Invited Paper

Optimal representation and processing of optical signals in quadratic-phase systems

Sercan Ö. Arık a,*, Haldun M. Ozaktas b

a Stanford University, Department of Electrical Engineering, 94305 Stanford, CA, USA
b Bilkent University, Department of Electrical Engineering, 06800 Bilkent, Ankara, Turkey

A R T I C L E   I N F O

Article history:
Received 6 November 2015
Received in revised form 11 December 2015
Accepted 12 December 2015
Available online 24 December 2015

Keywords:
ABCD systems
Quadratic-phase systems
Fourier optics
Sampling
Fractional Fourier transform
Linear canonical transform

A B S T R A C T

Optical fields propagating through quadratic-phase systems (QPSs) can be modeled as magnified fractional Fourier transforms (FRTs) of the input field, provided we observe them on suitably defined spherical reference surfaces. Non-redundant representation of the fields with the minimum number of samples becomes possible by appropriate choice of sample points on these surfaces. Longitudinally, these surfaces should not be spaced equally with the distance of propagation, but with respect to the FRT order. The non-uniform sampling grid that emerges mirrors the fundamental structure of propagation through QPSs. By providing a means to effectively handle the sampling of chirp functions, it allows for accurate and efficient computation of optical fields propagating in QPSs.

© 2015 Elsevier B.V. All rights reserved.

1. Introduction

Quadratic-phase systems (QPSs), also known as first-order optical systems or ABCD systems, are a very general family of optical systems encompassing arbitrary concatenations of various components such as thin lenses and sections of free space in the Fresnel approximation, as well as quadratic graded-index media [1–4]. Mathematically, QPSs are referred to as linear canonical transforms [5,7,6,8–11]. In this paper, we derive the optimal sampling grid for this general family. This non-redundant grid of sample points mirrors the physical structure of QPSs and enables their accurate and fast simulation. The fractional Fourier transform (FRT) plays a fundamental role in the analysis of QPSs. We also analyze the evolution of spatial information along the longitudinal direction and show that the spherical reference surfaces should be equally spaced with respect to the fractional Fourier transform order. Our results are relevant to work on both sampling and fast computation of light fields propagating in QPSs [12–26].

2. Decomposition of propagation in quadratic-phase systems

We will use \( \hat{f}(x) \) and \( \hat{f}(\sigma_x) \) to represent an optical signal in the space domain and the frequency domain, respectively. Although we work with functions of a single variable for sake of simpler analysis, our results can be generalized to two dimensions. It will be useful to introduce dimensionless variables \( u = x/s \) and \( \mu = s\sigma_x \) for space and frequency, where \( s \) is a scaling parameter with units of length. We now define \( \hat{f}(x) \equiv (1/\sqrt{s})f(u) \) and \( \hat{f}(\sigma_x) \equiv \sqrt{s}F(\mu) \). The functions \( f(u) \) and \( F(\mu) \) are the space- and frequency-domain functions that take dimensionless arguments. More information on this dimensional normalization process may be found in [1].

The FRT can be viewed as the “fractional operator power” of the common Fourier transform (FT). One way of defining the \( \alpha \)th order FRT of a function \( f(u) \), which we denote by \( f_{\alpha}(u) \), is

\[
f_{\alpha}(u) = A_{\alpha} \int_{-\infty}^{\infty} \exp(iu^2/2) \exp(iu \cos(\alpha/2) - \sin(\alpha/2)) f(u) du
\]

(1)

See [1] for subtleties in the definition of fractional operator powers as well as alternative ways of defining the FRT. Here \( A_{\alpha} = \sqrt{1 - i \cot(\alpha/2)} \). When \( \alpha = 1 \), we obtain the common FT, so that the FRT can be seen as a generalization of the common FT. Viewed in the space–frequency plane (phase space), the act of taking the FRT of a signal results in a \( \alpha = \alpha/2 \) rotation of its Wigner distribution. This can be expressed as a relation between the Wigner distribution of \( f(u) \) and the Wigner distribution of \( f_{\alpha}(u) \) as follows [27]:

\[
W_{f_{\alpha}}(u, \mu) = W_f(\mu \cos \alpha - \mu \sin \alpha, u \sin \alpha + \mu \cos \alpha).
\]

