Matrix-based Fourier analysis of matrix signals and systems for polarization optics

Citation for published version:
Wang, W 2024, 'Matrix-based Fourier analysis of matrix signals and systems for polarization optics', *Journal of the Optical Society of America A: Optics and Image Science, and Vision*, vol. 41, no. 10, pp. 1969-1978. https://doi.org/10.1364/JOSAA.532592

Digital Object Identifier (DOI):
10.1364/JOSAA.532592

Published In:
Journal of the Optical Society of America A: Optics and Image Science, and Vision

Research Article | Vol. 41, No. 10 / October 2024 / Journal of the Optical Society of America A | 1969

Journal of the Optical Society of America A
OPTICS, IMAGE SCIENCE, AND VISION

Matrix-based Fourier analysis of matrix signals and systems for polarization optics

WEI WANG
Institute of Photonics and Quantum Sciences, School of Engineering and Physical Sciences, Heriot-Watt University, Edinburgh EH14 4AS, UK (w.wang@hw.ac.uk)

Received 11 June 2024; revised 21 August 2024; accepted 29 August 2024; posted 29 August 2024; published 26 September 2024

Matrix functions are, of course, indispensable and of primary concern in polarization optics when the vector nature of light has been considered. This paper is devoted to investigating matrix-based Fourier analysis of two-dimensional matrix signals and systems. With the aid of the linearity and the superposition integral of matrix functions, the theory of linear invariant matrix systems has been constructed by virtue of six matrix-based integral transformations [i.e., matrix (direct) convolution, matrix (direct) correlation, and matrix element-wise convolution/correlation]. Properties of the matrix-based Fourier transforms have been introduced with some applications including the identity impulse matrix, matrix sampling theorem, width, bandwidth and their uncertainty relation for the matrix signal, and Haagerup's inequality for matrix normalization. The coherence time and the effective spectral width of the stochastic electromagnetic wave have been discussed as an application example to demonstrate how to apply the proposed mathematical tools in analyzing polarization-dependent Fourier optics.

https://doi.org/10.1364/JOSAA.532592

1. INTRODUCTION

As an intrinsic property of transverse waves, light polarization specifies the geometrical orientation of the electromagnetic oscillations and has attracted increasing interest due to its theoretical and practical importance. It is common practice in physics and engineering to describe spatio-temporal evolution and/or the optical system's response to a complicated stimulus of electromagnetic wave with multiple components by a matrix function. In particular, the matrix-multiplication-based operations, known as Jones calculus [1–8] and Mueller calculus [9,10], have provided simple approaches for calculation of light polarization after passing through some polarization elements (i.e., polarizers, waveplates, and rotators), and many important problems in polarization optics can be treated in a unified way.

The theory of linear systems and the Fourier analysis are also powerful tools that have found applications to diverse areas in modern science and technology. These two methods play key roles as the underlying analytical structure for our treatment in Fourier optics [11]. Despite their great importance, the combination of Fourier optics and the matrix theory for polarization optics has not yet received due attention, and only limited efforts have been made to develop matrix-based Fourier optics by taking the vector nature of light into account. As for the fusion of Fourier analysis and linear algebra for light polarization, there is limited previous work in the literature [12]. A Jones-matrix impulse response function was originated by Urbańczyk [13]. Matrix versions of the point-spread function and optical transfer function have been developed for polarization optics by McGuire and Chipman [14,15]. A Jones matrix treatment for polarization Fourier optics has been proposed by Moreno and co-workers [16]. The concept of the transmission cross-coefficient matrix has been introduced in a monography by Korotkova [17]. More recently, a new formalism for calculating Stokes imaging with partially coherent and partially polarized light has been proposed with the aid of matrix-based integral transformations [18]. On the other hand, Papoulis [19], Bracewell [20], and Gray and Goodman [21], in their renowned books, have given comprehensive review of the fundamental mathematical concepts of Fourier analysis in one and/or two dimensions, but the analysis is based on scalar function.

The purpose of this paper is to present some of the mathematical tools that are useful in describing linear phenomena in polarization optics and discuss some of the mathematical decompositions to be often employed in Fourier analysis of light polarization. Based on the linearity and the superposition integral of matrix functions, we extend the concepts of signals and systems in scalar optics to corresponding matrix versions. A new concept, referred to as linear matrix systems theory, will be developed, and transfer matrices are introduced for linear invariant matrix systems. After giving the definitions for several matrix-based integral transformations [i.e., matrix (direct) convolution, matrix (direct) correlation, and matrix element-wise convolution/correlation], we present the corresponding algebra for these matrix signals and matrix systems. Matrix-based

Fourier analysis in two dimensions is discussed with their basic properties of the matrix-based Fourier transform theorems. Although the formalism to be conducted is for matrix functions of two independent variables, which are, of course, of primary concern in optics, the reduction from two to one variable for temporal matrix signals and systems is quite straightforward. We hope that the formalism presented in this paper will lay a theoretical foundation for developing matrix Fourier optics (matrix information optics) and matrix-based information processing for electromagnetic waves.

2. LINEAR MATRIX SYSTEMS

For the purpose of discussion in this paper, we seek to introduce the words of matrix signal and matrix system in a way sufficiently general to the case of polarization optics. As shown in Fig. 1, a matrix system is defined to be a mapping of a set of input matrix functions into a set of output matrix functions with the same types of physical signal (i.e., electromagnetic wave). In the context of polarization optics, the matrix systems are polarization-sensitive optical systems, and the matrix signals, including the matrix version of stimuli (matrix inputs) and the matrix version of responses (matrix outputs), may be electric field vectors, (generalized) Stokes vectors, polarization matrices, or mutual coherence matrices of stochastic electromagnetic waves [22,23]. If attention is restricted to deterministic (nonrandom) systems, then a specific input must map to a unique output.

A convenient representation of a matrix system is a matrix version of operator, $\mathcal{L}\{\cdots\}$, which we imagine to operate on input matrix functions to produce output matrix functions of the same kind of physical signals with the same matrix dimension. Thus, if the $m \times n$ matrix function $\mathbf{G}^{\mathrm{In}}(x_1, y_1)$ represents the input to the system, and the $m \times n$ matrix $\mathbf{G}^{\mathrm{Out}}(x_2, y_2)$ represents the corresponding output, then by the definition of $\mathcal{L}\{\cdots\}$, the two matrix functions are related through

$$\mathbf{G}^{\mathrm{Out}}(x_2, y_2) = \mathcal{L}\left\{\mathbf{G}^{\mathrm{In}}(x_1, y_1)\right\}. \tag{1}$$

Without specifying more detailed properties of the operator $\mathcal{L}\{\cdots\}$, it is difficult to state more specific properties of the matrix system. In this paper, we are concerned primarily with a restricted class of systems, which are said to be linear due to the fact that the linearity assumption will yield simple and physically meaningful representations of such systems. Just in the same way as any linear system, where the superposition property plays a critical role, we can write a linear matrix system for all input matrix functions $\mathbf{G}^{\mathrm{In}}$ and $\mathbf{H}^{\mathrm{In}}$

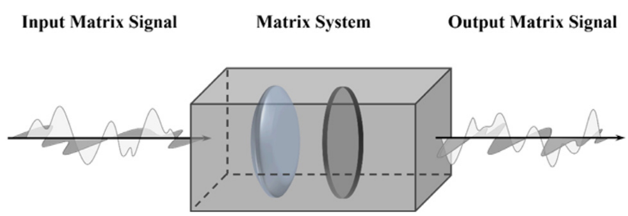

Fig. 1. Generalized model of matrix signals and a matrix system.

$$\mathcal{L}\left\{\alpha\mathbf{G}^{\mathrm{In}}(x, y) + \beta\mathbf{H}^{\mathrm{In}}(x, y)\right\} = \alpha\mathcal{L}\left\{\mathbf{G}^{\mathrm{In}}(x, y)\right\} + \beta\mathcal{L}\left\{\mathbf{H}^{\mathrm{In}}(x, y)\right\}, \tag{2}$$

where α and β are complex constants. The great advantage afforded by linearity is the ability to present the response of a matrix system to an arbitrary input in terms of the response to certain "elementary" matrix functions into which the input matrix signal has been decomposed. It is most important to find a simple and convenient means of decomposing the input in the same way as the case for a scalar signal; such a decomposition for a matrix signal is offered by the sifting property of the Dirac delta function and the identity matrix. That is,

$$\mathbf{G}^{\mathrm{In}}(x, y) = \iint_{-\infty}^{\infty} \boldsymbol{\delta}_m(x-\xi, y-\eta)\mathbf{G}^{\mathrm{In}}(\xi, \eta)\mathrm{d}\xi\,\mathrm{d}\eta, \tag{3}$$

where $\boldsymbol{\delta}_m(x, y) = \delta(x, y)\mathbf{I}_m$ is a $m \times m$ matrix referred to as the *identity impulse matrix* (IIM) or the *identity delta matrix* (IDM) with $\delta(x, y)$ being the Dirac delta function and $\mathbf{I}_m$ being the identity matrix of size m. Equation (3) may be regarded as expressing the input matrix signal $\mathbf{G}^{\mathrm{In}}$ as a linear combination of the displaced identity impulse matrices with the $m \times n$ matrices $\mathbf{G}^{\mathrm{In}}(\xi, \eta)$ being simply weighting matrices applied to the elementary matrices $\boldsymbol{\delta}_m(x_1-\xi, y_1-\eta)$ of the matrix decomposition. To find the response of the matrix system to the input matrix signal, we substitute Eq. (3) in Eq. (1) and invoke the linearity property in Eq. (2) to allow $\mathcal{L}\{\cdots\}$ to operate on the individual elementary matrices. Thus the operator can be brought within the integral, yielding

$$\mathbf{G}^{\mathrm{Out}}(x_2, y_2) = \iint_{-\infty}^{\infty} \mathcal{L}\left\{\boldsymbol{\delta}_m(x_1-\xi, y_1-\eta)\right\}\mathbf{G}^{\mathrm{In}}(\xi, \eta)\mathrm{d}\xi\,\mathrm{d}\eta. \tag{4}$$

