*\left(x-\frac{\xi}{2},t\right)\right\rangle_t \mathrm{e}^{-i2\pi p\xi}\,\mathrm{d}\xi, \tag{43}$$

where $\langle\cdot\rangle_t$ denotes taking the average through time. The *mutual coherence function* is defined as

$$\Gamma_{f_1 f_2}(x_1,x_2) := \langle f_1(x_1,t) f_2^*(x_2,t)\rangle_t. \tag{44}$$

the Wigner distribution function is equivalent to the Fourier transform of the mutual coherence function. Furthermore, if the function is independent of time, i.e. $f = f(x)$, then the equation above can be further reduced to the form below:

$$W_f(x,p) = \int_{-\infty}^{\infty} f\left(x+\frac{\xi}{2}\right) f^*\left(x-\frac{\xi}{2}\right) \mathrm{e}^{-i2\pi p\xi}\,\mathrm{d}\xi. \tag{45}$$

The Wigner distribution function transforms a function from its spatial domain x to its *spectrogram* (x, p), also known as its *phase space*.

5.1. Phase Space Characterization

When the signal represents a light wave in space, its spectrogram is equivalent to the phase space representation of light. Let us consider a few examples.

1. A point light source: $f = \delta(x - x_0)$, then $W_f(x, p) = \delta(x - x_0)$. This corresponds the fact that though situated at a single point, the point source emits light with equal intensity at all directions.
2. A plane wave traveling at an angle θ from the optical axis: $f = \mathrm{e}^{i\frac{2\pi}{\lambda}\sin\theta x}$, then $W_f = \delta\left(p - \frac{\sin\theta}{\lambda}\right)$. This corresponds to the fact that this plane wave travels with $p = \sin\theta/\lambda$.
3. Spherical wave with radius R: $f(x) = \mathrm{e}^{i\frac{\pi}{\lambda R}x^2}$, then $W_f = \delta\left(p - \frac{x}{\lambda R}\right)$.

As can be seen, the physical interpretation of the Wigner distribution function is apparent, having close ties with the phase space representation of light. An example is also given in Figure 7.

5.2. Relation with LCT

The following property of the Wigner distribution function with the linear canonical transformation is of utmost concern for us: for $M \in \mathrm{SL}(2,\mathbb{R})$ with unitless entries a, b, c and d, then

$$W_{\mathcal{O}_M\{f\}}(x,p) = W_f(dx - bp, -cx + ap). \tag{46}$$

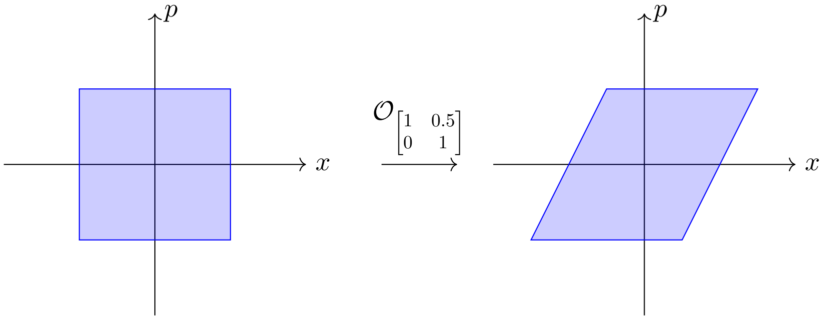


FIGURE 8: LCT on Wigner Distribution Function

The proof is identical to that of subsection 3.4. The above formula states that when f is transformed by the LCT described by the matrix M, the spectrogram of f is equivalently transformed by M. See Figure 8 for example.

With this visualization at hand, we can rethink the LCT as a geometric linear transformation on the phase space of light.

5.3. Transfer Function of Wave Profile in 3D

Previously in subsection 3.2, we've seen that the field $U_0(x')$ at the input plane of an optical system can be transferred to the output plane $U(x')$ by utilizing the LCT. The integral transformation even works for symmetric optical systems in three dimensions. However, the formula fails for asymmetric systems.

In this subsection, we will be developing a mathematical framework to transfer a monochromatic wave profile with wavelength λ at the input plane $U_0(\boldsymbol{x})$ (where $\boldsymbol{x} = [x, y]^\mathsf{T}$) to the output plane $U(\boldsymbol{x}')$ (where $\boldsymbol{x}' = [x', y']^\mathsf{T}$) where the system is described by

$$\begin{bmatrix}\boldsymbol{x}'\\ \lambda\boldsymbol{p}'\end{bmatrix} = \begin{bmatrix}\mathsf{A} & \mathsf{B}\\ \mathsf{C} & \mathsf{D}\end{bmatrix}\begin{bmatrix}\boldsymbol{x}\\ \lambda\boldsymbol{p}\end{bmatrix}. \tag{47}$$

$\boldsymbol{p}$ is the optical momentum as described in subsection 2.3, $\lambda\boldsymbol{p}$ is then the inclination of a light ray, and A, B, C, D are each a 2×2 matrix requiring that $\mathsf{A}\mathsf{D}^\mathsf{T} - \mathsf{B}\mathsf{C}^\mathsf{T} = \mathbb{1}$. A detailed proof mirrors that of subsection 3.4, and is provided in Appendix B. The main results are as follows.

Given the field $U_0(\boldsymbol{x})$ at the input plane, the field at the output plane of a paraxial system will be

$$U(\boldsymbol{x}') = \tilde{\mathscr{O}}_{\begin{bmatrix}\mathsf{A} & \mathsf{B}\\ \mathsf{C} & \mathsf{D}\end{bmatrix}}\{U_0(\boldsymbol{x})\}(\boldsymbol{x}') := \int_{\mathbb{R}^2} g(\boldsymbol{x}', \boldsymbol{x}) U_0(\boldsymbol{x})\,\mathrm{d}^2x, \tag{48}$$

where the *point-spread function* g is defined as

$$g(\boldsymbol{x}', \boldsymbol{x}) = \frac{1}{i\lambda\sqrt{\det(\mathsf{B})}}\exp\left(i\frac{\pi}{\lambda}\boldsymbol{x}^\mathsf{T}\mathsf{B}^{-1}\mathsf{A}\boldsymbol{x} - i\frac{2\pi}{\lambda}\boldsymbol{x}^\mathsf{T}\mathsf{B}^{-1}\boldsymbol{x}' + i\frac{\pi}{\lambda}\boldsymbol{x}'^\mathsf{T}\mathsf{D}\mathsf{B}^{-1}\boldsymbol{x}'\right) \tag{49}$$

for $\det(\mathsf{B}) \neq 0$; and we have

$$\begin{aligned} g(\boldsymbol{x}', \boldsymbol{x}) &= \frac{1}{\sqrt{\det(\mathsf{A}^{-1})}}\cdot\delta(\boldsymbol{x} - \mathsf{A}^{-1}\boldsymbol{x}')\exp\left(i\frac{\pi}{\lambda}\boldsymbol{x}'^\mathsf{T}\mathsf{C}\mathsf{A}^{-1}\boldsymbol{x}'\right)\\ &= \sqrt{\det(\mathsf{D})}\cdot\delta(\boldsymbol{x} - \mathsf{D}^\mathsf{T}\boldsymbol{x}')\exp\left(i\frac{\pi}{\lambda}\boldsymbol{x}'^\mathsf{T}\mathsf{C}\mathsf{D}^\mathsf{T}\boldsymbol{x}'\right) \end{aligned} \tag{50}$$

for $\det(\mathsf{B}) = 0$, with $\mathsf{A}\mathsf{D}^\mathsf{T} = \mathbb{1}$. This is the *generalized LCT*.

The above result is in relation with the Wigner distribution function since we can also obtain the transformation of it after passing through the optical system:

$$W_U(\boldsymbol{x}', \boldsymbol{p}') = \int_{\mathbb{R}^4} K(\boldsymbol{x}', \boldsymbol{p}', \boldsymbol{x}, \boldsymbol{p})\; W_{U_0}(\boldsymbol{x}, \boldsymbol{p})\,\mathrm{d}^2x\,\mathrm{d}^2p, \tag{51}$$

where

$$\begin{aligned} K(\boldsymbol{x}', \boldsymbol{p}', \boldsymbol{x}, \boldsymbol{p}) &= \int_{\mathbb{R}^4} g(\boldsymbol{x}' + \boldsymbol{\xi}'/2, \boldsymbol{x} + \boldsymbol{\xi}/2) g^*(\boldsymbol{x}' - \boldsymbol{\xi}'/2, \boldsymbol{x} - \boldsymbol{\xi}/2)\mathrm{e}^{-i2\pi(\boldsymbol{p}'^\mathsf{T}\boldsymbol{\xi}' - \boldsymbol{p}^\mathsf{T}\boldsymbol{\xi})}\,\mathrm{d}^2\xi'\,\mathrm{d}^2\xi\\ &= \delta\left(\boldsymbol{p}' - (\mathsf{D}\mathsf{B}^{-1}\boldsymbol{x}' - \mathsf{B}^{-\mathsf{T}}\boldsymbol{x})/\lambda\right)\times\delta\left(\boldsymbol{p} + (-\mathsf{B}^{-1}\boldsymbol{x}' + \mathsf{B}^{-1}\mathsf{A}\boldsymbol{x})/\lambda\right) \end{aligned} \tag{52}$$

is termed the *ray-spread function*. Like how the point-spread function describes the effect of a point source at the input plane, the ray-spread function describes how a ray (a point in the phase space) transforms.

