Unitary algorithm for non-separable linear canonical transforms applied to iterative phase retrieval

Liang Zhao, John T. Sheridan, *Fellow, OSA,* and John J. Healy, *Member, IEEE,*

***Abstract*—Phase retrieval is an important tool with broad applications in optics. The Gerchberg-Saxton algorithm has been a workhorse in this area for many years. The algorithm extracts phase information from intensities captured in two planes related by a Fourier transform. The ability to capture the two intensities in domains other than the image and Fourier plains adds flexibility; various authors have extended the algorithm to extract phase from intensities captured in two planes related by other optical transforms, e.g. by free space propagation or a fractional Fourier transform. These generalisation are relatively simple once a unitary discrete transform is available to propagate back and forth between the two measurement planes. In the absence of such a unitary transform, errors accumulate quickly as the algorithm propagates back and forth between the two planes. Unitary transforms are available for many separable systems, but there has been limited work reported on non-separable systems other than the gyrator transform. In this paper, we simulate a non-separable system in a unitary way by choosing an advantageous sampling rate related to the system parameters. We demonstrate a simulation of phase retrieval from intensities in the image domain and a second domain related to the image domain by a non-separable linear canonical transform. This work may permit the use of non-separable systems in many design problems.**

***Index Terms*—Linear canonical transforms, Phase retrieval, Fourier optics and signal processing, Image reconstruction techniques.**

I. Introduction

PHASE retrieval is the subject of continuing attention in the literature, and has widespread use in astronomy, electron microscopy, holographic imaging, optical design, wavefront sensing, and other areas [1], [2], [3]. The Gerchberg-Saxton algorithm is a classic iterative phase retrieval algorithm that has been studied extensively [4]. In its original form, it takes an input of the intensity of a wave field in two planes related by an optical Fourier transform. It has been extended to allow other transforms take the place of the optical Fourier transform, such as the Fresnel transform [5] (including by means of a chirp placed in the Fourier plane [6]), fractional Fourier transform [5], [7], and the 1-dimensional linear canonical transform (1D-LCT) [8]. Phase retrieval based on intensities captured in planes related by one specific non-separable transform, the gyrator transform, have also been reported [9], [10], [11], [12], [13], [14], [15]. The advantage of using these alternative systems is that increased flexibility for the system designer. There can be practical constraints on this choice, or the motivation can be to maximise the information captured by the camera and hence the resolution, window size, or even precision of the image [16]. It has been shown that the unitarity of the transform used is critical to the success of the algorithm [8], [17], [18], [19]. Details of the algorithm are given in Fig. 1. Many variations which differ on the known information and/or stopping criterion are also widely used.

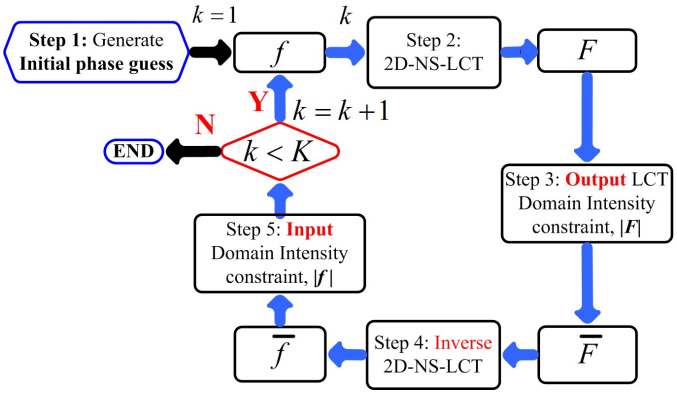

Fig. 1. Flowchart for the GS algorithm.

A. Definitions

We now introduce the optical transforms used in the algorithm. First, recall that the ABCD matrix characterises a first order optical system in both ray-tracing [20] and wave optics [21].

$$M = \begin{pmatrix} A & B \\ C & D \end{pmatrix}. \tag{1}$$

For a 2-dimensional system, M is a block matrix with 2×2 submatrices A, B, C and D. M must be a symplectic matrix. For a separable system, the ijth element of M is zero if i and j have different parities. Systems with circular symmetry have this property, which conveniently allows us to simulate the system using a one-dimensional transform, e.g. we can use the row-column algorithm to simulate a 2D Fourier transforming system using a 1D discrete Fourier transform algorithm.