Let the symbol $\mathbf{H}(x_2, y_2; \xi, \eta)$ denote the response of the matrix system at point (x_2, y_2) of the output space to a $\boldsymbol{\delta}_m$ matrix input at coordinates (ξ, η) of the input space, that is,

$$\mathbf{H}(x_2, y_2; \xi, \eta) = \mathcal{L}\left\{\boldsymbol{\delta}_m(x_1-\xi, y_1-\eta)\right\}. \tag{5}$$

The square matrix $\mathbf{H}$ with its dimension of $m \times m$ can be called the *impulse response matrix* or the *point-spread matrix* of the system. The system input and output can now be related by the simple equation

$$\mathbf{G}^{\mathrm{Out}}(x_2, y_2) = \iint_{-\infty}^{\infty} \mathbf{H}(x_2, y_2; \xi, \eta)\mathbf{G}^{\mathrm{In}}(\xi, \eta)\mathrm{d}\xi\,\mathrm{d}\eta. \tag{6}$$

The equation above can be understood as the matrix-based superposition integral, demonstrating the very important fact that a linear matrix system is completely characterized by its responses to identity impulse matrices. To completely specify the matrix output, the responses must in general be known for identity impulse matrices located at all possible points in the input plane. For the case of a linear polarization imaging system, this result has the interesting physical interpretation that the effects of polarization-sensitive elements (polarizers, wave plates, etc.) can be fully described by modifying the (possibly complex-valued) polarization images of the point sources located throughout the object optical field [18].

3. MATRIX-BASED CONVOLUTIONS/CORRELATIONS AND THEIR ALGEBRA

In our previous discussion on matrix linear systems, we have thus far considered the linear combination and/or superposition integral of matrix signals. In this section, we consider another means of combining matrix signals and systems: matrix (direct) convolutions, matrix (direct) correlations, and matrix element-wise convolution/correlation. The introduction of these matrix-based integral transformations can be regarded as a natural development and generalization of convolution or correlation integral and different types of matrix products. We begin with an introduction of the matrix convolution operation in the context of perhaps its most well-known and important applications: linear invariant matrix systems.

A. Matrix (Direct) Convolution

In optics, a linear polarization imaging system is space-invariant (or, equivalently, isoplanatic) if its impulse response matrix $\mathbf{H}(x_2, y_2; \xi, \eta)$ depends only on the distances $(x_2 - \xi)$ and $(y_2 - \eta)$. For such a matrix system, we can, of course, write $\mathbf{H}(x_2, y_2; \xi, \eta) = \mathbf{H}(x_2 - \xi, y_2 - \eta)$. Thus, an imaging system is space-invariant if the polarization images of a point source object change only in location, not in functional form, or in state-of-polarization. In practice, polarization imaging systems are seldom isoplanatic over their entire object field, but it is usually possible to divide that field into small regions (isoplanatic patches), within which the system is approximately invariant. Then, the superposition integral for linear invariant matrix system takes on a particularly simple form: $\mathbf{G}^{\mathrm{Out}}(x_2, y_2) = \iint_{-\infty}^{\infty} \mathbf{H}(x_2 - \xi, y - \eta)\mathbf{G}^{\mathrm{In}}(\xi, \eta)\mathrm{d}\xi\,\mathrm{d}\eta$, which is recognized as a two-dimensional convolution of the impulse response of the matrix system with the matrix signal of the object.

The *matrix convolution* $\mathbf{G}(x, y) * \mathbf{H}(x, y)$ denoted by the symbol $*$ gives the $m \times p$ matrix function, where $\mathbf{G}$ is an $m \times n$ matrix function and $\mathbf{H}$ is an $n \times p$ matrix function, respectively. That is,

$$[\mathbf{G} * \mathbf{H}]_{ij}(x, y) = \sum_{k=1}^{n} \iint_{-\infty}^{\infty} G_{ik}(\xi, \eta) H_{kj}(x - \xi, y - \eta)\mathrm{d}\xi\,\mathrm{d}\eta, \tag{7}$$

where the usual matrix multiplication has made use of when the matrix convolution was defined [15,18].

In linear algebra, the Kronecker product/matrix direct product/tensor product denoted by $\otimes$ is another type of matrix multiplication, giving the matrix of the tensor product of the linear map with respect to a standard choice of basis. In a similar way to the matrix convolution in Eq. (7), we can introduce and define the matrix direct convolution based on the matrix direct product. For an $m \times n$ matrix function $\mathbf{G}$ and a $p \times q$ matrix function $\mathbf{H}$, the *matrix direct convolution* $\mathbf{G}(x, y) \circledast \mathbf{H}(x, y)$ is the $(mp) \times (nq)$ block matrix function with its element defined by

$$[\mathbf{G} \circledast \mathbf{H}]_{cd}(x, y) = \iint_{-\infty}^{\infty} G_{ij}(\xi, \eta) H_{kl}(x - \xi, y - \eta)\mathrm{d}\xi\,\mathrm{d}\eta, \tag{8}$$

where $c = p(i-1) + k$ and $d = q(j-1) + l$. Here, the symbol $\circledast$ indicates the matrix direct convolution of two matrix functions with arbitrary sizes.

B. Matrix (Direct) Correlation

Note that correlation is another operation that strongly resembles convolution and shares very similar properties. The operation of correlation has found its applications in signal detection and estimation problems and in statistical communications theory to provide a measure of how similar a signal is to another signal. Just before, the matrix (direct) convolutions have been defined based on different types of matrix products. Therefore, it is natural to introduce and define matrix-based correlation operations for a matrix signal and/or system.

Suppose, as earlier, that we have two matrix functions $\mathbf{G}$ and $\mathbf{H}$ of respective sizes $m \times n$ and $n \times p$. The *matrix correlation* $\mathbf{G}(x, y) \star \mathbf{H}(x, y)$ (denoted by the symbol $\star$) gives the $m \times p$ matrix function with its element defined by

$$[\mathbf{G} \star \mathbf{H}]_{ij}(x, y) = \sum_{k=1}^{n} \iint_{-\infty}^{\infty} G_{ik}^{*}(\xi, \eta) H_{kj}(x + \xi, y + \eta)\mathrm{d}\xi\,\mathrm{d}\eta, \tag{9}$$

where the superscript of asterisk indicates the complex conjugate. The matrix autocorrelation of a matrix signal exists only when $\mathbf{G}$ is a square matrix function and is defined as correlation of $\mathbf{G}$ with itself: $\mathbf{G} \star \mathbf{G}$.

By analogy with the matrix direct convolution, the *matrix direct correlation* [15,18] can be introduced based on the matrix direct product. Given an $m \times n$ matrix function $\mathbf{G}$ and a $p \times q$ matrix function $\mathbf{H}$, their matrix direct correlation $\mathbf{G}(x, y) \circledstar \mathbf{H}(x, y)$ is an $(mp) \times (nq)$ matrix function with its element defined by

$$[\mathbf{G} \circledstar \mathbf{H}]_{cd}(x, y) = \iint_{-\infty}^{\infty} G_{ij}^{*}(\xi, \eta) H_{kl}(x + \xi, y + \eta)\mathrm{d}\xi\,\mathrm{d}\eta, \tag{10}$$

where $c = p(i-1) + k$ and $d = q(j-1) + l$. Similarly, the matrix direct autocorrelation $\mathbf{G} \circledstar \mathbf{G}$ is defined as the direct correlation of a matrix function $\mathbf{G}$ with itself, where $\mathbf{G}$ has arbitrary size without restriction of a square matrix as in the matrix autocorrelation.