One can see that if $n = 1$, then the results above coincides with that of eqn.(26). Additivity of the transforms also holds. These results extends beyond 4×4 ABCD matrices, it hold for all $2n \times 2n$ matrices with blocks satisfying the *symplectic relation* of $\mathsf{AD}^\mathsf{T} - \mathsf{BC}^\mathsf{T} = \mathbb{1}$. Such matrices belong in $\mathrm{Sp}(2n, \mathbb{R})$, the *symplectic group* of order $2n$ over $\mathbb{R}$.

Lastly, the action of the generalized LCT on a field is equivalent to a symplectic transformation on the phase space, i.e. for $\boldsymbol{q} = [\boldsymbol{x}^\mathsf{T}, \lambda \boldsymbol{p}^\mathsf{T}]^\mathsf{T}$ and $M \in \mathrm{Sp}(4, \mathbb{R})$,

$$W_{\tilde{\mathscr{O}}_M\{U\}}(\boldsymbol{q}) = W_U(M^{-1}\boldsymbol{q}). \tag{53}$$

The proof is provided in Appendix C. It should be noted that the above equation uses an abuse of notation by writing

$$W_f(\boldsymbol{x}, \boldsymbol{p}) \equiv W_f(\boldsymbol{q}) = W_f([\boldsymbol{x}^\mathsf{T}, \lambda \boldsymbol{p}^\mathsf{T}]^\mathsf{T}). \tag{54}$$

5.4. Coherence of Source

In all the previous discussions, we have only considered sources with a single wavelength, these are temporal coherent sources. Yet what will happen if U_0 consists of multiple different wavelengths? How will the light fields of different wavelengths interfere with each other?

Consider two sources of light:

$$\tilde{U}_1 = U_1(x) \cdot \mathrm{e}^{i2\pi \frac{c}{\lambda_1} t} \text{ and } \tilde{U}_2 = U_2(x) \cdot \mathrm{e}^{i2\pi \frac{c}{\lambda_2} t},$$

where c is the speed of light and λ_1, λ_2 being their respective wavelength. To find the Wigner distribution function of the total field $U = \tilde{U}_1 + \tilde{U}_2$, we must find the mutual coherence function of U:

$$\begin{aligned}\Gamma_{\tilde{U}_1\tilde{U}_2}(x_1, x_2) &= \left\langle \left(U_1(x_1)\mathrm{e}^{i2\pi \frac{c}{\lambda_1} t} + U_2(x_1)\mathrm{e}^{i2\pi \frac{c}{\lambda_2} t} \right) \cdot \left(U_1^*(x_2)\mathrm{e}^{-i2\pi \frac{c}{\lambda_1} t} + U_2^*(x_2)\mathrm{e}^{-i2\pi \frac{c}{\lambda_2} t} \right) \right\rangle_t \\ &= U_1(x_1)U_1^*(x_2) + U_2(x_1)U_2^*(x_2) + \left\langle U_1(x_1)U_2^*(x_2)\mathrm{e}^{i2\pi\left(\frac{c}{\lambda_1} - \frac{c}{\lambda_2}\right)t} + U_2(x_1)U_1^*(x_2)\mathrm{e}^{i2\pi\left(\frac{c}{\lambda_2} - \frac{c}{\lambda_1}\right)t} \right\rangle_t .\end{aligned} \tag{55}$$

One can immediately see that if $\lambda_1 \neq \lambda_2$, the exponential terms averages to zero in *principal value*. And then the Wigner distribution function is additive:

$$W_{\tilde{U}_1 + \tilde{U}_2} = W_{U_1} + W_{U_2} \qquad (\lambda_1 \neq \lambda_2). \tag{56}$$

However, if $\lambda_1 = \lambda_2$, *cross-terms* appears:

$$W_{\tilde{U}_1 + \tilde{U}_2} = W_{U_1} + W_{U_2} + \text{cross-terms} \qquad (\lambda_1 \neq \lambda_2). \tag{57}$$

This corresponds to the fact that incoherent light sources do not interfere with each other in average; while for coherent light sources, interference patterns emerge. Moreover, the Wigner distribution function satisfies the following projective property of

$$|U(\boldsymbol{x})|^2 = \int_{\mathbb{R}^2} W_U(\boldsymbol{x}, \boldsymbol{p})\, \mathrm{d}^2 p. \tag{58}$$

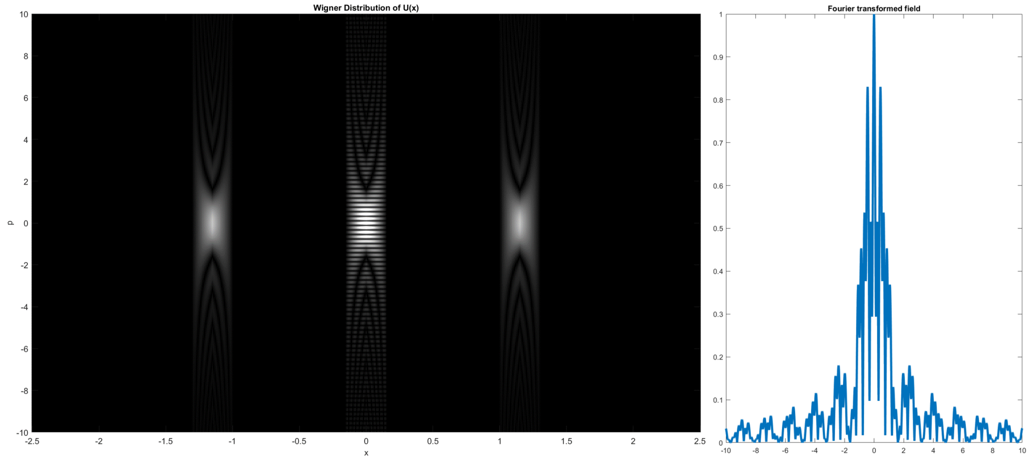
We can hence project all the energy from the phase space to the $\boldsymbol{x}$-plane and find out the intensity of the field as a function of $\boldsymbol{x}$. The proof is provided in Appendix C. See Figure 9 as an example and compare it with Figure 7. When the Wigner distribution function is integrated along $\boldsymbol{x}$, it is equivalent to the field being Fourier transformed, then integrated along $\boldsymbol{p}$: this has the physical meaning of the field intensity after passing through a thin lens and observed at the focal plane.

5.5. Terminologies

Previously, we call U the field or the electric field, and W_U the light field. But in literature, a more suitable for the latter would be the *augmented light field*.

The light field, often denoted by L, is a purely geometric term, denoting the direction and position of a ray bundled with its intensity, it is thus a four-dimensional *plenoptic function* $L : \mathbb{R}^4 \to \mathbb{R}_{\geq 0}$. However, this restricted the value of L to be non-negative, which is not the case for the Wigner distribution function.

The Wigner distribution function can be both positive and negative, i.e., $W_U : \mathbb{R}^4 \to \mathbb{R}$. Geometrically, this means that it allows for *negative intensity rays*. We hence call the Wigner distribution function an *augmented* light field. And it is this inclusion of negative intensity rays that allows for the Wigner distribution function to describe both the geometric ray optics and the physical wave optics.



(a) WDF of sum of two rectangular functions. (b) Integration of WDF along x.

FIGURE 9: Interference of Light Field

6. CONCLUSIONS

From the linear canonical transformation to the Wigner distribution function, we have seen that time-frequency analysis is closely related to optics.

By considering normals of wavefronts as rays and representing rays by a 4×1 vector, we can encode all information of a light field in a four-dimensional phase space. This four-dimensional phase space is in fact the Wigner distribution function of the field. Linear canonical transformation is, on the other hand, equivalent to a symplectic transformation on the phase space.

By considering these symplectic transforms of the phase space and projecting the transformed phase space back to the field strengths, we can then obtain all the diffraction formulas taught in optics. And as we have seen, not only do methods of time-frequency analysis suitably describes optics, but optics also inspires and motivates many of the time-frequency analysis methods.

REFERENCES

[Stu70] Jr. Stuart A. Collins. "Lens-System Diffraction Integral Written in Terms of Matrix Optics". In: *Journal of the Optical Society of America* (1970).

[Bas79] M. J. Bastiaans. "Wigner distribution function and its application to first-order optics". In: *J. Opt. Soc. Am.* 69.12 (Dec. 1979), pp. 1710–1716. doi: 10.1364/JOSA.69.001710. url: https://opg.optica.org/abstract.cfm?URI=josa-69-12-1710.

[Wol79] Kurt Bernardo Wolf. "Integral transforms in science and engineering". In: 1979. url: https://api.semanticscholar.org/CorpusID:118318210.

[Mp82] Keating Mp. "Advantages of a block matrix formulation for an astigmatic system." In: *Optometry and Vision Science* 59 (1982), pp. 851–857. url: https://api.semanticscholar.org/CorpusID:71539136.

[Dra97] D. Dragoman. "I: The Wigner Distribution Function in Optics and Optoelectronics". In: ed. by E. Wolf. Vol. 37. Progress in Optics. Elsevier, 1997, pp. 1–56. doi: https://doi.org/10.1016/S0079-6638(08)70336-6. url: https://www.sciencedirect.com/science/article/pii/S0079663808703366.

