Complex signal recovery from two fractional Fourier transform intensities: order and noise dependence

M. Günhan Ertosun *, Haluk Atlı, Haldun M. Ozaktas, Billur Barshan

Department of Electrical Engineering, Bilkent University, TR-06800 Bilkent, Ankara, Turkey

Received 27 March 2004; received in revised form 31 July 2004; accepted 6 September 2004

Abstract

The problem of recovering a complex signal from the magnitudes of two of its fractional Fourier transforms is addressed. This corresponds to phase retrieval from the transverse intensity profiles of an optical field at two arbitrary locations along the optical axis. The convergence of the iterative algorithm, the effects of noise or measurement errors, and their dependence on the fractional transform order are investigated. It is observed that in general, better results are obtained when the fractional transform order is close to unity and poorer results are obtained when the order is close to zero. It follows that to the extent that conditions allow, the fractional order between the two measurement planes should be chosen as close to unity (or other odd integer) as possible for best results.

© 2004 Elsevier B.V. All rights reserved.

PACS: 42.30.R; 42.30.K
Keywords: Fractional Fourier transform; Phase retrieval

1. Introduction

Phase retrieval from intensity information is a problem of great practical interest and has accordingly been extensively studied (for instance, see [1,2] and the references therein). Two variations of the problem are particularly common. In the first, only the magnitude of the Fourier transform of a function is known, but additional assumptions regarding the function, such as finite extent and non-negativity – are made. In the second, the magnitudes of both the function and its Fourier transform are known. Both problems are typically solved with iterative algorithms.

In this paper, we consider the generalization of the second of the above problems to the case where the magnitude (or intensity) of the signal is known at two arbitrary fractional Fourier domains (or in other words, we know the magnitude of the
fractional Fourier transform of the function at two orders). The $a$th order fractional Fourier domain is a generalization of the ordinary space and frequency domains. Just as the original function resides in the space domain and its Fourier transform resides in the frequency domain, the $a$th order fractional Fourier transform (FRT) of the function [3–7] resides in the $a$th order fractional Fourier domain [7–12]. The $a$th order FRT $f_a(u)$ of a function $f(u)$ is defined for $0 < |a| < 2$ as

$$f_a(u) = \int_{-\infty}^{\infty} \frac{\exp[-i\pi \text{sgn}(x)/4 - x/2]}{|\sin x|^{1/2}} \times \exp[i\pi (\cot x u^2 - 2 \csc x u' + \cot x u'^2)] f(u') du',$$

(1)

where $x = a\pi/2$. When $a = 0$, we have $f_0(u) = f(u)$ and when $a = \pm 2$, we have $f_\pm(u) = f(-u)$. When $a = 1$, we have $f_1(u) = F(u)$, the ordinary Fourier transform, and when $a = -1$, we have $f_{-1}(u) = F(-u)$, the ordinary inverse Fourier transform. The transform is additive in index: the $a_2$th transform of the $a_1$th transform of a function $f(u)$ is equal to the $a_2 + a_1$th transform and so forth. The FRT has a fast algorithm. Further properties and references are given in [13].

It has been shown that the propagation of optical waves can be characterized by the FRT, with the transverse amplitude of light going through FRTs of increasing order as light propagates along the optical axis. Mathematically, this result is expressed as a relationship between the FRT and the Fresnel integral, with the transform order being related to the distance of propagation [13,14]. Therefore, the problem of recovering a complex signal fully from its FRT magnitudes at two orders, can be used to solve the problem of recovering a complex field from two transverse intensity profiles at two arbitrary locations along the optical axis. In other words, if we cannot measure the phase at a certain plane, we can compensate by measuring the intensity at two planes.

Furthermore, optical systems involving arbitrary concatenations of thin lenses and sections of free space in the Fresnel approximation can also be modeled in terms of the FRT [13,14]. Such systems are known as quadratic-phase systems, ABCD optical systems, or other names. This means that the transverse amplitude of light at any arbitrary plane of such a system can be related to the transverse amplitude at any other plane through a FRT relationship. Therefore, the problem addressed in this paper can also be used for recovering the complex field from two transverse intensity profiles at two arbitrary locations in such a system. In other words, the problem we deal with does not require that the two planes be related through a Fourier transform or free-space propagation and is more general.

A fundamental algorithm used to recover phase from the magnitudes of a function and its Fourier transform is known as the Gerchberg–Saxton (GS) iterative algorithm [15]. Many refinements of this basic algorithm have been considered; for instance, see [1,2,16]. A variation of the algorithm to deal with arbitrary general linear systems (not necessarily unitary like the Fourier or Fractional Fourier transforms) has also been presented [17–19]. In this paper we will consider the fractional generalization of the two-magnitude phase retrieval problem and the GS algorithm in their most purest forms so as to reveal the effects of working with fractional domains as transparently as possible, and refrain from carrying over extensions or refinements proposed for the basic GS algorithm. We also do not make use of any additional a priori knowledge other than the magnitude at two orders (such as non-negativity and so forth).