C. Matrix Element-Wise Convolution/Correlation

In mathematics, the Hadamard product also known as the matrix element-wise product (denoted by $\odot$) is different from the matrix (direct) products. Taking in two matrices with the same dimensions, this operation returns a matrix of the multiplied corresponding elements and has found its wide applications in imaging processing, machine learning, and statistical analysis. In analogy with the matrix (direct) convolution and the matrix (direct) correlation introduced before, we can also propose the matrix element-wise convolution and the matrix element-wise correlation for matrix-based integral transformations.

For two matrix functions $\mathbf{G}$ and $\mathbf{H}$ with the same dimension $m \times n$, the *matrix element-wise convolution* $\mathbf{G}(x, y) \underset{\smile}{*} \mathbf{H}(x, y)$ provides a matrix function of the same dimension $m \times n$ with its element given by

$$[\mathbf{G} \underset{\smile}{*} \mathbf{H}]_{ij}(x, y) = \iint_{-\infty}^{\infty} G_{ij}(x - \xi, y - \eta) H_{ij}(\xi, \eta) \mathrm{d}\xi \mathrm{d}\eta. \tag{11}$$

Similarly, the *matrix element-wise correlation* $\mathbf{G}(x, y) \underset{\smile}{\star} \mathbf{H}(x, y)$ provides a different way of correlation between two equal-dimension matrix functions, and the corresponding element is defined by

$$[\mathbf{G} \underset{\smile}{\star} \mathbf{H}]_{ij}(x, y) = \iint_{-\infty}^{\infty} G_{ij}^{*}(\xi, \eta) H_{ij}(\xi + x, \eta + y) \mathrm{d}\xi \mathrm{d}\eta. \tag{12}$$

Once six matrix-based integral transformations (i.e., matrix convolution denoted by $*$, matrix correlation denoted by $\star$, matrix direct convolution denoted by $\circledast$, matrix direct correlation denoted by $\circledstar$, matrix element-wise convolution denoted by $\underset{\smile}{*}$, and matrix element-wise correlation denoted by $\underset{\smile}{\star}$) have been introduced for matrix signals and/or matrix systems, a natural question is whether these operations currently under consideration have useful algebraic properties such as the commutative law, the distributive law, and the associative law. The following results answer this question affirmatively.

D. Non-Commutativity or Commutativity

Note the fact that matrix-based integral transformations have been defined on the basis of matrix multiplications. The well-known non-commutativity [24] for matrix multiplication holds for the following matrix-based integral operations if the matrix product is defined:

$$\begin{aligned}
\mathbf{G}(x, y) * \mathbf{H}(x, y) &\neq \mathbf{H}(x, y) * \mathbf{G}(x, y), \\
\mathbf{G}(x, y) \star \mathbf{H}(x, y) &\neq \mathbf{H}(x, y) \star \mathbf{G}(x, y), \\
\mathbf{G}(x, y) \circledast \mathbf{H}(x, y) &\neq \mathbf{H}(x, y) \circledast \mathbf{G}(x, y), \\
\mathbf{G}(x, y) \circledstar \mathbf{H}(x, y) &\neq \mathbf{H}(x, y) \circledstar \mathbf{G}(x, y), \\
\mathbf{G}(x, y) \underset{\smile}{\star} \mathbf{H}(x, y) &\neq \mathbf{H}(x, y) \underset{\smile}{\star} \mathbf{G}(x, y).
\end{aligned} \tag{13}$$

Unlike the matrix (direct) convolution, the matrix (direct) correlation, and the matrix element-wise correlation, the matrix element-wise convolution is commutative:

$$\mathbf{G}(x, y) \underset{\smile}{*} \mathbf{H}(x, y) = \mathbf{H}(x, y) \underset{\smile}{*} \mathbf{G}(x, y). \tag{14}$$

E. Distributivity

Since the matrix product is distributive with respect to matrix addition, the distributive law holds for these six matrix-based integral transformations. That is, if $\mathbf{F}$, $\mathbf{G}$, $\mathbf{H}$ are matrix functions of respective sizes $m \times n$, $n \times p$ and $n \times p$, we have

$$\begin{aligned}
\mathbf{F}(x, y) * [\mathbf{G}(x, y) + \mathbf{H}(x, y)] &= \mathbf{F}(x, y) * \mathbf{G}(x, y) \\
&\quad + \mathbf{F}(x, y) * \mathbf{H}(x, y), \\
\mathbf{F}(x, y) \star [\mathbf{G}(x, y) + \mathbf{H}(x, y)] &= \mathbf{F}(x, y) \star \mathbf{G}(x, y) \\
&\quad + \mathbf{F}(x, y) \star \mathbf{H}(x, y).
\end{aligned} \tag{15}$$

When $\mathbf{G}$ and $\mathbf{H}$ are matrix functions of the same size, the distributive property for matrix direct convolution and matrix direct correlation can be expressed as

$$\begin{aligned}
\mathbf{F}(x, y) \circledast [\mathbf{G}(x, y) + \mathbf{H}(x, y)] &= \mathbf{F}(x, y) \circledast \mathbf{G}(x, y) \\
&\quad + \mathbf{F}(x, y) \circledast \mathbf{H}(x, y), \\
\mathbf{F}(x, y) \circledstar [\mathbf{G}(x, y) + \mathbf{H}(x, y)] &= \mathbf{F}(x, y) \circledstar \mathbf{G}(x, y) \\
&\quad + \mathbf{F}(x, y) \circledstar \mathbf{H}(x, y).
\end{aligned} \tag{16}$$

The matrix element-wise convolution and matrix element-wise correlation are also distributive. That is, if $\mathbf{F}$, $\mathbf{G}$, $\mathbf{H}$ are matrix functions of the same size $m \times n$, we have

$$\begin{aligned}
\mathbf{F}(x, y) \underset{\smile}{*} [\mathbf{G}(x, y) + \mathbf{H}(x, y)] &= \mathbf{F}(x, y) \underset{\smile}{*} \mathbf{G}(x, y) \\
&\quad + \mathbf{F}(x, y) \underset{\smile}{*} \mathbf{H}(x, y), \\
\mathbf{F}(x, y) \underset{\smile}{\star} [\mathbf{G}(x, y) + \mathbf{H}(x, y)] &= \mathbf{F}(x, y) \underset{\smile}{\star} \mathbf{G}(x, y) \\
&\quad + \mathbf{F}(x, y) \underset{\smile}{\star} \mathbf{H}(x, y).
\end{aligned} \tag{17}$$

F. Associativity

Given three matrix functions: $\mathbf{F}$, $\mathbf{G}$, and $\mathbf{H}$, the products $(\mathbf{FG})\mathbf{H}$ and $\mathbf{F}(\mathbf{GH})$ have been defined if and only if the number of columns of $\mathbf{F}$ equals the number of rows of $\mathbf{G}$, and the number of columns of $\mathbf{G}$ equals the number of the rows of $\mathbf{H}$. In this case, one has the associative property for the matrix convolution and the matrix correlation:

$$\begin{aligned}
\mathbf{F}(x, y) * [\mathbf{G}(x, y) * \mathbf{H}(x, y)] &= [\mathbf{F}(x, y) * \mathbf{G}(x, y)] * \mathbf{H}(x, y), \\
\mathbf{F}(x, y) \star [\mathbf{G}(x, y) \star \mathbf{H}(x, y)] &= [\mathbf{F}(x, y) \star \mathbf{G}(x, y)] \star \mathbf{H}(x, y).
\end{aligned} \tag{18}$$

The associativity for the matrix direct convolution and the matrix direct correlation is

$$\begin{aligned}
\mathbf{F}(x, y) \circledast [\mathbf{G}(x, y) \circledast \mathbf{H}(x, y)] &= [\mathbf{F}(x, y) \circledast \mathbf{G}(x, y)] \circledast \mathbf{H}(x, y), \\
\mathbf{F}(x, y) \circledstar [\mathbf{G}(x, y) \circledstar \mathbf{H}(x, y)] &= [\mathbf{F}(x, y) \circledstar \mathbf{G}(x, y)] \circledstar \mathbf{H}(x, y).
\end{aligned} \tag{19}$$

When $\mathbf{F}$, $\mathbf{G}$, and $\mathbf{H}$ are matrix functions with the same size, the matrix element-wise convolution and the matrix element-wise correlation are associative:

$$\begin{aligned}
\mathbf{F}(x, y) \underset{\smile}{*} [\mathbf{G}(x, y) \underset{\smile}{*} \mathbf{H}(x, y)] &= [\mathbf{F}(x, y) \underset{\smile}{*} \mathbf{G}(x, y)] \underset{\smile}{*} \mathbf{H}(x, y), \\
\mathbf{F}(x, y) \underset{\smile}{\star} [\mathbf{G}(x, y) \underset{\smile}{\star} \mathbf{H}(x, y)] &= [\mathbf{F}(x, y) \underset{\smile}{\star} \mathbf{G}(x, y)] \underset{\smile}{\star} \mathbf{H}(x, y).
\end{aligned} \tag{20}$$

All of these properties may be proved straightforwardly by combined manipulations of matrix addition, matrix multiplication, and linearity of integration, and these results also follow from the fact that matrices represent linear maps.