Non-separable systems are not as common, but can arise in many situations. For example, the following kinds of systems are non-separable:

- Are non-orthogonal, non-axially symmetric, or contain anamorphic lenses [22], [20];
- Involve coupling and/or shearing operations between components along the different dimensions, e.g., a co-ordinate transform, see Table II in [23]; and

L. Zhao was with the Insight Centre for Data Analytics, University College Dublin, Ireland.

J. J. Healy and J. T. Sheridan are with School of Electrical and Electronic Engineering and the IoE^2 Lab, University College Dublin, Ireland. e-mail: (john.healy@ucd.ie)

Manuscript received December 8, 2016; revised Month Day, Year.

- Involve rotations between any arbitrary planes in phase space, e.g., the gyrator transform [24], [25], [26], [27], [28];

Non-separable systems can be modelled using the two dimensional non-separable linear canonical transform (2D-NS-LCT). When $\det(B) \neq 0$, the continuous 2D-NS-LCT, $G(x, y)$, of a wave field $g(x, y)$ can be given by

$$\begin{aligned} G(x,y) &= L_M\{g(x',y')\}(x,y) \\ &= \int\int_{-\infty}^{\infty} g(x',y')K(x,x',y,y',M)dxdy \end{aligned} \tag{2}$$

where $K(x, x', y, y', M)$ is the kernel of the transform,

$$\begin{aligned} K(x,x',y,y',M) = &\frac{1}{\sqrt{j\det(B)}} \exp\left[j\pi \frac{k_1x^2 + k_2xy + k_3y^2}{2\det(B)}\right] \\ &\times \exp\frac{j2\pi}{\det(B)}\Big[(-b_{22}x + b_{12}y)x' \\ &+ (b_{21}x + b_{12}y)y' \\ &+ \frac{1}{2}(p_1x'^2 + p_2x'y' \\ &+ p_3y'^2)\Big] \end{aligned} \tag{3}$$

where

$$\begin{aligned} k_1 = d_{11}b_{22} - d_{12}b_{21}, \;\; p_1 = a_{11}b_{22} - a_{21}b_{12} \\ k_2 = 2(-d_{11}b_{12} + d_{12}b_{11}), \;\; p_2 = 2(a_{12}b_{22} - a_{22}b_{12}) \\ k_3 = -d_{21}b_{12} + d_{22}b_{11}, \;\; p_3 = -a_{12}b_{21} + a_{22}b_{11} \end{aligned} \tag{4}$$

For a full definition including the various special cases when $\det(B) = 0$, the reader is referred to the references [29], [30], [23], [31]. The inverse 2D-NS-LCT recovers $g(x, y)$ from $G(x', y')$ by,

$$g(x,y) = L_{M^{-1}}\{G(x',y')\}(x,y) \tag{5}$$

The continuous NS-LCT is additive [30], [23], meaning

$$L_{M_2}\{L_{M_1}\{g\}\} = L_{M_3}\{g\} \tag{6}$$

where $M_3 = M_2M_1$. When $M_2 = M_1^{-1}$, we obtain a statement that the transform is unitary

$$L_{M^{-1}}\{L_M\{g\}\} = L_{M^{-1}M}\{g\} = L_I\{g\} = g. \tag{7}$$

The difficulty is that the properties described in Eqs 6 and 7 are not easily preserved in discrete linear canonical transforms [8], [32], especially in the non-separable case [33], [23]. For phase retrieval, we require a discrete transform that approximates the 2D-NS-LCT [30], [34] and furthermore that is unitary.

II. A UNITARY ALGORITHM TO CALCULATE THE 2D NON-SEPARABLE LCT

For non-separable transforms, the output sampling points generally lie on a skewed grid, which may require interpolation to render [23]. This is evident in the decomposition given in Eq. 8, which is valid if B has an inverse. The decomposition shows that we can compute most non-separable transforms as a series of operations. A discrete input, e.g. simulated data or the output of a digital camera, is defined on a rectilinear grid of sample locations. The first three operations in the decomposition of Eq. 8, multiplication by a chirp, a Fourier transform and a second chirp multiplication, do not affect the locations of these points (note that we read the sequence of matrices right to left). Furthermore, they are unitary operations. The final operation, the affine transformation, can affect the locations of the samples. Interpolation operations are computationally intensive and a source of error, leading Ding *et al.* to develop algorithms that require fewer interpolations [34]. Our algorithm is based on the observation that for certain combinations of sampling rates and system parameters, the output sampling points lie on a Cartesian grid, see Fig. 2. This allows us to dispense with *any* interpolation, yielding a discrete, unitary transformation suitable for iterative calculations.