[SGD04] Sucharita Sanyal, Ajay Ghosh, and Kashi Nath Dey. "Fractional Fourier transform in optics – a new perspective". In: *Optik* 115.2 (2004), pp. 77–85. issn: 0030-4026. doi: https://doi.org/10.1078/0030-4026-00329. url: https://www.sciencedirect.com/science/article/pii/S0030402604703332.

[ABC05] Tatiana Alieva, Martin J. Bastiaans, and Maria Luisa Calvo. "Fractional Transforms in Optical Information Processing". In: *EURASIP J. Adv. Signal Process* 2005 (Jan. 2005), pp. 1498–1519. issn: 1110-8657. doi: 10.1155/ASP.2005.1498. url: https://doi.org/10.1155/ASP.2005.1498.

[ZL09] Zhengyun Zhang and Marc Levoy. "Wigner distributions and how they relate to the light field". In: *2009 IEEE International Conference on Computational Photography (ICCP)*. 2009, pp. 1–10. doi: 10.1109/ICCPHOT.2009.5559007.

[Alo11] Miguel A. Alonso. "Wigner functions in optics: describing beams as ray bundles and pulses as particle ensembles". In: *Advances in Optics and Photonics* 3 (2011), pp. 272–365. url: https://api.semanticscholar.org/CorpusID:121166749.

[JS16] Haldun M. Ozaktas John J. Healy M. Apler Kutay and John T. Sheridan. *Linear Canonical Transforms: Theory and Applications*. Springer, 2016.

APPENDICES

Appendix A. Formulation and Additivity of LCT

The following proof for additivity is adapted from [Wol79, p. 382–389] and [JS16, p. 7–12], both written by K.B. Wolf. Let us first consider two operators $\mathcal{Q}$ and $\mathcal{P}$ on functions f:

$$\mathcal{Q}f(x) := 2\pi i x f(x), \tag{A.1}$$

$$\mathcal{P}f(x) := \frac{\mathrm{d}}{\mathrm{d}x} f(x). \tag{A.2}$$

The two operators are conjugates via the Fourier transform operator $\mathscr{F}$ (as defined in eqn.(37)), i.e. $\mathscr{F}\mathcal{P} = \mathcal{Q}\mathscr{F}$, and can be seen as coordinates in a $\mathcal{Q}$-$\mathcal{P}$ phase space. Their commutator satisfies:

$$[\mathcal{Q}, \mathcal{P}]f = (\mathcal{Q}\mathcal{P} - \mathcal{P}\mathcal{Q})f = -2\pi i f \Leftrightarrow [\mathcal{Q}, \mathcal{P}] = -2\pi i \cdot \mathbb{1}, \tag{A.3}$$

where $\mathbb{1}$ is the identity operator.

Next, consider an invertible linear transformation $\mathcal{C}_M$ acting on $\mathcal{Q}$ and $\mathcal{P}$, defined as:

$$\begin{bmatrix} \mathcal{Q}' \\ \mathcal{P}' \end{bmatrix} := \mathcal{C}_M \begin{bmatrix} \mathcal{Q} \\ \mathcal{P} \end{bmatrix} \mathcal{C}_M^{-1} = \begin{bmatrix} d & -b \\ -c & a \end{bmatrix} \begin{bmatrix} \mathcal{Q} \\ \mathcal{P} \end{bmatrix} =: M \begin{bmatrix} \mathcal{Q} \\ \mathcal{P} \end{bmatrix} \tag{A.4}$$

where $M \in \mathbb{R}^{2\times 2}$. We impose the restriction $[\mathcal{Q}', \mathcal{P}'] = -2\pi i(ad-bc)\cdot\mathbb{1}$ on the transformation. Plug in the definition of $\mathcal{Q}'$ and $\mathcal{P}'$, we hence obtain the restriction on M:

$$-2\pi i(ad-bc)\cdot \mathbb{1} = [\mathcal{Q}', \mathcal{P}'] = [d\mathcal{Q} - b\mathcal{P}, -c\mathcal{Q} + a\mathcal{P}] = \mathcal{C}_M[\mathcal{Q}, \mathcal{P}]\mathcal{C}_M^{-1} = -2\pi i \cdot \mathbb{1} \Leftrightarrow ad - bc = 1 \Leftrightarrow M \in \mathrm{SL}(2, \mathbb{R}).$$

A natural question to ask would be: what is a realization of $\mathcal{C}_M$? We propose a realization of $\mathcal{C}_M$ as an integral transform with transformation kernel $C_M(x', x)$, i.e.,

$$(\mathcal{C}_M f)(x') := \int_{\mathbb{R}} C_M(x', x) f(x)\,\mathrm{d}x. \tag{A.5}$$

Consider the effects of $\mathcal{C}_M$ on $\mathcal{Q}f$ and $\mathcal{P}f$:

$$\mathcal{C}_M\mathcal{Q}f = (\mathcal{C}_M\mathcal{Q}\mathcal{C}_M^{-1})\mathcal{C}_M f = \mathcal{Q}'(\mathcal{C}_M f) = 2\pi i d x'(\mathcal{C}_M f)(x') - b\frac{\mathrm{d}}{\mathrm{d}x'}(\mathcal{C}_M f)(x'), \tag{A.6}$$

$$\mathcal{C}_M\mathcal{P}f = (\mathcal{C}_M\mathcal{P}\mathcal{C}_M^{-1})\mathcal{C}_M f = \mathcal{P}'(\mathcal{C}_M f) = -2\pi i c x'(\mathcal{C}_M f)(x') + a\frac{\mathrm{d}}{\mathrm{d}x'}(\mathcal{C}_M f)(x'), \tag{A.7}$$

plug in eqn.(A.5),

$$\begin{aligned}
\xrightarrow{\text{(A.6)}} \int_{\mathbb{R}} C_M(x', x)\cdot 2\pi i x f(x)\,\mathrm{d}x = \mathcal{C}_M\mathcal{Q}f &= \left(2\pi i d x' - b\frac{\mathrm{d}}{\mathrm{d}x'}\right)\int_{\mathbb{R}} C_M(x', x) f(x)\,\mathrm{d}x \\
&= \int_{\mathbb{R}} \left(2\pi i d x' - b\frac{\partial}{\partial x'}\right) C_M(x', x) f(x)\,\mathrm{d}x \\
\to 2\pi i x C_M(x', x) &= \left(2\pi i d x' - b\frac{\partial}{\partial x'}\right) C_M(x', x),
\end{aligned} \tag{A.8}$$

$$\begin{aligned}
\xrightarrow{\text{(A.7)}} \int_{\mathbb{R}} C_M(x', x)\left(\frac{\partial}{\partial x} f(x)\right)\,\mathrm{d}x = \mathcal{C}_M\mathcal{P}f &= \left(-2\pi i c x' + a\frac{\mathrm{d}}{\mathrm{d}x'}\right)\int_{\mathbb{R}} C_M(x', x) f(x)\,\mathrm{d}x \\
-\int_{\mathbb{R}} \left(\frac{\partial}{\partial x} C_M(x', x)\right) f(x)\,\mathrm{d}x &= \int_{\mathbb{R}} \left(-2\pi i c x' + a\frac{\partial}{\partial x'}\right) C_M(x', x) f(x)\,\mathrm{d}x \\
\to \frac{\partial}{\partial x} C_M(x', x) &= \left(2\pi i c x' - a\frac{\partial}{\partial x'}\right) C_M(x', x).
\end{aligned} \tag{A.9}$$

Note that during the steps towards eqn.(A.9), integration by parts were used with the assumption that

$$C_M(x', x) f(x)\big|_{x=\pm\infty} = 0. \tag{A.10}$$

This restricts the feasible f's that $\mathcal{C}_M$ can act on. The simple choice of $f \in L^2(\mathbb{R})$ suffices. It can be easily verified that the solutions to eqns.(A.8), (A.9) is in the null space of the following two partial differential operators:

$$\left(\frac{\partial}{\partial x'} + i\frac{2\pi}{b}x - i\frac{2\pi d}{b}x'\right) \text{ and } \left(\frac{\partial}{\partial x} - i\frac{2\pi a}{b}x + i\frac{2\pi}{b}x'\right), \tag{A.11}$$

having a general solution of the form:

$$C_M(x', x) = \theta_M \exp\left(i\pi\frac{a}{b}x^2 - i2\pi\frac{1}{b}xx' + i\pi\frac{d}{b}x'^2\right), \tag{A.12}$$

where θ_M is a variable coefficient to be determined later.

