Space–bandwidth ratio as a means of choosing between Fresnel and other linear canonical transform algorithms

John J. Healy* and John T. Sheridan

1. INTRODUCTION

The space–bandwidth product (SBP) is well established as a useful measure either of an optical system or a wave field [1]. The SBP of a signal, or of a set of signals, is defined as the product of the width of the region in space and the range of spatial frequencies within which the signal is nonzero; hence, it is dimensionless. (Note that, for convenience, we consider one-dimensional signals throughout this paper, but extension to higher dimensions is trivial.) Lohmann et al. pointed out that the SBP indicates the number of samples necessary to define a signal and described a similar definition of the SBP of a system as the number of degrees of freedom the system can handle [1]. The interpretation of the SBP as the number of samples necessary to define a signal has proven useful in numerical simulation of linear canonical transforms (LCTs) [2]. However, as a single number, the SBP is limited in what it can tell us about a signal. Let us illustrate its weakness: consider an arbitrary wave field. The SBP of this function is invariant under magnification, and yet numerical simulation of the propagation of a magnified version of the wave field is highly sensitive to the magnification. Lohmann et al. recognized that the SBP must be complimented with other information, suggesting that the shape of the phase space diagram (PSD) is critical, which has been borne out in more recent numerical simulation work [2].

We propose that, when the initial assumptions about a wave field are that it has a particular width and bandwidth (i.e., when the initial PSD is rectangular), the space–bandwidth ratio (SBR, the ratio of the width and bandwidth of the wave field) may be used to decide between the use of different algorithms in a very simple way. Note that a signal cannot have both finite width and finite bandwidth, so, in practice, one works with approximations, e.g., choosing the bandwidth to be a range of frequencies such that some arbitrarily large proportion of the signal’s power is contained in that range. (Note also that the initial PSD need not necessarily be rectangular, e.g., in [3], we found a skewed rectangle to be advantageous, and in [4], an ellipse is preferred as it is later transformed into a circle that simplifies a rotation in the argument.) Like the SBP, the SBR is a number. These two numbers, the SBP and the SBR, completely and uniquely define the PSD of the signal—a rectangle in the space–frequency plane, the width of which in space is the principle square root of the product of the SBP and the SBR, and the width of which in frequency is the principle square root of the SBP–SBR ratio. The SBP and SBR are particularly useful in practice because they are typically very accessible numbers theoretically and experimentally: for all discrete inputs (measurements), the input function consists of some number of samples, SBP, sampled with period \(\sqrt{\text{SBR/SP}}\). In this paper, we demonstrate how to use the SBR of a signal to choose between two commonly used Fresnel transform algorithms, the spectral method (SM) and the direct method (DM) [2,5], and a generalization of these algorithms to the LCT case. The DM and SM are described in Subsections 1.A and 1.B respectively. As we show later, the result for the Fresnel transform case is equivalent to a result of Mendlovic et al. [6], although they did not explicitly note the significance of the SBR. However, to our knowledge, the LCT case is entirely novel.

The remainder of this paper is structured as follows. In Subsection 1.A, we describe PSDs and define the SBP and the SBR. Subsection 1.B describes the DM algorithm for...
the numerical approximation of the Fresnel transform and its generalization for the LCT. Subsection 1.C describes the SM algorithm for the Fresnel transform and introduces a generalization SM algorithm for the LCT. We compare the performance of these two algorithms for the Fresnel transform in Section 2, demonstrating how the SBR may be used to determine which algorithm is more efficient and that this measure is equivalent to that of [6]. Section 3 generalizes the results of Section 2 for any LCT and provides an example of the use of such calculations.

A. PSDs and ABCD Matrices

PSDs are a simple representation of the approximate support of a signal in the space–spatial frequency plane [1,7]. Combined with the coordinate transforming effect of LCTs on phase space representations, they can be used to analyze the sampling requirements of numerical algorithms in a simple fashion [2].