4. MATRIX-BASED FOURIER TRANSFORM THEOREMS

In the previous sessions, we have extended the linear systems theory in scalar optics to corresponding matrix versions for polarization optics. Since the Fourier transform is perhaps the most important analytical tool needed for work in modern optics, we present a summary of the most important matrix-based Fourier transform theorems needed in practice. Here, we will adopt an operational approach, which is the characteristic of most engineering treatments. All of these properties can be derived and/or proved at great mathematical rigor with the aid of appropriate combination of relevant treatment for Fourier analysis of scalar function [11,19–21] and the matrix theory [24].

Here, we have chosen to use definitions of the matrix-based Fourier transforms that have a positive exponential kernel. Thus, the Fourier transform of a complex-valued matrix function $\mathbf{G}$ with its size $m \times n$ of two independent variables x and y will be represented by $\Im\{\mathbf{G}\}$ and is defined by

$$\Im\{\mathbf{G}(x, y)\} = \mathcal{G}(f_x, f_y) \triangleq \iint_{-\infty}^{\infty} \mathbf{G}(x, y) \exp[i2\pi(f_x x + f_y y)]\mathrm{d}x\mathrm{d}y. \tag{21}$$

Similarly, the inverse Fourier transform of a matrix function $\mathcal{G}$ of two variables f_x and f_y will be represented by $\Im^{-1}\{\mathcal{G}\}$ and is defined as

$$\Im^{-1}\{\mathcal{G}(f_x, f_y)\} = \mathbf{G}(x, y) \triangleq \iint_{-\infty}^{\infty} \mathcal{G}(f_x, f_y) \times \exp[-i2\pi(f_x x + f_y y)]\mathrm{d}f_x\mathrm{d}f_y. \tag{22}$$

The basic definitions in Eqs. (21) and (22) of the matrix-based (inverse) Fourier transformations lead to rich mathematical structure associated with the transform operations. We now consider a few of the basic mathematical properties of these transforms because these properties will find wide use for future application to polarization related Fourier optics. These properties are presented as mathematical theorems, followed by brief proofs and/or statements of their physical significance. Throughout this session, $\mathbf{G}$ and $\mathbf{H}$ represent matrix functions (generally complex valued), and $\mathcal{G}$ and $\mathcal{H}$ represent their Fourier transforms.

A. Matrix Linearity Theorems

Let $\mathbf{A}$ and $\mathbf{B}$ represent complex-valued constant matrices with the same size and $\mathbf{G}$ and $\mathbf{H}$ be matrix functions with the same size. If both the products $\mathbf{AG}$ and $\mathbf{BH}$ are defined (that is, the number of columns of $\mathbf{A}$ or $\mathbf{B}$ equals the number of rows of $\mathbf{G}$ or $\mathbf{H}$), then

$$\Im\{\mathbf{AG}(x, y) + \mathbf{BH}(x, y)\} = \mathbf{A}\mathcal{G}(f_x, f_y) + \mathbf{B}\mathcal{H}(f_x, f_y). \tag{23}$$

Based on matrix Kronecker (direct) products $\mathbf{A} \otimes \mathbf{G}$ and $\mathbf{B} \otimes \mathbf{H}$, we have

$$\Im\{\mathbf{A} \otimes \mathbf{G}(x, y) + \mathbf{B} \otimes \mathbf{H}(x, y)\} = \mathbf{A} \otimes \mathcal{G}(f_x, f_y) + \mathbf{B} \otimes \mathcal{H}(f_x, f_y). \tag{24}$$

Similarly, based on matrix Hadamard products $\mathbf{A} \odot \mathbf{G}$ and $\mathbf{B} \odot \mathbf{H}$, we have

$$\Im\{\mathbf{A} \odot \mathbf{G}(x, y) + \mathbf{B} \odot \mathbf{H}(x, y)\} = \mathbf{A} \odot \mathcal{G}(f_x, f_y) + \mathbf{B} \odot \mathcal{H}(f_x, f_y). \tag{25}$$

The matrix linearity theorems in Eqs. (23)–(25) follow directly from the linearity of the integral transformations that define the matrix-based Fourier transforms and indicate that the transform of a sum of two (or more) matrix functions weighted by constant matrices is simply the identically weighted sum of their individual transforms. It should be noted that the linearity presented in Eq. (2) is just a special case of the presented matrix linearity theorem where two matrices $\mathbf{A}$ and $\mathbf{B}$ have been reduced to two complex-valued constants.

B. Matrix Similarity Theorem

If a and b are real-valued constants, then

$$\Im\{\mathbf{G}(ax, by)\} = |ab|^{-1}\mathcal{G}\left(f_x/a, f_y/b\right). \tag{26}$$

The matrix similarity theorem can be understood as follows. A "stretch" of the coordinates in the space domain (x, y) for a matrix function results in a contraction of the coordinates in the frequency domain (f_x, f_y) with a change in the overall amplitude of the spectral matrix.

C. Matrix Shift Theorem

If a and b are real-valued constants, then

$$\Im\{\mathbf{G}(x - a, y - b)\} = e^{i2\pi(af_x + bf_y)}\mathcal{G}\left(f_x, f_y\right). \tag{27}$$

The matrix shift theorem states that translation of matrix signal in the space domain introduces a linear phase shift of its spectral matrix in the frequency domain. The matrix shift theorem in Eq. (27) can be proved with the aid of the shift theorem applied to each element in the matrix function.

D. Matrix Conjugate Theorem

If $\Im\{\mathbf{G}(x, y)\} = \mathcal{G}(f_x, f_y)$, then

$$\Im\{\mathbf{G}^*(x, y)\} = \mathcal{G}^*(-f_x, -f_y). \tag{28}$$

In particular, if $\mathbf{G}$ is a real-valued matrix function, then $\mathcal{G}$ is a Hermitian matrix fucntion with even symmetry (i.e., $\mathcal{G}(-f_x, -f_y) = \mathcal{G}^*(f_x, f_y)$). And if $\mathbf{G}$ is purely imaginary, then $\mathcal{G}$ is odd symmetric (i.e., $\mathcal{G}(-f_x, -f_y) = -\mathcal{G}^*(f_x, f_y)$).

E. Law of Energy Conservation

If $\Im\{\mathbf{G}(x, y)\} = \mathcal{G}(f_x, f_y)$, then

$$\iint_{-\infty}^{\infty} ||\mathbf{G}(x, y)||^2\mathrm{d}x\mathrm{d}y = \iint_{-\infty}^{\infty} ||\mathcal{G}(f_x, f_y)||^2\mathrm{d}f_x\mathrm{d}f_y, \tag{29}$$

where the symbol $||\cdots||$ represents the matrix's Frobenius norm [24]. Here is the brief proof of the law of energy conservation for a matrix signal. Applying the Rayleigh's (Parseval's) theorem, if follows directly for each element in a matrix signal that $\iint_{-\infty}^{\infty}|G_{lm}(x,y)|^2\mathrm{d}x\mathrm{d}y=\iint_{-\infty}^{\infty}|\mathcal{G}_{lm}(f_x,f_y)|^2\mathrm{d}f_x\mathrm{d}f_y$. After interchanging the order of integration and summation stemming from the squared Frobenius norm (i.e., $||\mathbf{G}||^2=\sum_l\sum_m|G_{lm}|^2$), we complete the derivation of the law of conservation of energy for a matrix signal. The integral on the left-hand side of this theorem can be interpreted as the energy contained in the matrix signal $\mathbf{G}(x,y)$. This in turn leads us to the idea that the quantity $||\boldsymbol{\mathcal{G}}(x,y)||^2$ can be interpreted as an energy density of the matrix signal in the frequency domain. Here, the term 'energy' has been used in a generalized sense of signal. Therefore, Eq. (29) can be understood as a natural extension of the Rayleigh's theorem (Parseval's theorem) for a matrix signal.

F. Matrix-Based Convolution Theorems

If $\Im\{\mathbf{G}(x,y)\}=\boldsymbol{\mathcal{G}}(f_x,f_y)$ and $\Im\{\mathbf{H}(x,y)\}=\boldsymbol{\mathcal{H}}(f_x,f_y)$, then

$$\Im\{\mathbf{G}(x,y)*\mathbf{H}(x,y)\}=\boldsymbol{\mathcal{G}}\left(f_x,f_y\right)\boldsymbol{\mathcal{H}}\left(f_x,f_y\right),$$
$$\Im\{\mathbf{G}(x,y)\circledast\mathbf{H}(x,y)\}=\boldsymbol{\mathcal{G}}\left(f_x,f_y\right)\otimes\boldsymbol{\mathcal{H}}\left(f_x,f_y\right),$$
$$\Im\{\mathbf{G}(x,y)\,\underset{\smile}{*}\,\mathbf{H}(x,y)\}=\boldsymbol{\mathcal{G}}\left(f_x,f_y\right)\odot\boldsymbol{\mathcal{H}}\left(f_x,f_y\right). \quad (30)$$

With the aid of the definitions for the matrix (direct) convolutions and the matrix element-wise convolution in Eqs. (7), (8), and (11), the matrix convolution theorems in Eq. (30) can be derived by applying the Fourier convolution theorem for scalar functions and interchanging the order of integration and summation to each matrix element. The physical significance of the matrix-based convolution theorems can be understood as follows. The matrix (direct) or element-wise convolution of two matrix functions in the space domain is entirely equivalent to the simpler operation of matrix (direct) or Hadamard product of their individual transforms.