$$\begin{aligned} M = \begin{pmatrix} A & B \\ C & D \end{pmatrix} = &\begin{pmatrix} B & 0 \\ 0 & (B^T)^{-1} \end{pmatrix} \begin{pmatrix} I & 0 \\ B^TD & I \end{pmatrix} \\ &\times \begin{pmatrix} 0 & I \\ -I & 0 \end{pmatrix} \begin{pmatrix} I & 0 \\ B^{-1}A & I \end{pmatrix}. \end{aligned} \tag{8}$$

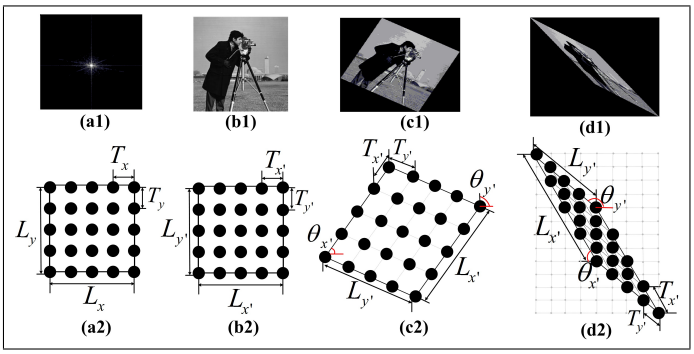

Fig. 2. (a1) The magnitude of an example input to system described by an LCT. (b1) The output if the system is separable. (c1) The output if the system is a typical non-separable system. (d1) The output of a non-separable system with appropriate sampling rates per [23]. Note how the samples lie on a Cartesian grid. (a2-d2) The corresponding sampling grids.

The conditions for which the output samples lie on a Cartesian grid are given in Tab. I; it is assumed that the same number of samples are used in both directions, and that uniform sampling is used - hence the constraints are on the width of the input image in x and y. These conditions can be derived using straightforward geometrical considerations, which we now outline.

We now sketch a proof of this property. While this may seem a little opaque, the geometrical interpretation is very simple. We are simply making sure that the co-ordinate transform lines up the samples on simple locations, as in Fig.

TABLE I
Constraints for samples lying on Cartesian grid

B	Constraints
Diagonal	None
Antidiagonal	None
Lower triangular	$L_y = l\frac{b_{22}}{b_{21}}L_x,\ l \in \mathbb{N}$
Upper triangular	$L_y = \frac{1}{l}\frac{b_{12}}{b_{11}}L_x,\ l \in \mathbb{N}$
Anti lower triangular	$L_y = \left\lvert\frac{b_{22}}{b_{21}}\right\rvert L_x$
Anti upper triangular	$L_y = l\frac{b_{12}}{b_{11}}L_x,\ l \in \mathbb{N}$
$b_{ij} \neq 0\ \forall i,j$	$\frac{b_{11}b_{22}}{b_{12}b_{21}} = 1 + \frac{1}{kl},\ L_y = l\frac{b_{22}}{b_{21}}L_x,\ k,l \in \mathbb{N}$

(2 d2). In [23], we showed that when $b_{22} \neq 0$ and $\det(B) \neq 0$, the sub-matrix B can be decomposed as follows.

$$B = \begin{pmatrix} 1 & \frac{b_{12}}{b_{22}} \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ \frac{b_{21}b_{22}}{\det(B)} & 1 \end{pmatrix} \begin{pmatrix} \frac{\det(B)}{b_{22}} & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & b_{22} \end{pmatrix} \quad (9)$$

Eq. 9 indicates that B can be decomposed into a pair of 1-D magnifications followed by a pair of 1-D sheering operations. The magnifications have no effect on whether the output samples lie on a Cartesian grid. Hence, we only need concern ourselves with the sheering operations.

We define $p = \frac{T_y}{T_x}$, where T_x and T_y are the sample spacing in x and y in the input plane.