(2)

Quadratic-phase systems (QPSs) are unitary. The input \( \hat{f}(x) \) leads to an output \( \hat{g}(x) \) given by
\[ \hat{\hat{g}}(x) = \sqrt{b} e^{-ia/4} \int_{-\infty}^{\infty} e^{bx^4} e^{-a(x^2 + b^2x^2)} f(x') dx'. \]  

(3)

They are often characterized by the ABCD parameters which satisfy

\[ AD - BC = 1 \]

and are defined as

\[ A = \frac{\lambda}{\beta}, \quad B = \frac{1}{\beta}, \quad C = -\beta + \frac{\pi}{\beta}, \quad D = \frac{a}{\beta}. \]

(4)

Consider a quasi-monochromatic optical signal with wavelength \( \lambda \), Fresnel diffraction and passage through a thin lens are special kinds of QPSs. For Fresnel propagation over a distance \( d \), we have \( A = D = 1, C = 0 \), and \( B = id \). For a thin lens with focal length \( f \) we have \( A = D = 1, B = 0, \) and \( C = -1/f \).

Quadratic-phase systems have many decompositions. A decomposition is the breaking down of the system into consecutive simpler parts. One of the possible decompositions involves three stages. The first is a FRT operation, the second is a magnification operation, and the final stage is a chirp multiplication operation [28–32]:

\[ \hat{g}(x) = e^{2\pi id/\beta - i\omega x^2/4} \exp \left( \frac{1}{2\beta} \int_{-\infty}^{\infty} \left( \frac{\ln x}{\lambda} \right) \left( \frac{x}{\lambda} \right) \frac{1}{\sm} \right). \]

Here \( a, M, \) and \( R \) are defined through

\[ \tan \left( \frac{a x}{2} \right) = \frac{1}{R^2 A}. \]

(6)

\[ M = \sqrt{A^2 + B^2} \]

(7)

\[ \frac{1}{R} = \frac{1}{s^2} \frac{B/A}{s^2} + \frac{C}{A}. \]

(8)

The ambiguity in inverting the tangent in Eq. (6) should be resolved by choosing \( a \) in \( 0 < a < 2 \) for \( B > 0 \) and in \( 2 < a < 4 \) for \( B < 0 \). We will make a number of observations on these equations. First, we note that the unit-magnitude \( e^{2\pi id/\beta - i\omega x^2/4} \) terms are constant and unimportant for our purposes. Furthermore, the chirp multiplication term can be eliminated, if we decide to observe the output not on a planar surface, but on a spherical reference surface with radius as given by the value of \( R \) above. In this case, we observe a magnified version of the FRT of the input, with the magnification given by the value of \( M \) above. A final note is that, Eq. (5) is valid no matter what we choose \( s \) to be. This decomposition will constitute the basis for our derivation of the optimal sampling grid.

3. Transverse sampling spacing

A discussion of sampling often begins with assumptions on the extent of the signals in both space and frequency, the latter often called the bandwidth. An alternative approach is to specify the extent of the signals in the space–frequency plane. Our beginning assumption will be to specify the space–frequency region to which the signal is confined. Here, "confined" means that a sufficiently large percentage of the total energy is contained in that region.

We will take the \( z = 0 \) plane as our input plane. We assume that the space–frequency region of confinement of the input signal at this plane is an ellipse with diameters denoted by \( \Delta x \) and \( \Delta \sigma_x \) (Fig. 1(a)). Note that this implies that the space extent of the signal is \( \Delta x \) and that the frequency extent of the signal is \( \Delta \sigma_x \). How many samples are needed to represent the signal? The sampling theorem of Nyquist–Shannon requires a sampling interval of \( 1/\Delta \sigma_x \).