Next, let us consider the composition of transformations with different M's. Let M_1, $M_2 \in \mathrm{SL}(2, \mathbb{R})$ and define

$$\begin{aligned}
&\mathcal{Q}'' = \mathcal{C}_{M_2}\left(\mathcal{C}_{M_1}\mathcal{Q}\mathcal{C}_M^{-1}\right)\mathcal{C}_{M_2}^{-1} = (c_2b_1 + d_2d_1)\mathcal{Q} - (a_2b_1 + b_2d_1)\mathcal{P} = \mathcal{C}_{M_1M_2}\mathcal{Q}\mathcal{C}_{M_1M_2}^{-1} =: \tilde{\mathcal{C}}_M\mathcal{Q}\tilde{\mathcal{C}}_M^{-1},\\
&\mathcal{P}'' = \mathcal{C}_{M_2}\left(\mathcal{C}_{M_1}\mathcal{P}\mathcal{C}_M^{-1}\right)\mathcal{C}_{M_2}^{-1} = -(c_2a_1 + d_2c_1)\mathcal{Q} + (a_2a_1 + b_2c_1)\mathcal{P} = \mathcal{C}_{M_1M_2}\mathcal{P}\mathcal{C}_{M_1M_2}^{-1} =: \tilde{\mathcal{C}}_M\mathcal{P}\tilde{\mathcal{C}}_M^{-1},\\
\Rightarrow\ &\tilde{\mathcal{C}}_M := \mathcal{C}_{M_2}\mathcal{C}_{M_1} = \varphi\mathcal{C}_{M_1M_2} \qquad (\varphi \in \mathbb{C}),
\end{aligned} \tag{A.13}$$

where

$$M = \begin{bmatrix} d & -b \\ -c & a \end{bmatrix} = \begin{bmatrix} c_2b_1 + d_2d_1 & -(a_2b_1 + b_2d_1) \\ -(c_2a_1 + d_2c_1) & a_2a_1 + b_2c_1 \end{bmatrix} = \begin{bmatrix} d_1 & -b_1 \\ -c_1 & a_1 \end{bmatrix}\begin{bmatrix} d_2 & -b_2 \\ -c_2 & a_2 \end{bmatrix} = M_1M_2.$$

Note that the combined transform $\tilde{\mathcal{C}}_M$ is not certainly of the form (A.5) and has a degree of freedom in φ, this is due to the sandwich product structure of the transformation. Next, we must prove the following equality:

$$(\tilde{\mathcal{C}}_M f)(x'') := (\mathcal{C}_{M_2}\mathcal{C}_{M_1}f)(x'') = \int_{\mathbb{R}}\left(\int_{\mathbb{R}} C_{M_2}(x'', x')C_{M_1}(x', x)\,\mathrm{d}x'\right)f(x)\,\mathrm{d}x \overset{?}{=} \varphi\int_{\mathbb{R}} C_{M_1M_2}(x'', x)f(x)\,\mathrm{d}x.$$

Plug in eqn.(A.12) and we have

$$\begin{aligned}
&\int_{\mathbb{R}} C_{M_2}(x'', x')C_{M_1}(x', x)\,\mathrm{d}x'\\
&\quad= \int_{\mathbb{R}} \theta_{M_2}\theta_{M_1}\exp\left(i\pi\frac{a_2}{b_2}x'^2 - i2\pi\frac{1}{b_2}x'x'' + i\pi\frac{d_2}{b_2}x''^2\right)\exp\left(i\pi\frac{a_1}{b_1}x^2 - i2\pi\frac{1}{b_1}xx' + i\pi\frac{d_1}{b_1}x'^2\right)\mathrm{d}x'\\
&\quad= \theta_{M_2}\theta_{M_1}\exp\left(i\pi\frac{d_2}{b_2}x''^2 + i\pi\frac{a_1}{b_1}x^2\right)\int_{\mathbb{R}}\exp\left(\left(i\pi\frac{a_2}{b_2} + i\pi\frac{d_1}{b_1}\right)x'^2 - \left(i2\pi\frac{x''}{b_2} + i2\pi\frac{x}{b_1}\right)x'\right)\mathrm{d}x'\\
&\quad\overset{?}{=} \varphi\theta_M\exp\left(i\pi\frac{a}{b}x^2 - i2\pi\frac{1}{b}xx'' + i\pi\frac{d}{b}x''^2\right) =: \varphi C_{M_1M_2}(x'', x),
\end{aligned}$$

Observe the following integral:

$$I(\alpha, \beta) := \int_{\mathbb{R}}\exp\left(i\alpha x'^2 - 2i\beta x'\right)\mathrm{d}x' = \sqrt{\frac{i}{\alpha}}\mathrm{e}^{-i\frac{\beta^2}{\alpha}}\int_\gamma \mathrm{e}^{-z^2}\,\mathrm{d}z,$$

where $z = \sqrt{\frac{\alpha}{i}}\left(x' - \frac{\beta}{\alpha}\right)$ and γ is the line traced out by z. Since α and β are reals (this restriction can be soften for complex numbers), the integrand is entire. One can use the M-L inequality along with Cauchy-Goursat theorem to argue for the convergence of the integral, effectively reducing it to that of a Gaussian integral. Therefore, we have

$$I(\alpha, \beta) = \sqrt{\frac{i}{\alpha}}\mathrm{e}^{-i\frac{\beta^2}{\alpha}} \cdot \sqrt{\pi}. \tag{A.14}$$

Then,

$$\begin{aligned}
\int_{\mathbb{R}} C_{M_2}(x'', x')C_{M_1}(x', x)\,\mathrm{d}x' &= \theta_{M_2}\theta_{M_1}\exp\left(i\pi\frac{d_2}{b_2}x''^2 + i\pi\frac{a_1}{b_1}x^2\right) \times I\left(\frac{a_2}{b_2} + \frac{d_1}{b_1}, \frac{x''}{b_2} + \frac{x}{b_1}\right)\\
&= \theta_{M_2}\theta_{M_1}\sqrt{\frac{ib_2b_1}{a_2b_1 + b_2d_1}}\exp\left(i\pi\frac{a}{b}x^2 - i2\pi\frac{1}{b}xx'' + i\pi\frac{d}{b}x''^2\right),
\end{aligned}$$

comparing the coefficients on both sides and we obtain

$$\theta_{M_2}\theta_{M_1}\sqrt{\frac{ib_2b_1}{a_2b_1+b_2d_1}}=\varphi\theta_M.$$

We would want to have $\varphi=1$ for the additivity property to hold. Thus, we can observe that a natural choice of the coefficient θ_M would be

$$\theta_M=\sqrt{\frac{1}{ib}}. \tag{A.15}$$

In the end, we see that the linear canonical transformation as defined in eqn.(3) is related to $\mathcal{C}_M$ by

$$\mathcal{O}_M=\mathcal{C}_{M^{-1}}, \tag{A.16}$$

we henceforth have

$$\mathcal{O}_{M_2}\mathcal{O}_{M_1}=\mathcal{C}_{M_2^{-1}}\mathcal{C}_{M_1^{-1}}=\mathcal{C}_{M_1^{-1}M_2^{-1}}=\mathcal{C}_{(M_1M_2)^{-1}}=\mathcal{O}_{M_2M_1}, \tag{A.17}$$

hence proven the additivity of the linear canonical transformation. ■

Moreover, the form of the kernel at the special case of $b=0$ can also be discussed. Since the delta function has the following limit definition of the Gaussian distribution:

$$\delta(x)=\lim_{\varepsilon\to0^+}\frac{1}{\sqrt{\pi\varepsilon}}\mathrm{e}^{-\frac{x^2}{\varepsilon}}. \tag{A.18}$$

By change of variables $\varepsilon=\frac{ib}{\pi a}$,

$$C_M(x',x)=\frac{1}{\sqrt{a}}\cdot\frac{1}{\sqrt{\pi\varepsilon}}\mathrm{e}^{-\frac{1}{\varepsilon}(x-x'/a)^2}\mathrm{e}^{i\pi\frac{d}{b}x'^2-i\pi\frac{a}{b}\frac{x'^2}{a^2}},$$

taking the limit of $\varepsilon\to0^+$ while keeping $b\to0$ on a path valid for eqn.(A.14) to hold:

$$\begin{aligned}C_M(x',x)&=\lim_{\varepsilon\to0^+}\frac{1}{\sqrt{a}}\mathrm{e}^{-\frac{1}{\varepsilon}(x-x'/a)^2}\mathrm{e}^{i\pi\left(\frac{d}{b}-\frac{1}{ab}\right)x'^2}\\&=\frac{1}{\sqrt{a}}\delta\left(x-\frac{x'}{a}\right)\mathrm{e}^{i\pi\frac{c}{a}x'^2}\\&=\sqrt{d}\delta(x-dx')\mathrm{e}^{i\pi cdx'^2}.\end{aligned}$$

Hence it is shown. ■

APPENDIX B. GENERAL TRANSFORMATION KERNEL FOR 3D OPTICAL SYSTEMS WAVE PROFILE

Here we shall motivate and derive from the optical relations a general formula for LCT.