G. Matrix-Based Correlation Theorems

If $\Im\{\mathbf{G}(x,y)\}=\boldsymbol{\mathcal{G}}(f_x,f_y)$ and $\Im\{\mathbf{H}(x,y)\}=\boldsymbol{\mathcal{H}}(f_x,f_y)$, then

$$\Im\{\mathbf{G}(x,y)\star\mathbf{H}(x,y)\}=\boldsymbol{\mathcal{G}}^*\left(f_x,f_y\right)\boldsymbol{\mathcal{H}}\left(f_x,f_y\right),$$
$$\Im\{\mathbf{G}(x,y)\circledstar\mathbf{H}(x,y)\}=\boldsymbol{\mathcal{G}}^*\left(f_x,f_y\right)\otimes\boldsymbol{\mathcal{H}}(f_x,f_y),$$
$$\Im\{\mathbf{G}(x,y)\,\underset{\smile}{\star}\,\mathbf{H}(x,y)\}=\boldsymbol{\mathcal{G}}^*\left(f_x,f_y\right)\odot\boldsymbol{\mathcal{H}}\left(f_x,f_y\right). \quad (31)$$

The matrix (direct) or element-wise correlation theorems above may be regarded as special cases of the matrix (direct) or element-wise convolution theorems where we conduct matrix (direct) or element-wise convolution between two matrix signals $\mathbf{G}^*(x,y)$ and $\mathbf{H}(-x,-y)$.

H. Matrix Derivative Theorem

Given a continuous matrix function $\mathbf{G}(x,y)$, which is everywhere differentiable, and its Fourier transform $\boldsymbol{\mathcal{G}}(f_x,f_y)$, then

$$\Im\left\{\frac{\partial^{p+q}}{\partial x^p\partial y^q}\mathbf{G}(x,y)\right\}=(-i2\pi f_x)^p(-i2\pi f_y)^q\boldsymbol{\mathcal{G}}(f_x,f_y). \quad (32)$$

This theorem follows directly after the derivative theorem has been applied to each element of a matrix function; therefore, Fourier transform of the higher-order partial and mixed derivative of the matrix function results in multiplication of the original spectral matrix by a constant $(-i2\pi f_x)^p(-i2\pi f_y)^q$.

I. Matrix-Based Fourier Integral Theorem

Learning from the Fourier integral theorem for a scalar function [11,19–21], we have that a Fourier transform followed by an inverse Fourier transform yields the original value of each element in the matrix function. At each point of discontinuity in any element of a matrix function, the two successive transforms yield the angular-averaged value of that element in a small neighborhood of that point. Therefore, except at points of discontinuity, the successive transformation and inverse transformation of a matrix function yields that matrix function

$$\Im\Im^{-1}\{\mathbf{G}(x,y)\}=\Im^{-1}\Im\{\mathbf{G}(x,y)\}=\mathbf{G}(x,y), \quad (33)$$

which is the matrix version of Fourier integral theorem.

Similar to the well-known Fourier analysis of scalar function [11], the above matrix-based Fourier transform theorems are of far more than just theoretical interest. They can save an enormous amount of work in the solution of Fourier analysis of matrix function since they provide the basic tools for the manipulation of matrix-based Fourier transformations.

5. SOME APPLICATIONS OF FOURIER ANALYSIS OF MATRIX SIGNALS AND SYSTEMS

In this section, we explore some applications of matrix-based Fourier analysis of two-dimensional signals and systems, as predicted by the theories presented in previous sections. Several important examples that arise frequently in optics, communication, and information sciences will be discussed for the purpose of illustration.

A. Identity Impulse Matrix

Similar to the Dirac delta function, the identity impulse matrix $\boldsymbol{\delta}_m(x,y)$ is a generalized matrix function and will find its considerable utility in engineering problems. As introduced in the Section 2, the identity impulse matrix is defined as scalar product of the Dirac delta function and the identity matrix of size m, so that $\boldsymbol{\delta}_m(x,y)$ is the $m\times m$ square matrix with the Dirac delta functions on the main diagonal and zeros elsewhere. The identity impulse matrix can be obtained as an inverse Fourier transform of a scalar product between $\exp[i2\pi(af_x+bf_y)]$ with the identity matrix

$$\boldsymbol{\delta}_m(x-a,y-b)=\Im^{-1}\{\exp[i2\pi(af_x+bf_y)]\mathbf{I}_m\}. \quad (34)$$

When $a = b = 0$, we have $\boldsymbol{\delta}_m(x, y) = \Im^{-1}\{\mathbf{I}_m\}$:

$$\Im\{\boldsymbol{\delta}_m(x, y)\} = \mathbf{I}_m. \quad (35)$$

A number of formal relationships involving the identity impulse matrix may be obtained directly from its definition and the matrix-based Fourier transform theorems. For example, the identity impulse matrix is constrained to satisfy the identity matrix

$$\iint_{-\infty}^{\infty} \boldsymbol{\delta}_m(x, y)\mathrm{d}x\mathrm{d}y = \mathbf{I}_m. \quad (36)$$

The identity impulse matrix satisfies the scaling property for nonzero real-valued scalars a and b

$$\boldsymbol{\delta}_m(ax, by) = \frac{1}{|ab|}\boldsymbol{\delta}_m(x, y). \quad (37)$$

The identity impulse matrix also obeys the so-called the *matrix sifting property* if the matrix product $\boldsymbol{\delta}_m\mathbf{G}$ is defined

$$\iint_{-\infty}^{\infty} \boldsymbol{\delta}_m(x - x_0, y - y_0)\mathbf{G}(x, y)\mathrm{d}x\mathrm{d}y = \mathbf{G}(x_0, y_0). \quad (38)$$

As a direct consequence of Eq. (38), we have the *matrix identity property*

$$\boldsymbol{\delta}_m(x, y) * \mathbf{G}(x, y) = \mathbf{G}(x, y), \quad (39)$$

which is fundamental in matrix-based signal processing since a matrix convolution between the identity impulse matrix and any matrix of appropriate size is simply the matrix function itself.

Following the conventional definition for the Dirac delta distribution [25], the derivative of the identity impulse matrix, denoted by $\boldsymbol{\delta}_m^{(p,q)}(x, y) = \partial_x^p \partial_y^q \boldsymbol{\delta}_m(x, y)$, is defined on matrix function $\mathbf{G}$ by $\iint_{-\infty}^{\infty} \boldsymbol{\delta}_m^{(p,q)}(x, y)\mathbf{G}(x, y)\mathrm{d}x\mathrm{d}y = (-1)^{p+q}\partial_x^p \partial_y^q \mathbf{G}(x, y)|_{x=y=0}$. Then the corresponding Fourier transform of the higher-order derivative of the identity impulse matrix is

$$\Im\{\boldsymbol{\delta}_m^{(p,q)}(x, y)\} = (-i2\pi f_x)^p(-i2\pi f_y)^q\mathbf{I}_m. \quad (40)$$

B. Matrix Sampling Theory

For both data processing and mathematical analysis purposes, it is of convenience to represent a $m \times n$ matrix function $\mathbf{G}(x, y)$ by $m \times n$ arrays of the sampled values taken on discrete points in the (x, y) plane. Here, we will extend and develop the Whittaker–Shannon sampling theorem by taking a class of bandlimited matrix function into account.

To develop a matrix version of the sampling theorem, we apply the identical rectangular sampling lattice $m \times n$ times to all elements in the matrix function $\mathbf{G}$. The sampled matrix function thus consists of $m \times n$ arrays of $\delta(x, y)$ functions, spaced at identical intervals of width X in the $\hat{x}$ direction and identical width Y in the $\hat{y}$ direction. The area under each Dirac delta function is proportional to the value of the corresponding matrix element at that particular point in the rectangular sampling lattices. Since the matrix function $\mathbf{G}$ is assumed to be bandlimited, the Frobenius norm of the spectral matrix $||\boldsymbol{\mathcal{G}}(f_x, f_y)||$ is nonzero over only a finite region of the frequency space. Let $2B_x$ and $2B_y$ represent the bandwidths in the $\hat{f}_x$ and $\hat{f}_y$ directions, respectively, of the smallest rectangle enclosing completely the finite region in the spatial frequency domain. The allowable spacings of the sampling lattices for exact recovery of the original matrix function $\mathbf{G}$ are $X \leq (2B_x)^{-1}$ and $Y \leq (2B_y)^{-1}$, respectively.