2. Review

The use of the GS algorithm in conjunction with FRTs has been reported in a number of earlier papers. In [20,21] the authors deal with beam shaping problems: while they do not consider the problem of phase retrieval, these papers are nevertheless relevant in that they use the GS algorithm to find the required profile of a phase-only filter.

A number of works deal with the retrieval of phase from FRT magnitudes at all orders, rather than just two. Such methods, based on tomography in the space-frequency plane, are clearly inefficient when applied to fully coherent or deterministic fields, since they use a great deal more information than needed, and require a very
large number of measurements. However, these methods constitute a powerful means of recovering the phase in the case of partially coherent light \[22,23\] \[13, pp. 378–380\].

The problem of retrieving the phase from the magnitude of two FRTs with closely spaced orders has been considered in \[24,25\]. This problem physically corresponds to retrieval of phase from knowledge of the intensity of light at two closely spaced locations on the optical axis \[26–29\], and has recently been generalized/extended in \[30,31\]. The two close-ordered FRT magnitudes allow approximation of the angular derivative of the fractional Fourier magnitudes, which permits one to find the instantaneous frequency and the phase. This approach is a special case of that considered in the present paper since it works only when the orders are close, whereas we allow arbitrary orders. However, as we will see, our method is less satisfactory when the orders are close, so that in such cases, this approach may be preferred and thus considered to complement the approach here.

The problem of phase retrieval from two FRT magnitudes is similar to the problem of phase retrieval from two Fresnel transform magnitudes (for instance, see \[32–34\]). Formulating the problem in terms of FRTs is consistent with the description of optical propagation through ABCD optical systems as continuous fractional Fourier transformation and evolution of the light field through increasing FRT orders. In addition to being mathematically purer, it has several advantages. The FRT satisfies a more complete and elegant set of basic properties. Since it corresponds to a pure rotation in the space-frequency plane (rather than shearing like the Fresnel transform), it is geometrically and numerically better behaved. Reference \[33\] discusses various Fresnel-based approaches and their applicability in the near, intermediate, and far fields, noting that fewer works on phase retrieval in the intermediate region have been published. The problem of phase retrieval from the magnitude at only a single Fresnel domain has also been considered \[35,36\]. These works make additional assumptions about the unknown signal, such as finite extent and non-negativity.

Reference \[37\] deals with optical systems involving the FRT which deviate from unitarity (rather than the pure FRT itself, which is always unitary) and shows that the Yang–Gu algorithm gives better results than the GS algorithm for nonunitary systems. In \[38–40\], the problem is formulated directly in the discrete domain, allowing the use of a finite recursive algorithm. The authors note that this approach has the advantage of not being dependent on the choice of initial phase. This algorithm has also found application in image encryption \[41\].

In the present paper, we consider the phase retrieval from two fractional Fourier magnitudes problem in a continuous framework. We do not make use of any additional a priori knowledge or hidden assumptions, employing the GS algorithm in its most basic form. We investigate the convergence of the algorithm, paying special attention to the dependence of convergence on the fractional transform orders. We also discuss the sensitivity of the outcome of the algorithm on noise or measurement errors, and the dependence of this sensitivity on the fractional transform orders. We observe that in general, both in terms of convergence and sensitivity, better results are obtained when the difference between the two orders is close to unity and poorer results are obtained when it is close to zero.

3. Results

The GS algorithm employed can be summarized as follows. For simplicity, in our examples we assume the magnitudes are known at the 0th and ath orders. The problem where the magnitudes are known at two orders a1 and a2 is totally identical if we identify a = a2 − a1. We begin by initializing the unknown phase function of f(u) = f0(u) to some initial value, such as zero or a constant. Then we take the ath order FRT of f(u). We leave the calculated phase intact but replace the magnitude with the known magnitude of f(u). Then we take the −ath (inverse ath) order FRT of this function and again leave the calculated phase intact and replace the magnitude with the known magnitude of f(u). The iteration cycle is then repeated. Since functions differing by a constant phase would have the same fractional
Fourier magnitudes, recovery is possible up to a constant phase factor. In plotting the final recovery errors in the following examples, we have eliminated this constant phase.