In this paper, we discuss two numerical algorithms, both of which consist of stages of magnification, Fourier transforms, and chirp multiplication. Magnification merely has to be accounted for and can be essentially be thought of as a free operation, chirp multiplication is an $O(N)$ operation, and the Fourier transform can be implemented using a commonly available fast Fourier transform (FFT) package of order $O(N \log N)$. Such multistage algorithms can be analyzed as decompositions of the ABCD matrix of a system since multiple concatenated systems may be characterized by the product of the ABCD matrices of the component systems [2]. The ABCD matrix of a system characterizes that system and defines the parameters of the LCT that models that system.

We now define the matrices that are used in this paper. The following operations are characterized by the following matrices.

- Free space propagation by a distance $z$, the Fresnel transform

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix} = \begin{pmatrix} 1 & \frac{\lambda z}{\lambda z} \\ 0 & 1 \end{pmatrix}. \quad (1)$$

- Propagation through a thin lens (chirp multiplication) of focal length $f$,

$$\begin{pmatrix} 1 \\ -1/df \\ 0 \\ 1 \end{pmatrix}. \quad (2)$$

- Magnification:

$$\begin{pmatrix} M \\ 0 \\ 0 \\ 1/M \end{pmatrix}. \quad (3)$$

- Fourier transform:

$$\begin{pmatrix} 0 \\ 1 \\ -1/0 \\ 0 \end{pmatrix}. \quad (4)$$

B. DM Algorithm

In this Subsection, we describe the DM algorithm. For the Fresnel transform case, this is described by the following operations [2]:

$$\begin{pmatrix} 1 & \frac{\lambda z}{\lambda z} \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ \frac{\lambda z}{\lambda z} & 0 \end{pmatrix} \begin{pmatrix} 0 \\ \frac{\lambda z}{\lambda z} \end{pmatrix} \begin{pmatrix} 0 \\ \frac{\lambda z}{\lambda z} \end{pmatrix}. \quad (5)$$

A pair of chirp multiplications flank a scaled Fourier transform. It is straightforward to generalize this algorithm to the LCT [2], yielding

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ B/\lambda z & 1 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ -1/0 \\ -1/0 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \end{pmatrix}. \quad (6)$$

The chirp multiplications have different factors, and the scaling has changed, but the core of the algorithm remains a single Fourier transform. As the chirps are of order $N$, the overall complexity of the algorithm is dominated by the FFT. Throughout the remainder of this paper, “the DM algorithm” will refer to the generalized algorithm for the LCT, i.e., Eq. (6), except where noted.

We now explicitly number the steps of the DM algorithm in order to facilitate referring to them in later sections.

D1. Consider a function, $f(x)$. Sample the function with the appropriate period, $T$. Thus, we produce $f[n] = f(nT)$, $n + [N/2] \in \{0, 1, \ldots, N\}$, where the square brackets indicate the operation of rounding down to the nearest integer.

D2. Multiply the sampled function by a chirp function, yielding $f[n] \exp[j\pi(nT)^2A/B]$. 

D3. Obtain $G[n]$, the FFT of $f[n] \exp[j\pi(nT)^2A/B]$. 

D4. The final (discrete) result is given by $G[n] \exp[j\pi(nT)^2D/B]$, where $T_y$ is the output sampling period, $T_y = 1/TB$. The continuous result, the LCT of $f(x)$, may be found using an appropriate reconstruction filter.

C. SM Algorithm

In this Subsection, we describe the SM algorithm [2] and introduce a generalization of it for the LCT. For the case of the Fresnel transform, the SM is given by

$$\begin{pmatrix} 1 & \frac{\lambda z}{\lambda z} \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}. \quad (7)$$

The generalization of this is less obvious than it was for the DM, as these three operations are not sufficient to implement any LCT. We propose the following generalization, which uses an additional magnification and a chirp multiplication

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 0 & 1 \end{pmatrix}. \quad (8)$$

For the Fresnel case, $C = 0$ and $A = 1$. Hence, the two leftmost matrices (the last two operations) reduce to imaging with unit magnification and can be neglected. It is interesting to note that, in this formulation, the SM is equivalent to using the DM with suitably modified parameters on the Fourier transform of the wave field. (This is evident if, neglecting the rightmost matrix in Eq. (8), a Fourier transform, we compare Eq. (8) with Eq. (6).) In [3], we determined that the DM requires high sampling rates for short propagation distances even when Ding’s LCT sampling theorem [8–10] is used in preference to Shannon sampling, although savings are available for longer propagation distances. The reason for these disparate sampling rates is that, unless the sampling rate is sufficiently high so as to well separate them, the replicas created in phase space by sampling are made to have overlapping support along the frequency axis due to the initial chirp multiplication. An algorithm that first Fourier transforms the signal moves these replicas onto the space axis, where they
are quickly (and typically automatically) removed by truncation.