When the sampling intervals X and Y are taken to have their maximum allowable values, exact recovery of a bandlimited matrix function can be achieved from appropriately spaced rectangular arrays of their sampled values of the matrix. That is,

$$\mathbf{G}(x, y) = \sum_p \sum_q \mathbf{G}\left(p/(2B_x), q/(2B_y)\right) \times \mathrm{sinc}(2B_x[x - p/(2B_x)])\mathrm{sinc}(2B_y[y - q/(2B_y)]), \quad (41)$$

which represents a matrix version of the Whittaker–Shannon sampling theorem. The exact recovery of the matrix function $\mathbf{G}$ is accomplished by injecting, at each sampling point, an interpolation function consisting of a product of sinc$(\cdots)$ functions, where each interpolated function is weighted by the matrix function $\mathbf{G}$ sampled at the corresponding point.

It is impossible for a bandlimited matrix function to be nonzero over only a finite region in the (x, y) plane simultaneously; most matrix functions in practice do eventually fall to very small values for all the elements. If $\mathbf{G}(x, y)$ is bandlimited and indeed has significant values over a finite region $-L_x \leq x \leq L_x$, $-L_y \leq y \leq L_y$, then it is possible to represent $\mathbf{G}(x, y)$ with good accuracy by a finite number of samples. In accord with the matrix sampling theorem, the total number of significant samples required to represent an $m \times n$ matrix function $\mathbf{G}(x, y)$ is seen to be

$$\mathcal{M} = mn(2L_x)(2L_y)/(XY) = 16mnL_xL_yB_xB_y, \quad (42)$$

which is referred to as the *matrix's space-bandwidth product* (MSBP) for the matrix function $\mathbf{G}(x, y)$. In the context of polarization optics, the MSBP can be regarded as the number of degrees of freedom of the given matrix function, providing a measure of the information-carrying capacity of a polarization-sensitive optical system, or setting a limit on the amount of recorded and reconstructed optical information in a matrix signal.

C. Mean-Squared Width/Bandwidth of Matrix Signal and Uncertainty Relation

In signal processing, the notions of the width of a signal or the bandwidth of its spectrum have been frequently encountered to measure the amount of space or spatial frequency required to contain most of the signal or spectrum. Note that some conventional measures of pulse width and bandwidth well-developed for a scalar signal break down since a matrix signal may have zero elements/components and therefore become inappropriate for measuring the overall width of an oscillatory matrix signal and/or wave packet described in a matrix function. To cope with this case, we are proposing yet another definition for a matrix signal in terms of the energy density through the centroid and the variance of the energy distribution. We shall refer to the mean-square departure from the centroid of $||\mathbf{G}||^2$ as the variance given by

$$(\Delta x^{\mathbf{G}})^2 = (\sigma_x^{||\mathbf{G}||^2})^2 \triangleq \frac{\iint_{-\infty}^{\infty} x^2 ||\mathbf{G}(x, y)||^2 \mathrm{d}x\mathrm{d}y}{\iint_{-\infty}^{\infty} ||\mathbf{G}(x, y)||^2 \mathrm{d}x\mathrm{d}y} - \left(\frac{\iint_{-\infty}^{\infty} x ||\mathbf{G}(x, y)||^2 \mathrm{d}x\mathrm{d}y}{\iint_{-\infty}^{\infty} ||\mathbf{G}(x, y)||^2 \mathrm{d}x\mathrm{d}y}\right)^2, \tag{43}$$

where $\Delta x^{\mathbf{G}}$ provides the spatial spread of $\mathbf{G}$ along the $\hat{x}$ direction about its centroid and is called the *mean-squared width* along the $\hat{x}$ direction, which is the standard derivation of the instantaneous power of the matrix signal. The corresponding frequency domain width of the spectral matrix $\Delta f_x^{\boldsymbol{\mathcal{G}}}$, which is the standard deviation of the energy spectrum, is defined as the square root of the variance

$$(\Delta f_x^{\boldsymbol{\mathcal{G}}})^2 = (\sigma_x^{||\boldsymbol{\mathcal{G}}||^2})^2 \triangleq \frac{\iint_{-\infty}^{\infty} f_x^2 ||\boldsymbol{\mathcal{G}}(f_x, f_y)||^2 \mathrm{d}f_x\mathrm{d}f_y}{\iint_{-\infty}^{\infty} ||\boldsymbol{\mathcal{G}}(f_x, f_y)||^2 \mathrm{d}x\mathrm{d}y} - \left(\frac{\iint_{-\infty}^{\infty} f_x ||\boldsymbol{\mathcal{G}}(f_x, f_y)||^2 \mathrm{d}f_x\mathrm{d}f_y}{\iint_{-\infty}^{\infty} ||\boldsymbol{\mathcal{G}}(f_x, f_y)||^2 \mathrm{d}f_x\mathrm{d}f_y}\right)^2. \tag{44}$$

The mean-squared width $\Delta y^{\mathbf{G}}$ and bandwidth $\Delta f_y^{\boldsymbol{\mathcal{G}}}$ of the matrix signal along the $\hat{y}$ direction can be introduced in a similar way. These definitions of widths and bandwidths are independent of the origin (not affected by translation), and they are strictly positive for all physical matrix functions.

With the aid of the law of energy conservation, the matrix's derivative theorem, and the Cauchy–Schwarz inequality, the space-bandwidth products of these newly introduced mean-squared width and bandwidth along two directions satisfy

$$(\Delta x^{\mathbf{G}})(\Delta f_x^{\boldsymbol{\mathcal{G}}}) \geq (4\pi)^{-1}$$
$$(\Delta y^{\mathbf{G}})(\Delta f_y^{\boldsymbol{\mathcal{G}}}) \geq (4\pi)^{-1}, \tag{45}$$

which provide the uncertainty relations for matrix signal. The minimum is attained only in the case of a Gaussian-shaped matrix signal [i.e., $\mathbf{G}(x, y)$ is a scalar product of a two-dimensional Gaussian function with a constant matrix]. When Eq. (45) is derived, we have followed the same derivation procedure used for the Garbor's uncertainty principle of scalar signal [19–21]. The basic result in Eq. (45) states that *it is impossible to arbitrarily concentrate a matrix signal in both space- and spatial frequency domains simultaneously* and therefore provides a lower bound to the width in one domain in terms of the inverse of the width in the other.

D. Haagerup's Inequality and Normalization for Matrix Direct Correlation

In the previous section, we have provided a mathematical definition for the matrix direct correlation as an integral transformation based on the matrix Kronecker (direct) product. The matrix direct correlation has found its applications in several areas of optics. For instance, the mutual coherence matrix and the generalized Stokes vector [17,23] in statistical optics have been redefined as the matrix direct cross correlation of electric field vectors at different spatio-temporal points [26]. In physical image formation, the optical transfer matrix (OTM) has been introduced as the matrix direct autocorrelation of the generalized pupil matrix [18]. In order to derive one of the principal properties of matrix direct correlation, an upper bound on the magnitude of the matrix direct correlation in terms of the energies of the matrix functions, we would like to introduce the Haagerup's inequality [27], which is one of the most useful inequalities in matrix mathematics. As a generalization of the Cauchy–Schwarz inequality to the matrix space, the Haagerup's inequality states for two matrix functions $\mathbf{X}(\xi, \eta)$ and $\mathbf{Y}(\xi, \eta)$ that

$$\left\| \iint_{-\infty}^{\infty} \mathbf{X} \otimes \mathbf{Y} \mathrm{d}\xi \mathrm{d}\eta \right\| \leq \left\| \iint_{-\infty}^{\infty} \mathbf{X} \otimes \mathbf{X}^* \mathrm{d}\xi \mathrm{d}\eta \right\|^{1/2} \times \left\| \iint_{-\infty}^{\infty} \mathbf{Y} \otimes \mathbf{Y}^* \mathrm{d}\xi \mathrm{d}\eta \right\|^{1/2}, \tag{46}$$

with equality if and only if $\mathbf{Y} = \alpha \mathbf{X}^*$, where α is a complex constant. Letting $\mathbf{X}(\xi, \eta) = \mathbf{G}^*(\xi, \eta)$ and $\mathbf{Y}(\xi, \eta) = \mathbf{H}(\xi + x, \eta + y)$, we have

$$\|\mathbf{G}(x, y) \circledast \mathbf{H}(x, y)\| \leq \left\| \mathbf{G}(x, y) \circledast \mathbf{G}(x, y)|_{x=y=0} \right\|^{1/2} \times \left\| \mathbf{H}(x, y) \circledast \mathbf{H}(x, y)|_{x=y=0} \right\|^{1/2}, \tag{47}$$

where $\mathbf{F} \circledast \mathbf{F}|_{x=y=0}$ for ($\mathbf{F} = \mathbf{G}$ or $\mathbf{H}$) indicates that the matrix direct autocorrelation is evaluated at zero offset (i.e., $x = y = 0$). The Haagerup's inequality leads us to define a normalized matrix of the form

$$[\widehat{\mathbf{G} \circledast \mathbf{H}}](x, y) \triangleq \frac{\mathbf{G}(x, y) \circledast \mathbf{H}(x, y)}{\left\| \mathbf{G} \circledast \mathbf{G}|_{x=y=0} \right\|^{1/2} \left\| \mathbf{H} \circledast \mathbf{H}|_{x=y=0} \right\|^{1/2}}. \tag{48}$$

From the inequality of Eq. (47), we readily see that $||[\widehat{\mathbf{G} \circledast \mathbf{H}}](x, y)|| \leq 1$. It is the Haagerup's inequality that justified the definition of normalization for matrix direct correlation given in Eq. (48). In fact, such way of matrix normalization has been adopted to normalize the (generalized) Stokes vector, the mutual coherence matrix, and the polarization matrix in statistical optics [26] and to define the optical transfer matrix (OTM) for the Stokes imaging with incoherent light of arbitrary polarization [18].