Under the effect of a paraxial optical system, a light field of $U_0(\boldsymbol{x})$ is transformed into $U(\boldsymbol{x}')$ by the point-spread function $g(\boldsymbol{x}', \boldsymbol{x})$:

$$U(\boldsymbol{x}') = \int_{\mathbb{R}^2} g(\boldsymbol{x}', \boldsymbol{x}) U_0(\boldsymbol{x}) \, \mathrm{d}^2 x.$$

To relate the field with the rays, we shall turn to Wigner distribution function, and see how it transforms. Since we have

$$\begin{aligned}
W_U(\boldsymbol{x}', \boldsymbol{p}') &= \int_{\mathbb{R}^2} U(\boldsymbol{x}' + \boldsymbol{\xi}'/2) U^*(\boldsymbol{x}' - \boldsymbol{\xi}'/2) \mathrm{e}^{-i2\pi \boldsymbol{p}'^\mathsf{T} \boldsymbol{\xi}'} \, \mathrm{d}^2 \xi' \\
&= \int_{\mathbb{R}^2} \left(\int_{\mathbb{R}^2} g(\boldsymbol{x}' + \boldsymbol{\xi}'/2, \boldsymbol{x}_1) U_0(\boldsymbol{x}_1) \, \mathrm{d}^2 x_1 \right) \left(\int_{\mathbb{R}^2} g^*(\boldsymbol{x}' - \boldsymbol{\xi}'/2, \boldsymbol{x}_2) U_0^*(\boldsymbol{x}_2) \, \mathrm{d}^2 x_2 \right) \mathrm{e}^{-i2\pi \boldsymbol{p}'^\mathsf{T} \boldsymbol{\xi}'} \, \mathrm{d}^2 \xi'.
\end{aligned}$$

Substitute $\boldsymbol{x}_1 = \boldsymbol{x} + \boldsymbol{\xi}/2$ and $\boldsymbol{x}_2 = \boldsymbol{x} - \boldsymbol{\xi}/2$, with Jacobian equal to 1, we have

$$\begin{aligned}
W_U(\boldsymbol{x}', \boldsymbol{p}') &= \int_{\mathbb{R}^6} g(\boldsymbol{x}' + \boldsymbol{\xi}'/2, \boldsymbol{x} + \boldsymbol{\xi}/2) g^*(\boldsymbol{x}' - \boldsymbol{\xi}'/2, \boldsymbol{x} - \boldsymbol{\xi}/2) \cdot \mathrm{e}^{-i2\pi \boldsymbol{p}'^\mathsf{T} \boldsymbol{\xi}'} \cdot U_0(\boldsymbol{x} + \boldsymbol{\xi}/2) U_0^*(\boldsymbol{x} - \boldsymbol{\xi}/2) \, \mathrm{d}^2 \xi' \, \mathrm{d}^2 x \, \mathrm{d}^2 \xi \\
&= \int_{\mathbb{R}^6} g(\boldsymbol{x}' + \boldsymbol{\xi}'/2, \boldsymbol{x} + \boldsymbol{\xi}/2) g^*(\boldsymbol{x}' - \boldsymbol{\xi}'/2, \boldsymbol{x} - \boldsymbol{\xi}/2) \cdot \mathrm{e}^{-i2\pi \boldsymbol{p}'^\mathsf{T} \boldsymbol{\xi}'} \\
&\qquad \cdot \left(\int_{\mathbb{R}^2} U_0(\boldsymbol{x} + \boldsymbol{\eta}/2) U_0^*(\boldsymbol{x} - \boldsymbol{\eta}/2) \delta(\boldsymbol{\eta} - \boldsymbol{\xi}) \, \mathrm{d}^2 \eta \right) \mathrm{d}^2 \xi' \, \mathrm{d}^2 x \, \mathrm{d}^2 \xi \\
&= \int_{\mathbb{R}^8} g(\boldsymbol{x}' + \boldsymbol{\xi}'/2, \boldsymbol{x} + \boldsymbol{\xi}/2) g^*(\boldsymbol{x}' - \boldsymbol{\xi}'/2, \boldsymbol{x} - \boldsymbol{\xi}/2) \cdot \mathrm{e}^{-i2\pi \boldsymbol{p}'^\mathsf{T} \boldsymbol{\xi}'} \\
&\qquad \cdot U_0(\boldsymbol{x} + \boldsymbol{\eta}/2) U_0^*(\boldsymbol{x} - \boldsymbol{\eta}/2) \left(\int_{\mathbb{R}^2} \mathrm{e}^{-i2\pi \boldsymbol{p}^\mathsf{T} (\boldsymbol{\eta} - \boldsymbol{\xi})} \, \mathrm{d}^2 p \right) \mathrm{d}^2 \eta \, \mathrm{d}^2 \xi' \, \mathrm{d}^2 x \, \mathrm{d}^2 \xi \\
&= \int_{\mathbb{R}^4} \left(\int_{\mathbb{R}^4} g(\boldsymbol{x}' + \boldsymbol{\xi}'/2, \boldsymbol{x} + \boldsymbol{\xi}/2) g^*(\boldsymbol{x}' - \boldsymbol{\xi}'/2, \boldsymbol{x} - \boldsymbol{\xi}/2) \cdot \mathrm{e}^{-i2\pi (\boldsymbol{p}'^\mathsf{T} \boldsymbol{\xi}' - \boldsymbol{p}^\mathsf{T} \boldsymbol{\xi}} \, \mathrm{d}^2 \xi' \, \mathrm{d}^2 \xi \right) \\
&\qquad \cdot \left(\int_{\mathbb{R}^2} U_0(\boldsymbol{x} + \boldsymbol{\eta}/2) U_0^*(\boldsymbol{x} - \boldsymbol{\eta}/2) \mathrm{e}^{-i2\pi \boldsymbol{p}^\mathsf{T} \boldsymbol{\eta}} \, \mathrm{d}^2 \eta \right) \mathrm{d}^2 x \, \mathrm{d}^2 p \\
&=: \int_{\mathbb{R}^4} K(\boldsymbol{x}', \boldsymbol{p}', \boldsymbol{x}, \boldsymbol{p}) \cdot W_{U_0}(\boldsymbol{x}, \boldsymbol{p}) \, \mathrm{d}^2 x \, \mathrm{d}^2 p.
\end{aligned} \tag{B.1}$$

We hence obtain the transformation kernel (i.e., the ray-spread function) of the Wigner distribution function as

$$K(\boldsymbol{x}', \boldsymbol{p}', \boldsymbol{x}, \boldsymbol{p}) = \int_{\mathbb{R}^4} g(\boldsymbol{x}' + \boldsymbol{\xi}'/2, \boldsymbol{x} + \boldsymbol{\xi}/2) g^*(\boldsymbol{x}' - \boldsymbol{\xi}'/2, \boldsymbol{x} - \boldsymbol{\xi}/2) \mathrm{e}^{-i2\pi (\boldsymbol{p}'^\mathsf{T} \boldsymbol{\xi}' - \boldsymbol{p}^\mathsf{T} \boldsymbol{\xi})} \, \mathrm{d}^2 \xi' \, \mathrm{d}^2 \xi. \tag{B.2}$$

If we further assume that (1) our system is a quadratic-phase system, viz., paraxial approximation is assumed, and (2) the ray $[0, 0, 0, 0]^\mathsf{T}$ doesn't change after passing through the system, then we can set

$$g(\boldsymbol{x}', \boldsymbol{x}) = \theta_M \cdot \exp\left(\frac{i}{2} \begin{bmatrix} \boldsymbol{x}'^\mathsf{T} & \boldsymbol{x}^\mathsf{T} \end{bmatrix} \begin{bmatrix} L_{oo} & L_{oi} \\ L_{io} & L_{ii} \end{bmatrix} \begin{bmatrix} \boldsymbol{x}' \\ \boldsymbol{x} \end{bmatrix} \right). \tag{B.3}$$

Where the matrix

$$L = \begin{bmatrix} L_{oo} & L_{oi} \\ L_{io} & L_{ii} \end{bmatrix} \in \mathbb{R}^{4 \times 4} \tag{B.4}$$

is symmetric, and θ_M is a coefficient depending on the ABCD matrix M. Then the ray-spread function will be

$$\begin{aligned}
K(\boldsymbol{x}', \boldsymbol{p}', \boldsymbol{x}, \boldsymbol{p}) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \exp\left(i \left(\boldsymbol{x}'^\mathsf{T} L_{oo} \boldsymbol{\xi}' + \boldsymbol{x}'^\mathsf{T} L_{oi} \boldsymbol{\xi} + \boldsymbol{x}^\mathsf{T} L_{io} \boldsymbol{\xi}' + \boldsymbol{x}^\mathsf{T} L_{ii} \boldsymbol{\xi} \right) \right) \mathrm{e}^{-i2\pi (\boldsymbol{p}'^\mathsf{T} \boldsymbol{\xi}' - \boldsymbol{p}^\mathsf{T} \boldsymbol{\xi})} \, \mathrm{d}^2 \xi' \, \mathrm{d}^2 \xi \\
&= \delta\left(\boldsymbol{p}' - \frac{1}{2\pi} (L_{oo} \boldsymbol{x}' + L_{oi} \boldsymbol{x}) \right) \times \delta\left(\boldsymbol{p} + \frac{1}{2\pi} (L_{io} \boldsymbol{x}' + L_{ii} \boldsymbol{x}) \right).
\end{aligned} \tag{B.5}$$