As before, we explicitly number the steps of the SM algorithm in order to facilitate reference in later sections.

S1. Sample the function, \( f(x) \), with the appropriate period, \( T \). Thus, we produce \( f(nT) \), as in step D1.

S2. Obtain \( F[n] \), the FFT of \( f[n] \).

S3. Multiply \( F[n] \) by a chirp function, \( g[n] = F[n]\exp[\cdot j\pi(nT_f)^2B/A] \), where \( T_f \) is the sampling period in the Fourier domain, \( T_f = 1/T \).

S4. Obtain \( G[n] \), the FFT of \( g[n] \).

S5. Multiply \( G[n] \) by a chirp function, yielding \( h[n] = G[n]\exp[j\pi(nT_z)^2C/A] \), where \( T_z \) is the sampling rate in the Fourier domain, \( T_z = TA \). The continuous result, the LCT of \( f(x) \), may be found using an appropriate reconstruction filter.

2. COMPARISON OF THE FRESNEL ALGORITHMS

In this Section, we determine the sampling rate required to calculate the Fresnel transform of a function with a given SBP and SBR for a range of propagation distances, choosing the more efficient of the two algorithms under consideration. The resulting sampling rates show the expected well-known behavior: the SM requires fewer samples for short propagation distances, but the two algorithms require the same number of samples at longer propagation distances, which amounts to an advantage for the DM, which only employs the FFT once. The question is then what is meant by "short" propagation distances. We determine this transition point in terms only of the SBR and the wavelength.

One of the advantages of both of these algorithms is that one need not consider the role of replicas because the sampling analysis of \([2]\) is sufficient for algorithms that consist only of chirp multiplications, Fourier transforms, and scaling, as explicitly shown for the DM in \([11]\). The equivalent analysis for the SM is straightforward. We now briefly describe the sampling analysis of \([2]\) for the sake of completeness. The initial signal is assumed to have a given extent and bandwidth, indicating a rectangular PSD. The coordinates of the four corners of this matrix are placed in a \( 2 \times 4 \) matrix called the corner coordinate matrix (CCM). The CCM is multiplied by the ABCD matrix of each operation of the algorithm in turn, giving the coordinates of the corners of the PSD at each stage of the algorithm. The smallest possible rectangle is fitted around the PSD at each stage, giving the intermediate sampling requirements at every stage of the algorithm. (We note that this requires finding the points that are furthest apart in the spatial direction and the points furthest apart in the spatial frequency direction, resulting in the max{} and absolute value operations appearing in the analyses in this paper.)

Consider a rectangular aperture \( 2 \text{ cm} \) wide illuminated by an on-axis plane wave of wavelength \( \lambda = 500 \text{ nm} \). Of the power of the wave field, 99.87% (chosen as an arbitrarily high proportion of the total) is contained within the bandwidth \( 50,000 \text{ m}^{-1} \), which we therefore take as the finite bandwidth of the function. In Table 1, we compare the number of samples required to calculate the Fresnel transform of this function for various distances using either the SM or DM. These have been obtained using the CCM method of analysis of \([2]\). What constitutes a short propagation distance in this case?