6. EXAMPLE

So far, we have introduced the linear invariant matrix systems theory and the matrix-based Fourier analysis of general matrix signals and systems without specifying the optical signals and/or systems in terms of matrix functions for polarization optics. We now turn to an example that uses the theories developed before for a discussion about the coherence time and effective spectral width of the electromagnetic wave, though the problem of interest is matrix-based Fourier analysis in one dimension.

The concept of the coherence time has been found particularly useful as a measure of the time interval in which appreciable correlations of light vibrations will persist. As one of the basic quantities in statistical optics [17,22,23,28], the temporal coherence matrix of the optical disturbance is given by $\mathbf{\Gamma}(\tau) = \langle \mathbf{E}(t+\tau) \otimes \mathbf{E}^\dagger(t)\rangle$ with the angular bracket representing an infinite time average and the superscript † denoting the operation of conjugate and transpose. In terms of the matrix direct correlation, the temporal coherence matrix can be rewritten as $\mathbf{\Gamma}(\tau) = \mathbf{E}(\tau) \circledast \mathbf{E}^{\mathrm{T}}(\tau)$ with the superscript T denoting the matrix transpose. Since the temporal coherence matrix and the power spectral density matrix form a Fourier transform pair, we have the corresponding power spectral density matrix of the light $\boldsymbol{\mathcal{G}}(\nu) = \Im\{\mathbf{\Gamma}(\tau)\}$.

Based on our previous discussion on the width and bandwidth of matrix signal, it is natural and convenient to define the coherence time as the normalized root-mean-square width of the squared norm of $\mathbf{\Gamma}(\tau)$

$$(\Delta\tau)^2 = \frac{\int_{-\infty}^{\infty} \tau^2 ||\mathbf{\Gamma}(\tau)||^2 \mathrm{d}\tau}{\int_{-\infty}^{\infty} ||\mathbf{\Gamma}(\tau)||^2 \mathrm{d}\tau}, \tag{49}$$

where $||\mathbf{\Gamma}(\tau)||^2$ can be understood as the coherenergy density of stochastic electromagnetic wave [29]. When Eq. (49) is written, we have made use of the property that the average value $\bar{\tau} = \int_{-\infty}^{\infty} \tau ||\mathbf{\Gamma}(\tau)||^2 \mathrm{d}\tau / \int_{-\infty}^{\infty} ||\mathbf{\Gamma}(\tau)||^2 \mathrm{d}\tau$ is zero because $||\mathbf{\Gamma}(\tau)||^2$ is an even function of τ.

Similarly, we may define the effective spectral width $\Delta\nu$ (the bandwidth) of the light as the normalized root-mean-square width of the spectral density matrix $\boldsymbol{\mathcal{G}}(\nu)$ by the formula

$$(\Delta\nu)^2 = \frac{\int_0^{\infty} (\nu - \bar{\nu})^2 ||\boldsymbol{\mathcal{G}}(\nu)||^2 \mathrm{d}\nu}{\int_0^{\infty} ||\boldsymbol{\mathcal{G}}(\nu)||^2 \mathrm{d}\nu}, \tag{50}$$

where $\bar{\nu}$ may be identified with the mean frequency of the electromagnetic wave given by

$$\bar{\nu} = \frac{\int_0^{\infty} \nu ||\boldsymbol{\mathcal{G}}(\nu)||^2 \mathrm{d}\nu}{\int_0^{\infty} ||\boldsymbol{\mathcal{G}}(\nu)||^2 \mathrm{d}\nu}. \tag{51}$$

With the aid of the uncertainty relation for the matrix signal given by Eq. (45), we obtain the *reciprocity inequality* for a stochastic electromagnetic wave: $(\Delta\tau)(\Delta\nu) \geq (4\pi)^{-1}$. Note that the equality sign will apply if and only if $\mathbf{\Gamma}(\tau)$ is a Gaussian matrix function. However, even if the power spectral density matrix $\boldsymbol{\mathcal{G}}(\nu)$ is Gaussian matrix centered at the mean frequency $\bar{\nu}$, the corresponding temporal coherence matrix $\mathbf{\Gamma}(\tau)$ can be expressible as a scalar product of $\exp(-i2\pi\bar{\nu}\tau)$ and another Gaussian matrix function. This property derived from the matrix shift theorem in Eq. (27) is a result of our choice of spectral line shapes that are symmetrical about $\bar{\nu}$. Thus in the present context, the equality sign in the reciprocity inequality relation never strictly holds. However, when the power spectral density matrix $\boldsymbol{\mathcal{G}}(\nu)$ is approximately of Gaussian matrix and its mean frequency $\bar{\nu}$ is large compared to the effective spectral width $\Delta\nu$, then the inequality may be replaced by the order of magnitude relation: $(\Delta\tau)(\Delta\nu) \sim (4\pi)^{-1}$.

Note the fact that the time-domain generalized Stokes vector (Stokes self-coherence vector) $\mathbf{S}(\tau)$ is an alternative representation of $\mathbf{\Gamma}(\tau)$ for characterizing temporal polar-coherence of optical wave. Based on time average, $\mathbf{S}(\tau)$, defined as $\mathbf{S}(\tau) = \mathbf{A}\langle \mathbf{E}(t+\tau) \otimes \mathbf{E}^*(t)\rangle$ with $\mathbf{A}$ being a constant matrix for unitary transformation [30], can be understood as the mixed algebra of a matrix product and a matrix direct correlation (i.e., $\mathbf{S}(\tau) = \mathbf{A}[\mathbf{E}(\tau) \circledast \mathbf{E}(\tau)]$). Similarly, we have the spectral Stokes vector $\boldsymbol{\mathcal{S}}(\nu)$, which is the matrix-based Fourier transform of the time-domain generalized Stokes vector [i.e., $\boldsymbol{\mathcal{S}}(\nu) = \Im\{\mathbf{S}(\tau)\}$]. Therefore, we can redefine the coherence time and the effective spectral width of electromagnetic wave in terms of $\mathbf{S}(\tau)$ and $\boldsymbol{\mathcal{S}}(\nu)$. They are

$$(\Delta\tau)^2 = \frac{\int_{-\infty}^{\infty} \tau^2 ||\mathbf{S}(\tau)||^2 \mathrm{d}\tau}{\int_{-\infty}^{\infty} ||\mathbf{S}(\tau)||^2 \mathrm{d}\tau}, \tag{52}$$

$$(\Delta\nu)^2 = \frac{\int_0^{\infty} (\nu - \bar{\nu})^2 ||\boldsymbol{\mathcal{S}}(\nu)||^2 \mathrm{d}\nu}{\int_0^{\infty} ||\boldsymbol{\mathcal{S}}(\nu)||^2 \mathrm{d}\nu}, \tag{53}$$

with the mean frequency $\bar{\nu} = \int_0^{\infty} \nu ||\boldsymbol{\mathcal{S}}(\nu)||^2 \mathrm{d}\nu / \int_0^{\infty} ||\boldsymbol{\mathcal{S}}(\nu)||^2 \mathrm{d}\nu$. By an obvious modification as explained before, these quantities satisfy the same reciprocity inequality: $(\Delta\tau)(\Delta\nu) > (4\pi)^{-1}$. Furthermore, these two sets of matrix-based inverse Fourier transform relationships (i.e., $\mathbf{\Gamma}(\tau) = \Im^{-1}\{\boldsymbol{\mathcal{G}}(\nu)\}$ and $\mathbf{S}(\tau) = \Im^{-1}\{\boldsymbol{\mathcal{S}}(\nu)\}$) can be regarded as the matrix version of the Wiener–Khinchin theorem for stochastic electromagnetic wave.