The first operations that alter the grid spacing are the Fourier transform (see Eq. (8)) and the two scaling operations (Eq. (9)). After these operations, the sampling grid has spacing L'_x/N and L'_y/N, where,

$$L'_x = \frac{\det(B)}{b_{22}T_x}, \quad (10)$$

and

$$L'_y = \frac{b_{22}}{T_y} = \frac{b_{22}}{pT_x}. \quad (11)$$

At this point, we have a rectangular grid of samples with known spacings. Basic geometry will obtain the locations of the samples in such a grid after a sheering in x and a second sheering in y, yielding conditions on the ratio of the sampling input rates p in terms of the parameters of the skewing operations. Given a sampling point (x, y), it is mapped by the two skewing operations to the location $(x - s_2y, y(1 + s_1s_2) - s_1x)$, where s_1 is the parameter relating to the first skewing (in y), and s_2 is the parameter relating to the second skewing (in x). We need only constrain the two co-ordinates to be any integer multiple of the sampling period to ensure the new sample location lies on the grid. From such considerations, we can show that a Cartesian output grid can be obtained providing the following conditions are met:

$$kp = \frac{b_{12}b_{22}}{\det(B)}, \quad (12)$$

and

$$\frac{p}{l} = \frac{b_{22}}{b_{21}}, \quad (13)$$

where $k, l \in \mathbb{N}$. From these conditions, we can find the constraints given in Tab. (I), except for the 2nd and 6th cases, when the decomposition of B given in Eq. (9) is invalid because $b_{22} = 0$.

We are preparing detailed proofs for publication in a more verbose format.

III. Demonstration of phase retrieval for a non-separable system

In this section, we demonstrate the efficacy of our method. For the sake of a simplified calculation, we choose a system characterised by the following matrix.

$$M = \begin{pmatrix} 0 & B \\ B^{-1} & 0 \end{pmatrix}, \quad (14)$$

where $B = \left(\begin{smallmatrix} 1 & 0.5 \\ 0.5 & 0.5 \end{smallmatrix}\right)$. From Eq. (8), this 2D-NS-LCT system can be evaluated by $M = \left(\begin{smallmatrix} B & 0 \\ 0 & (B^T)^{-1} \end{smallmatrix}\right)\left(\begin{smallmatrix} 0 & I \\ -I & 0 \end{smallmatrix}\right)$, i.e. a 2D Fourier transform followed by a coordinate transform. As noted previously, the omitted chirp multiplications don't alter the sample locations and so the example is not compromised by this simplification. The output will appear on a Cartesian grid for this system providing the sampling rates in x and y are the same, per the last condition in Tab. I.

The phase retrieval results obtained by both the proposed algorithm and the standard algorithm are shown in Fig. 3. Both algorithms consist of a Fourier transform followed by a co-ordinate transform. In the proposed algorithm, we use the sampling rates specified by Tab. (I) (arbitrarily choosing $k = l = 1$), which reduces the co-ordinate transform to pixel swapping. In the standard algorithm, we perform the co-ordinate transform by means of MATLAB®'s `imwarp` function. In Fig. 3(a), we plot the log base 10 of the mean squared error of the recovered image against the number of iterations, k. After the 1000st iteration, this metric is -307 dB for our proposed algorithm, which is much less than the corresponding -1.57 dB for the reference algorithm. The image reconstructed by our algorithm is shown in Fig. 3(b2), and is recognisably a skewed version of the cameraman image. The reference algorithm fails to produce a recognisable image, see Fig. 3(c2). Hence, we have demonstrated phase retrieval for a general non-separable LCT for the first time.

IV. Conclusion

The calculation of 2D non-separable LCTs is complicated by the presence of an affine transformation which results in a skewed sampling grid. The interpolation required to cope with this skewed grid adds time and error to the simulation of these optical propagation problems. Iterative calculations are especially sensitive to this problem. We have shown that we can choose the sampling rate to obtain a simplified affine transformation, which conveniently maps to a Cartesian grid.

We have made use of this result to perform iterative phase retrieval in a non-separable optical system. The complexity of our calculation is fundamentally limited by the Fourier transform stages. However, the required sampling rates (per Tab. (I)) may prove onerous in some practical cases.

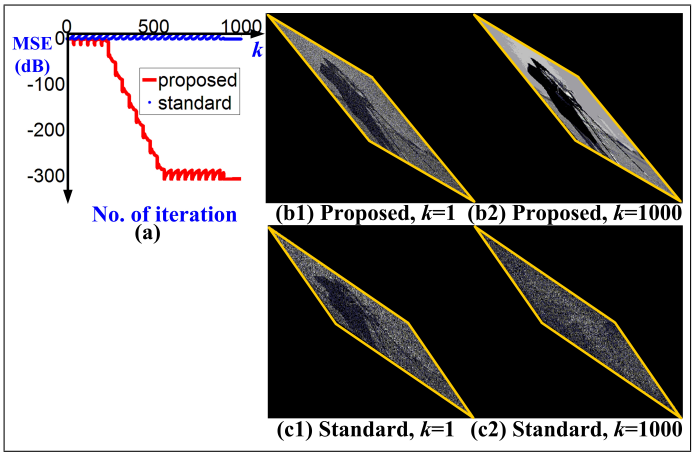

Fig. 3. Simulation result: (a) 10 $\log_{10}$(MSE) of the recovered images; (b1) and (b2) are the reconstructed image after the 1st and 1000th iteration by the proposed algorithm; (c1) and (c2) are the reconstructed image after the 1st and 1000th iteration by the standard algorithm.