But from the matrix optics formalism, we know that the following relation must hold:

$$\begin{bmatrix} \boldsymbol{x}' \\ \frac{\lambda}{2\pi} (L_{oo} \boldsymbol{x}' + L_{oi} \boldsymbol{x}) \end{bmatrix} = \begin{bmatrix} \mathsf{A} & \mathsf{B} \\ \mathsf{C} & \mathsf{D} \end{bmatrix} \begin{bmatrix} \boldsymbol{x} \\ -\frac{\lambda}{2\pi} (L_{io} \boldsymbol{x}' + L_{ii} \boldsymbol{x}) \end{bmatrix}, \tag{B.6}$$

hence,

$$\begin{cases} \boldsymbol{x}' = \mathsf{A}\boldsymbol{x} - \frac{\lambda}{2\pi}\mathsf{B}L_{io}\boldsymbol{x}' - \frac{\lambda}{2\pi}\mathsf{B}L_{ii}\boldsymbol{x} \\ \frac{\lambda}{2\pi}L_{oo}\boldsymbol{x}' + \frac{\lambda}{2\pi}L_{oi}\boldsymbol{x} = \mathsf{C}\boldsymbol{x} - \frac{\lambda}{2\pi}\mathsf{D}L_{io}\boldsymbol{x}' - \frac{\lambda}{2\pi}\mathsf{D}L_{ii}\boldsymbol{x} \end{cases}$$

holds for all pairs of $(\boldsymbol{x}, \boldsymbol{x}')$. The relations reduce further to

$$\begin{cases} 2\pi\mathbb{1} = -\mathsf{B}L_{io} \\ 2\pi\mathsf{A} - \mathsf{B}L_{ii} = 0 \\ L_{oo} = -\mathsf{D}L_{io} \\ L_{oi} = 2\pi\mathsf{C} - \mathsf{D}L_{ii} \end{cases}$$

The four relations translates simply to

$$L_{io} = -\frac{2\pi}{\lambda}\mathsf{B}^{-1} \tag{B.7}$$

$$L_{ii} = \frac{2\pi}{\lambda}\mathsf{B}^{-1}\mathsf{A} \tag{B.8}$$

$$L_{oo} = \frac{2\pi}{\lambda}\mathsf{D}\mathsf{B}^{-1} \tag{B.9}$$

$$L_{oi} = \frac{2\pi}{\lambda}(\mathsf{C} - \mathsf{D}\mathsf{A}^{\mathsf{T}}\mathsf{B}^{-\mathsf{T}}). \tag{B.10}$$

Moreover, since ${L_{ii}}^{\mathsf{T}} = L_{ii}$, ${L_{oo}}^{\mathsf{T}} = L_{oo}$, and ${L_{io}}^{\mathsf{T}} = L_{oi}$, we have

$$\mathsf{B}^{-1}\mathsf{A} = \mathsf{A}^{\mathsf{T}}\mathsf{B}^{-\mathsf{T}} \tag{B.11}$$

$$\mathsf{D}\mathsf{B}^{-1} = \mathsf{A}^{\mathsf{T}}\mathsf{B}^{-\mathsf{T}} \tag{B.12}$$

$$-\mathsf{B}^{-\mathsf{T}} = \mathsf{C} - \mathsf{D}\mathsf{A}^{\mathsf{T}}\mathsf{B}^{-\mathsf{T}} \to \mathsf{A}\mathsf{D}^{\mathsf{T}} - \mathsf{B}\mathsf{C}^{\mathsf{T}} = \mathbb{1}. \tag{B.13}$$

The relation (B.13) is termed the *symplectic relation*. Matrices of size $2n \times 2n$ that satisfies the symplectic relation form the group of *symplectic matrices*, denoted as $\mathrm{Sp}(2n, \mathbb{R})$.

From the discussions above, we arrived at a partial result of

$$\begin{aligned} g(\boldsymbol{x}', \boldsymbol{x}) &= \theta_M \cdot \exp\left(\frac{i}{2}\begin{bmatrix}{\boldsymbol{x}'}^{\mathsf{T}} & \boldsymbol{x}^{\mathsf{T}}\end{bmatrix}\begin{bmatrix}L_{oo} & L_{oi} \\ L_{io} & L_{ii}\end{bmatrix}\begin{bmatrix}\boldsymbol{x}' \\ \boldsymbol{x}\end{bmatrix}\right) \\ &= \theta_M \cdot \exp\left(i\frac{\pi}{\lambda}{\boldsymbol{x}'}^{\mathsf{T}}\mathsf{D}\mathsf{B}^{-1}\boldsymbol{x}' - i\frac{2\pi}{\lambda}\boldsymbol{x}^{\mathsf{T}}\mathsf{B}^{-1}\boldsymbol{x}' + i\frac{\pi}{\lambda}x^{\mathsf{T}}\mathsf{B}^{-1}\mathsf{A}\boldsymbol{x}\right). \end{aligned} \tag{B.14}$$

Our task now would be to determine the coefficient θ_M, it should be chosen so that the additivity of the transform holds. Consider the composition of transforms:

$$\begin{aligned} \int_{\mathbb{R}^2} g_{M_2}(\boldsymbol{x}'', \boldsymbol{x}')g_{M_1}(\boldsymbol{x}', \boldsymbol{x})\,\mathrm{d}^2x' &= \theta_{M_2}\theta_{M_1}\exp\left(i\frac{\pi}{\lambda}{\boldsymbol{x}''}^{\mathsf{T}}\mathsf{D}_2\mathsf{B}_2^{-1}\boldsymbol{x}'' + i\frac{\pi}{\lambda}\boldsymbol{x}^{\mathsf{T}}\mathsf{B}_1^{-1}\mathsf{A}_1\boldsymbol{x}\right) \\ &\quad\cdot\int_{\mathbb{R}^2}\exp\left(i\frac{\pi}{\lambda}{\boldsymbol{x}'}^{\mathsf{T}}(\mathsf{B}_2^{-1}\mathsf{A}_2 + \mathsf{D}_1\mathsf{B}_1^{-1})\boldsymbol{x}' - \imath\frac{2\pi}{\lambda}({\boldsymbol{x}''}^{\mathsf{T}}\mathsf{B}_2^{-\mathsf{T}} + \boldsymbol{x}^{\mathsf{T}}\mathsf{B}_1^{-1})\boldsymbol{x}'\right)\,\mathrm{d}^2x' \\ &\stackrel{?}{=} \theta_M\exp\left(i\frac{\pi}{\lambda}\boldsymbol{x}^{\mathsf{T}}(\mathsf{A}_2\mathsf{B}_1 + \mathsf{B}_2\mathsf{D}_1)^{-1}(\mathsf{A}_2\mathsf{A}_1 + \mathsf{B}_2\mathsf{C}_1)\boldsymbol{x} - i\frac{2\pi}{\lambda}\boldsymbol{x}^{\mathsf{T}}(\mathsf{A}_2\mathsf{B}_1 + \mathsf{B}_2\mathsf{D}_1)^{-1}\boldsymbol{x}''\right. \\ &\quad\left.+i\frac{\pi}{\lambda}(\mathsf{C}_2\mathsf{B}_1 + \mathsf{D}_2\mathsf{D}_1)(\mathsf{A}_2\mathsf{B}_1 + \mathsf{B}_2\mathsf{D}_1)^{-1}\boldsymbol{x}''\right), \end{aligned}$$

with $M = M_2M_1$. The integral above is of the form

$$\begin{aligned} I(\mathbb{A}, \boldsymbol{\beta}) &= \int_{\mathbb{R}^2}\exp(i{\boldsymbol{x}'}^{\mathsf{T}}\mathbb{A}\boldsymbol{x}' - 2i\boldsymbol{\beta}^{\mathsf{T}}\boldsymbol{x}')\,\mathrm{d}^2x' \\ &= \exp\left(i\left(4\boldsymbol{\beta}^{\mathsf{T}}(\mathbb{A} + \mathbb{A}^{\mathsf{T}})^{-1}\mathbb{A}(\mathbb{A} + \mathbb{A}^{\mathsf{T}})^{-1}\boldsymbol{\beta} - 4\boldsymbol{\beta}^{\mathsf{T}}(\mathbb{A} + \mathbb{A}^{\mathsf{T}})^{-1}\boldsymbol{\beta}\right)\right)\cdot\frac{i\pi}{\sqrt{\det(\mathbb{A})}}, \end{aligned} \tag{B.15}$$

where $\mathbb{A} \in \mathbb{R}^{2\times 2}$ and $\boldsymbol{\beta} \in \mathbb{R}^2$. Apply eqn.(B.15) to the composition of transforms above:

$$\begin{aligned}
\int_{\mathbb{R}^2} g_{M_2}(\boldsymbol{x}'', \boldsymbol{x}') g_{M_1}(\boldsymbol{x}', \boldsymbol{x})\, \mathrm{d}^2 x' &= \theta_{M_2}\theta_{M_1} \exp\left(i\frac{\pi}{\lambda}\boldsymbol{x}''^{\mathsf{T}} \mathsf{D}_2\mathsf{B}_2^{-1}\boldsymbol{x}'' + i\frac{\pi}{\lambda}\boldsymbol{x}^{\mathsf{T}}\mathsf{B}_1^{-1}\mathsf{A}_1\boldsymbol{x} \right) \\
&\quad \cdot I\left(\frac{\pi}{\lambda}(\mathsf{B}_2^{-1}\mathsf{A}_2 + \mathsf{D}_1\mathsf{B}_1^{-1}), \frac{\pi}{\lambda}(\mathsf{B}_2^{-1}\boldsymbol{x}'' + \mathsf{B}_1^{-\mathsf{T}}\boldsymbol{x}) \right) \\
&= \theta_{M_2}\theta_{M_1} \cdot \exp\left(i\frac{\pi}{\lambda}\boldsymbol{x}''^{\mathsf{T}} \mathsf{D}_2\mathsf{B}_2^{-1}\boldsymbol{x}'' + i\frac{\pi}{\lambda}\boldsymbol{x}^{\mathsf{T}}\mathsf{B}_1^{-1}\mathsf{A}_1\boldsymbol{x} \right) \cdot \frac{i\lambda}{\sqrt{\det(\mathsf{B}_2^{-1}\mathsf{A}_2 + \mathsf{D}_1\mathsf{B}^{-1})}} \\
&\quad \cdot \exp\left(-i\frac{\pi}{\lambda}(\boldsymbol{x}''^{\mathsf{T}}\mathsf{B}_2^{-\mathsf{T}} + \boldsymbol{x}^{\mathsf{T}}\mathsf{B}_1^{-1})(\mathsf{B}_2^{-1}\mathsf{A}_2 + \mathsf{D}_1\mathsf{B}_1^{-1})^{-1}(\mathsf{B}_2^{-1}\boldsymbol{x}'' + \mathsf{B}_1^{-\mathsf{T}}\boldsymbol{x}) \right) \\
&= \theta_{M_2}\theta_{M_1} \cdot i\lambda \sqrt{\frac{\det(\mathsf{B}_2)\det(B_1)}{\det(\mathsf{A}_2\mathsf{B}_1 + \mathsf{B}_2\mathsf{D}_1)}} \\
&\quad \cdot \exp\left(i\frac{\pi}{\lambda}\left[\boldsymbol{x}''^{\mathsf{T}} \underbrace{\left(\mathsf{D}_2\mathsf{B}_2^{-1} - \mathsf{B}_2^{-\mathsf{T}}(\mathsf{B}_2^{-1}\mathsf{A}_2 + \mathsf{D}_1\mathsf{B}_1^{-1})^{-1}\mathsf{B}_2^{-1}\right)}_{\text{(a)}} \boldsymbol{x}'' \right.\right. \\
&\qquad + \boldsymbol{x}^{\mathsf{T}} \underbrace{\left(\mathsf{B}_1^{-1}\mathsf{A}_1 - \mathsf{B}_1^{-1}(\mathsf{B}_2^{-1}\mathsf{A}_2 + \mathsf{D}_1\mathsf{B}_1^{-1})^{-1}\mathsf{B}_1^{-\mathsf{T}}\right)}_{\text{(b)}} \boldsymbol{x} \\
&\qquad \left.\left. -2\boldsymbol{x}^{\mathsf{T}} \underbrace{\left(\mathsf{B}_1^{-1}(\mathsf{B}_2^{-1}\mathsf{A}_2 + \mathsf{D}_1\mathsf{B}_1^{-1})^{-1}\mathsf{B}_2^{-1}\right)}_{\text{(c)}} \boldsymbol{x}'' \right]\right).
\end{aligned}$$

To let the additivity of transforms hold, we must let

$$\theta_M = \frac{1}{i\lambda\sqrt{\det(\mathsf{B})}}. \tag{B.16}$$

Furthermore, to make sure that the additivity holds, we need to check whether the exponents satisfy the additivity, too. First note that

$$M = \begin{bmatrix} \mathsf{A} & \mathsf{B} \\ \mathsf{C} & \mathsf{D} \end{bmatrix} = \begin{bmatrix} \mathsf{A}_2\mathsf{A}_1 + \mathsf{B}_2\mathsf{C}_1 & \mathsf{A}_2\mathsf{B}_1 + \mathsf{B}_2\mathsf{D}_1 \\ \mathsf{C}_2\mathsf{A}_1 + \mathsf{D}_2\mathsf{C}_1 & \mathsf{C}_2\mathsf{B}_1 + \mathsf{D}_2\mathsf{D}_1 \end{bmatrix} = M_2 M_1.$$

Then,

$$\text{(c)} = \left(\mathsf{B}_2(\mathsf{B}_2^{-1}\mathsf{A}_2 + \mathsf{D}_1\mathsf{B}_1^{-1})\mathsf{B}_1\right)^{-1} = (\mathsf{A}_2\mathsf{B}_1 + \mathsf{B}_2\mathsf{D}_1)^{-1} = \mathsf{B}^{-1};$$

$$\begin{aligned}
\text{(a)} &= \mathsf{B}_2^{-\mathsf{T}}\mathsf{D}_2{}^{\mathsf{T}} - \mathsf{B}_2^{-\mathsf{T}}\mathsf{B}_1(\mathsf{A}_2\mathsf{B}_1 + \mathsf{B}_2\mathsf{D}_1)^{-1} = \mathsf{B}_2^{-\mathsf{T}}\left[\mathsf{D}_2{}^{\mathsf{T}}(\mathsf{A}_2\mathsf{B}_1 + \mathsf{B}_2\mathsf{D}_1) - \mathsf{B}_1\right](\mathsf{A}_2\mathsf{B}_1 + \mathsf{B}_2\mathsf{D}_1)^{-1} \\
&= (\mathsf{B}_2^{-\mathsf{T}}\mathsf{D}_2{}^{\mathsf{T}}\mathsf{A}_2\mathsf{B}_1 + \mathsf{D}_2\mathsf{B}_2^{-1}\mathsf{B}_2\mathsf{D}_1 - \mathsf{B}_2^{-\mathsf{T}}\mathsf{B}_1)(\mathsf{A}_2\mathsf{B}_1 + \mathsf{B}_2\mathsf{D}_1)^{-1} = \left[\mathsf{B}_2^{-\mathsf{T}}(\mathsf{D}_2{}^{\mathsf{T}}\mathsf{A}_2 - \mathbb{1}\mathsf{B}_1 + \mathsf{D}_2\mathsf{D}_1\right](\mathsf{A}_2\mathsf{B}_1 + \mathsf{B}_2\mathsf{D}_1)^{-1} \\
&= (\mathsf{C}_2\mathsf{B}_1 + \mathsf{D}_2\mathsf{D}_1)(\mathsf{A}_2\mathsf{B}_1 + \mathsf{B}_2\mathsf{D}_1)^{-1} = \mathsf{D}\mathsf{B}^{-1}.
\end{aligned}$$

$$\begin{aligned}
\text{(b)} &= \mathsf{A}_1{}^{\mathsf{T}}\mathsf{B}_1^{-\mathsf{T}} - (\mathsf{A}_2\mathsf{B}_1 + \mathsf{B}_2\mathsf{D}_1)^{-1}\mathsf{B}_2\mathsf{B}_1^{-\mathsf{T}} = (\mathsf{A}_2\mathsf{B}_1 + \mathsf{B}_2\mathsf{D}_1)^{-1}\left[(\mathsf{A}_2\mathsf{B}_1 + \mathsf{B}_2\mathsf{D}_1)\mathsf{A}_1{}^{\mathsf{T}} - \mathsf{B}_2\right]\mathsf{B}_1^{-\mathsf{T}} \\
&= (\mathsf{A}_2\mathsf{B}_1 + \mathsf{B}_2\mathsf{D}_1)^{-1}(\mathsf{A}_2\mathsf{B}_1\mathsf{B}_1^{-1}\mathsf{A}_1 + \mathsf{B}_2\mathsf{D}_1\mathsf{A}_1{}^{\mathsf{T}}\mathsf{B}_1^{-\mathsf{T}} - \mathsf{B}_2\mathsf{B}_1^{-\mathsf{T}}) = (\mathsf{A}_2\mathsf{B}_1 + \mathsf{B}_2\mathsf{D}_1)^{-1}\left[\mathsf{A}_2\mathsf{A}_1 + \mathsf{B}_2(\mathsf{D}_1\mathsf{A}_1{}^{\mathsf{T}} - \mathbb{1})\mathsf{B}_1^{-\mathsf{T}}\right] \\
&= (\mathsf{A}_2\mathsf{B}_1 + \mathsf{B}_2\mathsf{D}_1)^{-1}(\mathsf{A}_2\mathsf{A}_1 + \mathsf{B}_2\mathsf{C}_1) = \mathsf{B}^{-1}\mathsf{A}.
\end{aligned}$$

Hence we show that for $\det(\mathsf{B}) \neq 0$, the additivity holds for the point-spread function of the form

$$g(\boldsymbol{x}', \boldsymbol{x}) = \frac{1}{i\lambda\sqrt{\det(\mathsf{B})}} \exp\left(i\frac{\pi}{\lambda}\boldsymbol{x}^{\mathsf{T}}\mathsf{B}^{-1}\mathsf{A}\boldsymbol{x} - i\frac{2\pi}{\lambda}\boldsymbol{x}^{\mathsf{T}}\mathsf{B}^{-1}\boldsymbol{x}' + i\frac{\pi}{\lambda}\boldsymbol{x}'^{\mathsf{T}}\mathsf{D}\mathsf{B}^{-1}\boldsymbol{x}' \right). \tag{B.17}$$

As for the case where $\det(\mathsf{B}) = 0$, first note that the delta function can be defined as the limit to the multivariate (two-dimensional) Gaussian:

$$\delta(\boldsymbol{x} - \boldsymbol{\mu}) = \lim_{|\Sigma| \to 0} \frac{1}{\pi\sqrt{\det(\Sigma)}} \mathrm{e}^{-(\boldsymbol{x}-\boldsymbol{\mu})^\mathsf{T}\Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu})}. \tag{B.18}$$