### Table 1. Number of Samples Required by the DM and the SM to Calculate the Fresnel Transform of a Rectangular Aperture of Width 0.02 m, Bandwidth 50000 m\(^{-1}\) for Various Propagation Distances for Light of Wavelength 500 nm

<table>
<thead>
<tr>
<th>Propagation Distance (m)</th>
<th>DM</th>
<th>SM</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.01</td>
<td>81000</td>
<td>1013</td>
</tr>
<tr>
<td>0.5</td>
<td>2500</td>
<td>1625</td>
</tr>
<tr>
<td>0.75</td>
<td>2067</td>
<td>1938</td>
</tr>
<tr>
<td>0.8 ((Z_r))</td>
<td>2000</td>
<td>2000</td>
</tr>
<tr>
<td>1</td>
<td>2000</td>
<td>2000</td>
</tr>
</tbody>
</table>

We repeat the analysis used to produce Table 1, but for arbitrary wavelength and propagation distance and write the initial CCM in terms of the SBP (0.02 \( \times 50,000 = 1000 \)) and the SBR (0.02/50,000 = \( 4 \times 10^{-7} \)) only. Without loss of generality, we assume that all of the variables are positive. For the SM, we obtain the following required number of samples:

\[
N_{\text{SM}}^{\text{Fresnel}} = \text{SBP}\left(1 + \frac{\lambda Z}{\text{SBR}}\right).
\]

For the DM, we obtain

\[
N_{\text{DM}}^{\text{Fresnel}} = \text{SBP}\left(\max\left\{\left(1 + \frac{\text{SBR}}{\lambda Z}\right), \left(1 + \frac{\lambda Z}{\text{SBR}}\right)\right\}\right).
\]

Hence, the number of samples required by the two methods differs only when

\[
z < Z_f = \frac{\text{SBR}}{\lambda}.
\]

In the example given, the number of samples required differs only when \( z < 0.8 \text{ m} \), which is consistent with the findings shown in Table 1. Recall that we constrained all of the parameters in this analysis to be positive.

Both the DM and SM require the use of an FFT and a chirp multiplication, whereas the remaining operation in the DM is a second chirp multiplication; for the SM, it is an FFT. As the FFT has computational complexity \( O(N \log N) \) compared with the chirp multiplication’s complexity \( O(N) \), the DM is the faster algorithm to compute for a given number of samples. However, as we have seen, the DM and SM can require different numbers of samples for a given calculation. Mendlovic et al. proposed that the choice between the use of the two algorithms for Fresnel transform calculations be made by means of a simple inequality \([5]\). If \( z > Z_r \), then the DM should be applied; otherwise, the DM requires more samples than the SM, and so the SM should be preferred. The inequality arose from a PSD-based analysis of the sampling requirements for the two algorithms, and, while the methodology is somewhat different from ours, comparison of their definition of \( Z_r \) (in their notation) shows that Eq. \( (11) \) is completely equivalent.

3. COMPARISON OF LCT ALGORITHMS

The LCT case has not previously been examined in the literature. We now present the results of the analogous analysis for the LCT case to that for the Fresnel case that produced Eqs. \( (9) \) and \( (10) \). The DM requires the following number of
samples:

\[ N_{\text{DM}}^{\text{LCT}} = \text{SBP} \left( \max \left\{ \left| 1 \pm \frac{A}{B} \frac{\text{SBR}}{C} \right| \left( A \sqrt{\text{SBR}} \pm B \sqrt[12]{\frac{1}{\text{SBR}}} \right) \right\} \right) \times \left( C \sqrt{\text{SBR}} \pm D \sqrt[12]{\frac{1}{\text{SBR}}} \right). \]  

(12)

Note that, for the Fresnel transform parameters, Eq. (12) reduces to Eq. (10). The SM yields

\[ N_{\text{SM}}^{\text{LCT}} = \text{SBP} \left( \max \left\{ \left| 1 \pm \frac{B}{A} \frac{\text{SBR}}{C} \right| \left( A \sqrt{\text{SBR}} \pm B \sqrt[12]{\frac{1}{\text{SBR}}} \right) \right\} \right) \times \left( C \sqrt{\text{SBR}} \pm D \sqrt[12]{\frac{1}{\text{SBR}}} \right). \]  

(13)

For the Fresnel transform case, Eq. (13) reduces to Eq. (9).