While we have demonstrated convergence, we have not proven it. In fact, it is not generally possible to mathematically prove convergence to zero: even Feinups famous paper [17] only proves that the error does not increase for the Gerschberg-Saxton algorithm; that result applies no matter the propagation algorithm used, providing it is unitary, and so it applies to the algorithm in this paper.

ACKNOWLEDGMENT

LZ thanks the UCD-China Scholarship Council (CSC) joint scholarship and SPIE Optics and Photonics Education Scholarship. JJH thanks the NUI Fellowship in the Sciences. The authors also acknowledge the support of Science Foundation Ireland, the Irish Research Council and Enterprise Ireland under the National Development Plan.

REFERENCES

[1] J. R. Fienup, "Phase retrieval algorithms: a personal tour," *Appl. Opt.*, vol. 52, no. 1, pp. 45–56, Jan 2013.

[2] Y. Shechtman, Y. C. Eldar, O. Cohen, H. N. Chapman, J. Miao, and M. Segev, "Phase retrieval with application to optical imaging: A contemporary overview," *IEEE Signal Process. Mag.*, vol. 32, no. 3, pp. 87–109, May 2015.

[3] C.-L. Guo, S. Liu, and J. T. Sheridan, "Iterative phase retrieval algorithms. i: optimization," *Appl. Opt.*, vol. 54, no. 15, pp. 4698–4708, May 2015.

[4] R. W. Gerchberg and W. O. Saxton, "A practical algorithm for the determination of phase from image and diffraction plane pictures," *Optik*, vol. 35, p. 237, 1972.

[5] Z. Zalevsky, R. G. Dorsch, and D. Mendlovic, "Gerchberg–Saxton algorithm applied in the fractional Fourier or the Fresnel domain," *Opt. Lett.*, vol. 21, no. 12, pp. 842–844, Jun 1996.

[6] C. Falldorf, M. Agour, C. v. Kopylow, and R. B. Bergmann, "Phase retrieval by means of a spatial light modulator in the Fourier domain of an imaging system," *Appl. Opt.*, vol. 49, no. 10, pp. 1826–1830, Apr 2010.

[7] B. M. Hennelly, D. Kelly, A. Corballis, and J. T. Sheridan, "Phase retrieval using theoretically unitary discrete fractional Fourier transform," in *Proc. SPIE*, vol. 5908, 2005, pp. 59 080D–59 080D–12.

[8] L. Zhao, J. J. Healy, and J. T. Sheridan, "Unitary discrete linear canonical transform: analysis and application," *Appl. Opt.*, vol. 52, no. 7, pp. C30–C36, 2013.

[9] H. Li and Y. Wang, "Double-image encryption based on iterative gyrator transform," *Opt. Comm.*, vol. 281, no. 23, pp. 5745 – 5749, 2008. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0030401808008523

[10] A. Cámara, T. Alieva, J. A. Rodrigo, and M. L. Calvo, "Tomographic reconstruction of the Wigner distribution of non-separable beams," *PIERS Cambridge*, 2010.

[11] J. A. Rodrigo, H. Duadi, T. Alieva, and Z. Zalevsky, "Multi-stage phase retrieval algorithm based upon the gyrator transform," *Opt. Express*, vol. 18, no. 2, pp. 1510–1520, Jan 2010. [Online]. Available: http://www.opticsexpress.org/abstract.cfm?URI=oe-18-2-1510

[12] Z. Liu, Q. Guo, L. Xu, M. A. Ahmad, and S. Liu, "Double image encryption by using iterative random binary encoding in gyrator domains," *Opt. Express*, vol. 18, no. 11, pp. 12 033–12 043, May 2010. [Online]. Available: http://www.opticsexpress.org/abstract.cfm?URI=oe-18-11-12033

[13] Z. Liu, L. Xu, Q. Guo, C. Lin, and S. Liu, "Image watermarking by using phase retrieval algorithm in gyrator transform domain," *Opt. Comm.*, vol. 283, no. 24, pp. 4923 – 4927, 2010. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0030401810007716