Over a spatial extent of \( \Delta x \) this means \( N = \Delta x/(1/\Delta \sigma_x) = \Delta x \sigma_x \) samples. This is known as the space–bandwidth product. The same derivation can be repeated in dimensionless coordinates. Now the ellipse diameters (and thus also the space and frequency extents) are \( \Delta x/s \) and \( \Delta \sigma_x/s \), which lead to the same value of \( N \). The space–bandwidth product is invariant under scalings and does not depend on \( s \).

We now consider a QPS between the \( z = 0 \) and \( z = d \) planes characterized by the parameters \( ABCD \). For example, in a system made up of lenses separated by sections of free space, these \( ABCD \) parameters will depend on the focal lengths and locations of the lenses. We will first determine the spatial extent of the output signal \( \hat{g}(x) \) observed at \( z = d \), given an input signal \( \hat{f}(x) \) at \( z = 0 \). It is known that the Wigner distributions of \( \hat{f}(x) \) and \( \hat{g}(x) \) have the following relationship [13,33–36]:

\[ \hat{W}_g(x, \sigma_x) = \hat{W}_f(dx - B \sigma_x - Cx + A \sigma_x). \]

(9)

Based on our assumption that the initial space–frequency distribution of the signal is well-confined to an elliptical region with diameters \( \Delta x \) and \( \Delta \sigma_x \), and using Eq. (9), it is possible to show that the output space–frequency distribution will have a spatial extent

\[ \Delta x' = \sqrt{\Delta x^2 D^2 + \Delta \sigma_x^2 B^2}. \]

(10)
and a frequency extent
\[ \Delta \sigma_x = \sqrt{\Delta x^2 C^2 + \Delta \sigma_f^2}. \] (11)

The derivation of these results are similar to those in [37] and are thus not repeated. On the output plane, the space–bandwidth product is \( N' = \Delta x \Delta \sigma_x \). It can be shown that we always have \( N' \geq N \), with equality in the trivial case \( B = C = 0 \) and \( A = 1/D \) (simple scaling system). We have used double primes since we are saving the use of single primes for the output on the spherical reference surfaces. The linearly distorted region of confinement on the output plane at \( z = d \) is illustrated in Fig. 1(b).

Now we wish to obtain an expression for the spatial extent \( \Delta x' \) of the output field on the spherical reference surface. We refer to Eq. (5). First we choose a scale parameter \( s \) and map the ellipse enclosing most of the signal energy, into a dimensionless space–frequency plane. Here, the diameters are \( \Delta x/s \) and \( s \Delta \sigma_x \). Now, we apply a FRT of order \( a \), which results in a rotation by an angle \( a = as/2 \). Using Eq. (2), we find that the rotated ellipse has a spatial extent of \( \sqrt{\Delta \sigma_x \sin \alpha} \), \( \Delta \sigma_x \) in dimensionless coordinates. To express this in dimensionless coordinates, we multiply by \( s \). Finally, we multiply with the magnification \( M \) to obtain the spatial extent of the output signal on the spherical reference surface:

\[ \Delta x' = M \sqrt{(s^2 \Delta \sigma_x \sin \alpha)} + (\Delta x \cos \alpha)^2. \] (12)

Following a similar argument, we can derive the spatial frequency extent \( \Delta \sigma'_x \) of the output signal on the spherical reference surface as \( \sqrt{s \Delta \sigma_x \cos \alpha} + (\Delta \sigma_x \sin \alpha)^2 \). Dividing by \( s \) to express in dimensionless coordinates, and finally dividing by the magnification \( M \), we obtain

\[ \Delta \sigma'_x = \frac{1}{M} \sqrt{(s \Delta \sigma_x \cos \alpha)} + (\Delta \sigma_x \sin \alpha)^2. \] (13)