Unfortunately, Eqs. (12) and (13) depend on \( \text{sgn}(A/B) \) and \( \text{sgn}(C/D) \) because of the \( \text{max} \) operation. Thus, expanding the second terms in these equations fails to yield simple insights. Let us consider a specific example: consider a system with the parameters \( A = 0.6, B = 2, C = -0.17, \) and \( D = 1.1. \) Figure 1 contains plots for these ABCD parameters of \( N_{\text{DM}}^{\text{LCT}}/N \) and \( N_{\text{SM}}^{\text{LCT}}/N, \) the ratios of the number of samples required for the calculation of each algorithm to the number of samples required to represent the input, as functions of SBR. The curves cross when SBR = 3.33. We have indicated that \( Z_i \), the transition distance, is equivalent to Eq. (11). Consistent with the idea of a transition distance for a given SBR, we can speak of a transition SBR, \( SBR_t, \) for a given system (i.e., a given ABCD matrix). When \( SBR > SBR_t, \) the DM requires more samples than the SM, so the latter should be preferred. For \( SBR < SBR_t, \) the DM is preferable. Note that the form of these plots depends on the ABCD parameters. For some parameters, the two curves are identical, meaning that the SM is preferable in all cases.

In this Section, we have considered only two numerical algorithms for calculating the LCT. As stated, these were chosen partly because analyses of algorithms involving only chirps, magnifications, and Fourier transforms do not need to additionally account for replicas [3,11]. The two algorithms were also chosen as their special cases have been extensively analyzed in the Fresnel transform literature [2]. However, many other LCT algorithms are presented in the literature [2,4,10], and these can also be analyzed and compared in this fashion. In particular, analysis of the algorithm presented in [10] is straightforward because that algorithm does not depend on the decomposition of the ABCD matrix, and thus it requires the number of samples indicated by the common terms in Eqs. (12) and (13). In the example in Fig. 1, the relative increase in the number of samples for this algorithm lies along the same curve as the DM below the transition SBR, and the same curve as the SM above it (i.e., the lower curve for all SBRs). Thus, as it is a single \( O(N \log N) \) operation requiring the minimal number of samples for any calculation, the algorithm presented in [10] is theoretically preferable to both the SM and DM algorithms for all input wave fields. However, in practice, commonly available FFTs have received a great deal of optimization, and hence the SM and DM algorithms are faster in practice than the algorithm of [10].

4. CONCLUSION

The SBP and SBR are interesting at least partly because they are very accessible variables. Every sampled representation of a wave field for a numerical simulation have associated with it a sampling rate and must truncate the wave field (field of view). Engineers typically do not want to have to engage in involved analysis to determine the preferable algorithm: a rule of thumb or simple equation is more convenient and offers the possibility of transparent automation. Equations (11)–(13) provide such a rule.

Note that the transitional distances/SBRs described in this paper are not sharp cutoffs. The advantage of only requiring one FFT means that the DM is advantageous slightly earlier than indicated in the kinds of analyses we have carried out. However, quantifying the meaning of “earlier” is not trivial. Certainly, we can find curves that describe the total count of operations or just multiplications [12]. However, the actual run time of the calculations is specific to each combination of platform and implementation as well as to the optical system parameters. In situations where a large number of transformations of different inputs for the same ABCD system must be calculated [13], experimentation to determine the precise ranges where various algorithms are advantageous is clearly advisable.

The SBP has become a useful part of the optical engineer’s conceptual toolbox, a simple, powerful description of optical systems and signals. Whether the SBR can find applications outside the one suggested in this paper and become as powerful an idea as the SBP remains to be seen.

ACKNOWLEDGMENTS

J. J. Healy wishes to acknowledge the support of the Irish Research Council for Science, Engineering and Technology and the Optica Mathematica project DGAPA-UNAM IN-105008 and SEP-CONACYT 79899. J. T. Sheridan acknowledges the support of Enterprise Ireland and Science Foundation Ireland through the Research Innovation and Proof of Concept Funds and the Basic Research and Research Frontiers Programs and the Irish Research Council for Science, Engineering and Technology.

The authors would like express their gratitude to the reviewers, who were exceptionally thorough and greatly improved the clarity of the paper.
REFERENCES