[14] Z. Liu, L. Xu, C. Lin, J. Dai, and S. Liu, "Image encryption scheme by using iterative random phase encoding in gyrator transform domains," *Opt. Las. Eng.*, vol. 49, no. 4, pp. 542 – 546, 2011. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0143816610002708

[15] Z. Liu, C. Guo, J. Tan, Q. Wu, L. Pan, and S. Liu, "Iterative phase-amplitude retrieval with multiple intensity images at output plane of gyrator transforms," *J. Optics*, vol. 17, no. 2, p. 025701, 2015. [Online]. Available: http://stacks.iop.org/2040-8986/17/i=2/a=025701

[16] A. Ozcelikkale, H. M. Ozaktas, and E. Arikan, "Signal recovery with cost-constrained measurements," *IEEE Transactions on Signal Processing*, vol. 58, no. 7, pp. 3607–3617, July 2010.

[17] J. R. Fienup, "Phase retrieval algorithms: A comparison," *Appl. Opt.*, vol. 21, no. 15, pp. 2758–2769, 1982.

[18] G.-Z. Yang, B.-Z. Dong, B.-Y. Gu, J.-Y. Zhuang, and O. K. Ersoy, "Gerchberg–Saxton and Yang–Gu algorithms for phase retrieval in a nonunitary transform system: a comparison," *Appl. Opt.*, vol. 33, no. 2, pp. 209–218, Jan 1994.

[19] Y. Zhang, B.-Z. Dong, B.-Y. Gu, and G.-Z. Yang, "Beam shaping in the fractional Fourier transform domain," *J. Opt. Soc. Am. A*, vol. 15, no. 5, pp. 1114–1120, May 1998.

[20] G. Kloos, *Matrix methods for optical layout*. SPIE, Bellingham, USA, 2007.

[21] J. W. Goodman, *Introduction to Fourier optics, 3rd Ed.* Roberts and Company Publishers, 2005.

[22] A. E. Siegman, *Lasers*. University Science Books, Palo Alto, USA, 1986.

[23] L. Zhao, J. J. Healy, and J. T. Sheridan, "The 2d non-separable linear canonical transform: Sampling theorem and unitary discretization," *J. Opt. Soc. Am. A*, vol. 31, no. 12, pp. 2631–2641, 2014.

[24] J. A. Rodrigo, T. Alieva, and M. L. Calvo, "Optical system design for orthosymplectic transformations in phase space," *J. Opt. Soc. Am. A*, vol. 23, no. 10, pp. 2494–2500, 2006.

[25] ——, "Experimental implementation of the gyrator transform," *J. Opt. Soc. Am. A*, vol. 24, pp. 3135–3139, 2007.

[26] ——, "Gyrator transform: Properties and applications," *Opt. Express*, vol. 15, pp. 2190–2203, 2007.

[27] ——, "Applications of gyrator transform for image processing," *Opt. Commun.*, vol. 278, pp. 279–284, 2007.

[28] S.-C. Pei and J.-J. Ding, "Properties, digital implementation, applications, and self-image phenomena of the gyrator transform," in *17th European Signal processing conference*, 2009, pp. 441–445.

[29] S.-C. Pei, "Two-dimensional affine generalized fractional Fourier transform," *IEEE Trans. Sig. Proc.*, vol. 49, pp. 878–897, 2001.

[30] A. Koç, H. M. Ozaktas, and L. Hesselink, "Fast and accurate computation of two-dimensional non-separable quadratic-phase integrals," *J. Opt. Soc. Am. A*, vol. 27, pp. 1288–1302, 2010.

[31] J. J. Healy, M. A. Kutay, H. M. Ozaktas, and J. T. Sheridan, *Linear Canonical Transforms*. Springer: New York, NY, USA, 2016.

[32] L. Zhao, J. J. Healy, and J. T. Sheridan, "Constraints on additivity of the 1d discrete linear canonical transform," *Appl. Opt.*, vol. 54, no. 33, pp. 9960–9965, 2015.

[33] ——, "Unitary implementation of the discrete two dimensional non-separable linear canonical transform," in *Proc. SPIE*, vol. 9216, 2014, pp. 921 610–921 610.

[34] J.-J. Ding, S.-C. Pei, and C.-L. Liu, "Improved implementation algorithms of the two-dimensional non-separable linear canonical transform," *J. Opt. Soc. Am. A*, vol. 29, pp. 1615–1624, 2012.