Now, it is possible to obtain the space–bandwidth product on the spherical reference surface: \( N' = \Delta x \Delta \sigma'_x \) from the above expressions:

\[ N' = (s^2 \Delta \sigma_x \sin^2 \alpha + \Delta \sigma_x^2 \Delta x^2 \sin^2 \alpha + \cos^4 \alpha) + \Delta \sigma_x^2 \sin^2 \alpha \cos^4 \alpha \times s^4 / 2. \] (14)

Calculating the minimum value of Eq. (14) by equating its derivative with respect to \( s \) to 0, we find that the minimum occurs at \( s = \sqrt{\Delta x/\Delta \sigma_x} \).

(15)

With this value of \( s \), we always have \( N' = N \), regardless of \( ABCD \). In other words, the choice of \( s \) allows the output signal to be fully represented on the spherical reference surface, with the same number of samples as the input signal. In contrast, since \( N' \) is equal to \( N \) only for very special values of \( ABCD \), the representation of the signal on the planar output plane would require a greater number of samples than information theoretically necessary. Thus, the use of spherical reference surfaces allows non-redundant representation of signals with the minimum possible number of samples. Since this is possible if and only if the scale parameter \( s = \sqrt{\Delta x/\Delta \sigma_x} \), we refer to this value of \( s \) as the natural scale parameter. What is special about this value of \( s \)? It makes the spatial extent \( \Delta x/s \) and the frequency extent \( s \Delta \sigma_x \) in dimensionless phase-space both equal to \( \sqrt{\Delta x} \Delta \sigma_x = \sqrt{N} \). For this value of \( s \), the ellipse is reduced to a circle with diameter \( \sqrt{N} \). A rotated circle is still the same circle. So the FRT operation, whose effect in phase space is to rotate, does not change this circular region in any way. The consequence of this is that in dimensional coordinates, the output space extent is nothing more complicated than the input space extent multiplied with \( M \), and the output frequency extent is nothing more than the input frequency extent divided by \( M \). This leaves the space–bandwidth product unchanged.

These observations can also be seen from Fig. 1(c), which shows the space–frequency region-of-confinement of the signal on the spherical reference surface. With the natural scale parameter, not only are the area of the ellipses in Fig. 1(a) and (c) (and the area of their bounding rectangles) equal, but the space–bandwidth product and thus number of samples for the region in Fig. 1(c) is equal to that in Fig. 1(a). This becomes possible because the natural scale parameter equates the vertical and horizontal extents of the ellipse in dimensionless coordinates, reducing it to a circle. Consequently, no matter how much the space–frequency region is rotated by the FRT operation, the circle remains a circle and upon going back to dimensional coordinates we obtain an ellipse with axes aligned with the space and frequency axes as in Fig. 1(c). In contrast, the sheared space–frequency region in Fig. 1(b), although having an area equal to the area of the region in Fig. 1(a), exhibits a space–bandwidth product which is larger, requiring a greater number of samples. In other words, shearing leads to a space–bandwidth product and number of samples needlessly greater than the true number of degrees of freedom of the signal (given by the area of the region, which is invariant). Use of the natural scale parameter on the appropriate spherical reference surface avoids shearing and thus keeps the number of samples equal to the true number of degrees of freedom of the signal.

The optical field on the output spherical reference surface is simply related to the field on the input plane through a magnified FRT. If we choose the natural scale parameter, the magnification has a simple effect. It increases the spacing of the samples by a factor of \( M \) from \( \Delta \sigma_x \) to \( M \Delta \sigma_x \) from the values. We refer to these spherical reference surface samples as the natural sampling grid. Based on the Nyquist–Shannon theorem, these \( N \) samples allow interpolation of the continuous transverse optical field at any point along the QPS. They contain precisely the information we need to reconstruct the continuous signal without any redundancy.