The definitions of the coherence time and the effective spectral width that we have just discussed are useful when the light is quasi-monochromatic and when its spectral density matrix or spectral Stokes vector has single and reasonably well-defined peak for each element. It becomes more difficult to provide useful definitions of these quantities when $\boldsymbol{\mathcal{G}}(\nu)$ or $\boldsymbol{\mathcal{S}}(\nu)$ has several peaks (as is the case with multimode laser light). There are a multitude of definitions of the coherence time and effective spectral width of light. Following Mandel [31], we can develop a consideration of the extent within a unit cell of photon phase space to define the coherence time for the electromagnetic wave by

$$\Delta\tau = \int_{-\infty}^{\infty} ||\boldsymbol{\gamma}(\tau)||^2 \mathrm{d}\tau = \int_{-\infty}^{\infty} ||\boldsymbol{\gamma}_S(\tau)||^2 \mathrm{d}\tau, \tag{54}$$

where $\boldsymbol{\gamma}(\tau) = \mathbf{\Gamma}(\tau)/||\mathbf{\Gamma}(0)||$ and $\boldsymbol{\gamma}_S(\tau) = \mathbf{S}(\tau)/||\mathbf{S}(0)||$ are the normalized temporal coherence matrix and the normalized Stokes self-coherence vector, respectively, which are justified by the Haagerup's inequality.

Meanwhile, the normalized power spectral density matrix $\hat{\boldsymbol{\mathcal{G}}}(\nu) = \Im\{\boldsymbol{\gamma}(\tau)\} = \boldsymbol{\mathcal{G}}(\nu)/||\int_0^{\infty} \boldsymbol{\mathcal{G}}(\nu)\mathrm{d}\nu||$ and the normalized spectral Stokes vector $\hat{\boldsymbol{\mathcal{S}}}(\nu) = \Im\{\boldsymbol{\gamma}_S(\tau)\} = \boldsymbol{\mathcal{S}}(\nu)/||\int_0^{\infty} \boldsymbol{\mathcal{S}}(\nu)\mathrm{d}\nu||$ can be expressed as their matrix-based Fourier transform, respectively. Therefore, the effective spectral width is then defined as

$$\Delta\nu = \left[\int_0^{\infty} ||\hat{\boldsymbol{\mathcal{G}}}(\nu)||^2 \mathrm{d}\nu\right]^{-1} = \left[\int_0^{\infty} ||\hat{\boldsymbol{\mathcal{S}}}(\nu)||^2 \mathrm{d}\nu\right]^{-1}. \tag{55}$$

With these definitions for the electromagnetic wave, the coherence time is always the reciprocal of the effective spectral width, i.e., $(\Delta\tau)(\Delta\nu) \sim 1$.

7. CONCLUSION

When the vector nature of an optical wave with multiple components is the main concern, the methods of Fourier analysis of matrix functions become indispensable and play a key role in the analysis of both linear and nonlinear phenomena in polarization optics. In this paper, we have developed the matrix-based Fourier analysis as the mathematical tools to be often employed in the analysis of light polarization. We first discussed linear matrix systems and presented mathematical decompositions of matrix signals and matrix systems in two dimensions. After giving the definitions of the matrix (direct) convolution, the matrix (direct) correlation, and the matrix element-wise convolution/correlation as matrix-based integral transformations, we presented some of the basic mathematical properties of the matrix-based Fourier transforms with brief statements of their physical significance. With the aid of the matrix-based Fourier transform theorems, the identity impulse matrix, the matrix sampling theorem, the width and bandwidth of matrix signals, the matrix space-bandwidth product, the uncertainty relation, and the Haagerup's inequality for normalization of matrix direct correction have been discussed as some applications of the matrix-based Fourier analysis. To demonstrate how to apply the proposed mathematical tools in analyzing polarization optics, the coherence time and the effective spectral width of the stochastic electromagnetic wave have been discussed as an application example. The extensions of Fourier analysis and linear systems theory from scalar to matrix functions provided a new richness to the mathematical theory, introducing many new properties that have no direct counterpart in the conventional theory of scalar Fourier optics. Therefore, the formulism and fundamental properties presented in this paper serve as a mathematical foundation in the analysis of polarization-dependent phenomena in optics, information, and communication. Further development of matrix-based Fourier analysis for polarization optics will lead to a new direction referred to as *matrix Fourier optics*.

Funding. Scottish Univeristies Physics Alliance (SSG040).

Acknowledgment. The author thanks Mr. Y. C. Tuo for his help with preparing the figure for this work.

Disclosures. The author declares no conflicts of interest.

Data availability. No data were generated or analyzed in the presented research.

REFERENCES

1. R. C. Jones, "A new calculus for the treatment of optical systems I. Description and discussion of the calculus," J. Opt. Soc. Am. **31**, 488–493 (1941).
2. H. Hurwitz and R. C. Jones, "A new calculus for the treatment of optical systems II. Proof of three general equivalence theorems," J. Opt. Soc. Am. **31**, 493–499 (1941).
3. R. C. Jones, "A new calculus for the treatment of optical systems III. The Sohncke theory of optical activity," J. Opt. Soc. Am. **31**, 500–503 (1941).
4. R. C. Jones, "A new calculus for the treatment of optical systems. IV," J. Opt. Soc. Am. **32**, 486–493 (1942).
5. R. C. Jones, "A new calculus for the treatment of optical systems V. A more general formulation, and description of another calculus," J. Opt. Soc. Am. **37**, 107–110 (1947).
6. R. C. Jones, "A new calculus for the treatment of Optical Systems VI. Experimental determination of the matrix," J. Opt. Soc. Am. **37**, 110–112 (1947).
7. R. C. Jones, "A new calculus for the treatment of optical systems. VII. Properties of the N-matrices," J. Opt. Soc. Am. **38**, 671–685 (1948).
8. R. C. Jones, "New calculus for the treatment of optical systems. VIII. Electromagnetic theory," J. Opt. Soc. Am. **46**, 126–131 (1956).
9. N. G. Parke, "Optical algebra," J. Math. Phys. **28**, 131 (1949).
10. E. Collett, *Field Guide to Polarization: v. FG05* (SPIE, 2005).
11. J. W. Goodman, *Introduction to Fourier Optics*, 4th ed. (W. H. Freeman, 2017).
12. N. A. Rubin, Z. Shi, and F. Capasso, "Polarization in diffractive optics and metasurfaces," Adv. Opt. Photon. **13**, 836–970 (2021).
13. W. Urbańczyk, "Optical transfer function for imaging systems which change the state of light polarization," Opt. Acta **33**, 53–62 (1986).
14. J. P. McGuire and R. A. Chipman, "Diffraction image formation in optical systems with polarization aberrations. I: Formulation and example," J. Opt. Soc. Am. A **7**, 1614–1626 (1990).
15. R. A. Chipman, W. S. T. Lam, and G. Young, *Polarized Light and Optical Systems* (CRC Press, 2019).
16. I. Moreno, M. J. Yzuel, J. Campos, *et al.*, "Jones matrix treatment for polarization Fourier optics," J. Mod. Opt. **51**, 2031–2038 (2004).
17. O. Korotkova, *Theoretical Statistical Optics* (World Scientific, 2021).
18. W. Wang and M. Takeda, "Matrix-based integral transformations for Stokes imaging with partially polarized and partially coherent light," J. Opt. Soc. Am. A **41**, 774–788 (2024).
19. A. Papoulis, *The Fourier Integral and Its Applications* (McGraw-Hill, 1968).
20. R. N. Bracewell, *The Fourier Transform and Its Applications*, 2nd ed. (McGraw-Hill, 1994).
21. R. M. Gray and J. W. Goodman, *Fourier Transforms: An Introduction for Engineers* (Kluwer Academic, 1995).
22. L. Mandel and E. Wolf, *Optical Coherence and Quantum Optics* (Cambridge University, 1995).
23. E. Wolf, *Introduction to the Theory of Coherence and Polarization of Light* (Cambridge University, 2007).
24. R. A. Horn and C. R. Johnson, *Matrix Analysis* (Cambridge University, 2012).
25. I. M. Gelfand and G. E. Shilov, *Generalized Functions* (Academic, 1966).
26. W. Wang, S. G. Hanson, and M. Takeda, "Autocorrelation functions and power spectral densities of the Stokes parameters in a polarization speckle pattern," J. Opt. Soc. Am. A **40**, 165–174 (2023).
27. G. Pisier, *Introduction to Operator Space Theory* (Cambridge University, 2003).
28. J. W. Goodman, *Statistical Optics*, 2nd ed. (Wiley-Blackwell, 2015).
29. W. Wang and M. Takeda, "Coherence current, coherence vortex, and the conservation law of coherence," Phys. Rev. Lett. **96**, 223904 (2006).
30. E. L. O'Neill, *Introduction to Statistical Optics* (Addison-Wesley, 1963).
31. L. Mandel, "Fluctuations of photon beams: the distribution of the photo-electrons," Proc. Phys. Soc. London **74**, 233–243 (1959).