Hence, by factoring:

$$\begin{aligned} g(\boldsymbol{x}', \boldsymbol{x}) &= \frac{\sqrt{\det(\mathsf{A}^{-1})}}{\pi\sqrt{\det\left(\frac{i\lambda}{\pi}\mathsf{A}^{-1}\mathsf{B}\right)}} \exp\left(-(\boldsymbol{x} - \mathsf{A}^{-1}\boldsymbol{x}')^\mathsf{T}\left(\frac{i\lambda}{\pi}\mathsf{A}^{-1}\mathsf{B}\right)^{-1}(\boldsymbol{x} - \mathsf{A}^{-1}\boldsymbol{x}')\right) \exp\left(i\frac{\pi}{\lambda}\boldsymbol{x}'^\mathsf{T}\underbrace{(\mathsf{D}\mathsf{B}^{-1} - \mathsf{A}^{-\mathsf{T}}\mathsf{B}^{-1})}_{=\mathsf{C}\mathsf{A}^{-1}}\right) \\ &\xrightarrow{|\Sigma| \to 0} \sqrt{\det \mathsf{A}^{-1}}\, \delta(\boldsymbol{x} - \mathsf{A}^{-1}\boldsymbol{x}')\ \exp\left(i\frac{\pi}{\lambda}\boldsymbol{x}'^\mathsf{T}\mathsf{C}\mathsf{A}^{-1}\boldsymbol{x}'\right) \qquad \text{(B.19)} \\ &= \sqrt{\det \mathsf{D}}\, \delta(\boldsymbol{x} - \mathsf{D}^\mathsf{T}\boldsymbol{x}')\ \exp\left(i\frac{\pi}{\lambda}\boldsymbol{x}'^\mathsf{T}\mathsf{C}\mathsf{D}^\mathsf{T}\boldsymbol{x}'\right). \qquad \text{(B.20)} \end{aligned}$$

Thus it is shown. ■

Appendix B.1. Generalized LCT

Note that the generalized LCT $\tilde{\mathcal{O}}_M\{f(\boldsymbol{x})\}(\boldsymbol{x}')$ for unitless $\boldsymbol{x}$ and $\boldsymbol{x}' \in \mathbb{R}^n$ with unitless

$$M = \begin{bmatrix} \mathsf{A} & \mathsf{B} \\ \mathsf{C} & \mathsf{D} \end{bmatrix} \in \mathrm{Sp}(2n, \mathbb{R})$$

is defined as

$$\tilde{\mathcal{O}}_{\left[\begin{smallmatrix} \mathsf{A} & \mathsf{B} \\ \mathsf{C} & \mathsf{D} \end{smallmatrix}\right]}\{f(\boldsymbol{x})\}(\boldsymbol{x}') := \begin{cases} \dfrac{1}{\sqrt{i^n \det(\mathsf{B})}} \displaystyle\int_{\mathbb{R}^n} \exp\left(i\pi\boldsymbol{x}^\mathsf{T}\mathsf{B}^{-1}\mathsf{A}\boldsymbol{x} - i2\pi\boldsymbol{x}^\mathsf{T}\mathsf{B}^{-1}\boldsymbol{x}' + i\pi\boldsymbol{x}'^\mathsf{T}\mathsf{D}\mathsf{B}^{-1}\boldsymbol{x}'\right) f(\boldsymbol{x})\, \mathrm{d}^n x & , \det(\mathsf{B}) \neq 0 \\ \dfrac{1}{\sqrt{\det(\mathsf{A})}} \mathrm{e}^{i\pi\boldsymbol{x}'^\mathsf{T}\mathsf{C}\mathsf{A}^{-1}\boldsymbol{x}'} f(\mathsf{A}^{-1}\boldsymbol{x}') & , \det(\mathsf{B}) = 0 \end{cases}. \tag{B.21}$$

The proof is trivial and follows similar steps as the prior.

APPENDIX C. PROPERTIES OF WIGNER DISTRIBUTION FUNCTION

Appendix C.1. Projective Property

For a function $f(\boldsymbol{x}) : \mathbb{R}^n \to \mathbb{R}$, we have

$$\begin{aligned}\int_{\mathbb{R}^n} W_f(\boldsymbol{x},\boldsymbol{p})\,\mathrm{d}^n p &= \int_{\mathbb{R}^n}\int_{\mathbb{R}^n} f(\boldsymbol{x}+\boldsymbol{\xi}/2)f^*(\boldsymbol{x}-\boldsymbol{\xi}/2)\mathrm{e}^{-i2\pi\boldsymbol{p}^{\mathsf{T}}\boldsymbol{\xi}}\,\mathrm{d}^n\xi\,\mathrm{d}^n p\\ &= \int_{\mathbb{R}^n} f(\boldsymbol{x}+\boldsymbol{\xi}/2)f^*(\boldsymbol{x}-\boldsymbol{\xi}/2)\left(\int_{\mathbb{R}^n}\mathrm{e}^{-i2\pi\boldsymbol{p}^{\mathsf{T}}\boldsymbol{\xi}}\,\mathrm{d}^n p\right)\mathrm{d}^n\xi\\ &= \int_{\mathbb{R}^n} f(\boldsymbol{x}+\boldsymbol{\xi}/2)f^*(\boldsymbol{x}-\boldsymbol{\xi}/2)\cdot\delta(\boldsymbol{\xi})\,\mathrm{d}^n\xi\\ &= f(\boldsymbol{x})f^*(\boldsymbol{x}) = |f(\boldsymbol{x})|^2.\end{aligned}$$

Thus it is proven. ■

Appendix C.2. Relation with LCT

We shall prove eqn.(53), the effect of LCT on the Wigner distribution function. Similarly, we shall denote $\boldsymbol{q} = [\boldsymbol{x}^{\mathsf{T}}, \lambda\boldsymbol{p}^{\mathsf{T}}]^{\mathsf{T}}$ and

$$M = \begin{bmatrix}\mathsf{A} & \mathsf{B}\\ \mathsf{C} & \mathsf{D}\end{bmatrix} \in \mathrm{Sp}(2n,\mathbb{R}).$$

It can be shown that

$$M^{-1} = \begin{bmatrix}\mathsf{D}^{\mathsf{T}} & -\mathsf{B}^{\mathsf{T}}\\ -\mathsf{C}^{\mathsf{T}} & \mathsf{A}^{\mathsf{T}}\end{bmatrix}.$$

First notice that the derivations from eqn.(B.1) all the way up to eqn.(B.10) doesn't depend on the dimension n, hence they hold for all $n \geq 1$. Then by combining eqns.(B.1) and (A.7), we have:

$$\begin{aligned}W_{\tilde{\mathcal{O}}_M\{f\}}(\boldsymbol{q}') &= \int_{\mathbb{R}^{2n}} \underbrace{\delta\left(\boldsymbol{p}' - (\mathsf{D}\mathsf{B}^{-1}\boldsymbol{x}' - \mathsf{B}^{-\mathsf{T}}\boldsymbol{x})/\lambda\right)}_{=\frac{\lambda^n}{\det(\mathsf{B}^{-1})}\delta(\boldsymbol{x}-\mathsf{D}^{\mathsf{T}}\boldsymbol{x}'+\lambda\mathsf{B}^{\mathsf{T}}\boldsymbol{p}')} \times \underbrace{\delta\left(\boldsymbol{p} + (-\mathsf{B}^{-1}\boldsymbol{x}' + \mathsf{B}^{-1}\mathsf{A}\boldsymbol{x})/\lambda\right)}_{=\delta(\boldsymbol{p}+\mathsf{B}^{-1}\mathsf{A}\boldsymbol{x}/\lambda-\mathsf{B}^{-1}\boldsymbol{x}'/\lambda)} \cdot W_f(\boldsymbol{x},\boldsymbol{p})\,\mathrm{d}^n x\,\mathrm{d}^n p\\ &= \lambda^n\det(\mathsf{B})\int_{\mathbb{R}^n}\delta(\boldsymbol{p}+\mathsf{C}^{\mathsf{T}}\boldsymbol{x}'/\lambda - \lambda\mathsf{A}^{\mathsf{T}}\boldsymbol{p}'/\lambda)W_f(\mathsf{D}^{\mathsf{T}}\boldsymbol{x}' - \lambda\mathsf{B}^{\mathsf{T}}\boldsymbol{p}',\boldsymbol{p})\,\mathrm{d}^n p\\ &= W_f(\mathsf{D}^{\mathsf{T}}\boldsymbol{x}' - \lambda\mathsf{B}^{\mathsf{T}}\boldsymbol{p}', -\mathsf{C}^{\mathsf{T}}\boldsymbol{x}'/\lambda + \boldsymbol{A}^{\mathsf{T}}\boldsymbol{p}') = W_f(M^{-1}\boldsymbol{q}')\end{aligned}$$

Hence it is proven. ■

It should be noted that the derivation above uses an abuse of notation by writing

$$W_f(\boldsymbol{x},\boldsymbol{p}) \equiv W_f(\boldsymbol{q}) = W_f([\boldsymbol{x}^{\mathsf{T}}, \lambda\boldsymbol{p}^{\mathsf{T}}]^{\mathsf{T}}). \tag{C.1}$$