The discrete Fourier transform begins with the samples of a function and delivers the samples of its common Fourier transform, to a good approximation. The discrete FRT does the same for the FRT [38–43]. It follows that if we are given the values of the samples of the input field, the discrete FRT can be used to approximately compute the values of the field on the natural sampling grid. Based on this we can state that, calculation of the field at any position in a quadratic-phase system requires nothing more than discrete fractional Fourier transformation.

Our approach allows us to work with the smallest number of samples \( N \) to represent the signal. Computation of the output samples from the input samples is fast; it takes \( \sim N \) log \( N \) time [38]. Moreover, computational accuracy arising from the use of the discrete FRT is not different than that arising from the use of the FFT to compute the common Fourier transform [31]. To summarize, we have shown that we can represent and compute the output field by using an identical number of samples as required for the input.

This is accomplished through a particular scaling strategy and by choosing to observe the output on spherical reference surfaces. Had we not made these choices, the unitarity and space–frequency-area (and number of degrees of freedom) preserving property of QPSs would not suffice to avoid information loss with this number of samples.

4. Longitudinal sampling spacing

Until now we concentrated on transverse sample spacing. What about the longitudinal spacing between reference surfaces? For simplicity, in this derivation we will use the QPS representation with the parameters \( \alpha, \beta, \gamma \) instead of the \( ABCD \) parameters. Consider the integral in Eq. (3). We can truncate the
infinite limits to \(\pm \Delta x/2\) on both sides, because the signal is assumed to have an extent of \(\Delta x\) around zero. We will see that depending on the parameters and integrand structure, the integral can be limited even more tightly. The instantaneous frequency of the chirp function inside the integral is found by differentiating its phase and multiplying by the chirp function inside the integral is found by differentiating its parameter \(\gamma\), resulting in the change of \(\Delta f = \Delta x/2\). This yields
\[
\Delta f = \frac{\pi \gamma}{\Delta x/2}.
\]

The non-uniform grid of samples points we have proposed has two constituents: the curved reference surfaces and the separation of samples on them. The parameters \(x, M, R\) define the curved surface structure, which mirrors the physical structure of the QPS. The parameters \(\Delta x\) and \(\Delta z\) define \(s\) and \(\delta a\), which mirrors the information content of the set of signals. The special value 
\[
M = \frac{s^2}{R^2} + 1
\]

obtained from Eq. (7), and with
\[
M = \frac{s^2}{R^2} + 1
\]

obtained from Eq. (7), we eventually obtain
\[
\delta a = \frac{16}{\pi \Delta x \Delta z} = \frac{16}{\pi N}
\]

Note that while \(\delta a\) depends on \(z\), the value of \(\delta a\) does not depend on \(z\) nor any of the system parameters. It is a constant for given signal space–frequency extents. Thus the spherical reference surfaces should be spaced equally with respect to the FRT order, rather than with respect to the distance. This is a generalization to QPSs, of the result we derived for Fresnel diffraction in [44], where we refer the reader for further discussion.

5. Discussion and conclusion

The non-uniform grid of samples points we have proposed has two constituents: the curved reference surfaces and the separation of samples on them. The parameters \(x, M, R\) define the curved surface structure, which mirrors the physical structure of the QPS. The parameters \(\Delta x\) and \(\Delta z\) define \(s\) and \(\delta a\), which mirrors the information content of the set of signals. The special value
\[
M = \frac{s^2}{R^2} + 1
\]

obtained from Eq. (7), and with
\[
M = \frac{s^2}{R^2} + 1
\]

obtained from Eq. (7), we eventually obtain
\[
\delta a = \frac{16}{\pi \Delta x \Delta z} = \frac{16}{\pi N}
\]

Notice that while \(\delta a\) depends on \(z\), the value of \(\delta a\) does not depend on \(z\) nor any of the system parameters. It is a constant for given signal space–frequency extents. Thus the spherical reference surfaces should be spaced equally with respect to the FRT order, rather than with respect to the distance. This is a generalization to QPSs, of the result we derived for Fresnel diffraction in [44], where we refer the reader for further discussion.

References