Our first example is shown in Fig. 1. A very large percentage of the signal energy is contained in a space-frequency region of radius 4, so that in all domains we restrict our attention to the interval $[-4,4]$. The Nyquist theorem implies a sampling interval in each domain equal to the inverse of the extent of the signal in the orthogonal domain, which leads to a sampling interval of $1/8$ and a space-bandwidth product of 64 in all domains. Either the discrete FRT matrix defined in [42] or the algorithm described in [43] can be employed to compute the FRTs to a good approximation; the former approach was employed here. The GS algorithm uses the magnitude of the function displayed and that of its FRT as its input. The phase is initialized to zero. (Very similar results are obtained when the phase is initialized to other constant values.) The results are shown in Fig. 2, which shows the difference between two consecutive iterates $\phi_k(u), \phi_{k-1}(u)$, defined as

$$\sqrt{\frac{1}{8} \int_{-4}^{4} |\phi_k(u) - \phi_{k-1}(u)|^2 du}. \quad (2)$$

In this particular example, the convergence is smooth and this difference converges to very small values within about a 100 iterations. We observe that in general, convergence is faster when the order is close to unity and worse when it is close to zero. For instance, for the 100th iteration, the difference decreases with the FRT order as 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.8, 0.7, 0.9, 1.0, with only the 0.7th and 0.8th orders spoiling the monotonic dependence. This is understandable, since if the order was zero, the function and its FRT become identical and we do not have sufficient information to retrieve the phase. Table 1 shows the final error between the 10th/100th iterate and the original phase function $\phi(u)$, defined as

$$\sqrt{\frac{1}{8} \int_{-4}^{4} |\phi_k(u) - \phi(u)|^2 du}. \quad (3)$$

![Fig. 1. Magnitude (a) and phase (b) of first example function.](image-url)
We again observe that the error is smaller when the order is close to unity and the error is larger when the order is close to zero. In fact, apart from the 0.8th and 0.9th orders, the final error exhibits a monotonic dependence on order.

We now consider the effects of adding various amounts of zero-mean white Gaussian noise to the magnitudes of both the function and its FRT. The signal-to-noise ratio (SNR) is defined as

\[
\text{SNR} = \frac{\int_{-\frac{\pi}{4}}^{\frac{\pi}{4}} |f(u)|^2 \, du}{\int_{-\frac{\pi}{4}}^{\frac{\pi}{4}} \sigma^2 \, du},
\]

where \(\sigma^2\) is the variance of the zero-mean Gaussian random variable characterizing each sample of the noise process. Table 2 gives the final errors between the hundredth iterate and the original phase function for different amounts of added noise. As expected, the final error increases as the noise increases for any given order. The decrease in final error with decreasing noise is less for small orders and more for higher orders. We also observe that, as a general pattern, a given amount of noise degrades the result more when the order is closer to zero and less when the order is closer to unity. When SNR = 1, all the final errors are very large and do not depend much on the order. As the SNR increases, we begin to observe a decrease in the final error with increasing FRT order. This decrease with order is most pronounced for the largest SNR values. The errors in this table have been computed according to the modified error formula.

![Difference between two consecutive iterates for first example.](image)

Table 1

<table>
<thead>
<tr>
<th>Order</th>
<th>0.1</th>
<th>0.2</th>
<th>0.3</th>
<th>0.4</th>
<th>0.5</th>
<th>0.6</th>
<th>0.7</th>
<th>0.8</th>
<th>0.9</th>
<th>1.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>10 iterations</td>
<td>0.3209</td>
<td>0.2509</td>
<td>0.1778</td>
<td>0.1327</td>
<td>0.1017</td>
<td>0.0681</td>
<td>0.0464</td>
<td>0.0395</td>
<td>0.0447</td>
<td>0.0001</td>
</tr>
<tr>
<td>100 iterations</td>
<td>0.1893</td>
<td>0.0825</td>
<td>0.0470</td>
<td>0.0250</td>
<td>0.0199</td>
<td>0.0088</td>
<td>0.0004</td>
<td>0.0013</td>
<td>0.0003</td>
<td>0.0001</td>
</tr>
</tbody>
</table>
in order to exclude the regions where the magnitude of the original function is identically zero. When the magnitude is identically zero, we cannot expect to retrieve the phase, which is indeterminate. Due to this indeterminacy, the algorithm is very sensitive to even the smallest amount of noise and results in random-like and highly erroneous results outside the interval \([-2, 2]\). Including these in the final error results in meaningless results. Of course, in practice we may not know that the actual noiseless magnitude was exactly zero outside this interval, and formal or informal application of additional constraints or knowledge would be needed to decide that the random-like results obtained outside \([-2, 2]\) constitute an amplification of noise and not a reconstruction of the phase in that region.

Our second example is shown in Fig. 3. Again, a significant percentage of the signal energy is assumed to be constrained in the space-frequency region of radius 4, so we restrict ourselves to the interval \([-4, 4]\) and the space-bandwidth product is 64. The phase is again initialized to zero (with similar results for other constant values). The convergence is displayed in Fig. 4, leading to similar

Table 2
Final recovery error in the presence of noise for first example. The SNR values correspond to noise variances of 0.5065², 0.1602², 0.0506², 0.0160², 0.0051², respectively

<table>
<thead>
<tr>
<th>Order</th>
<th>0.1</th>
<th>0.2</th>
<th>0.3</th>
<th>0.4</th>
<th>0.5</th>
<th>0.6</th>
<th>0.7</th>
<th>0.8</th>
<th>0.9</th>
<th>1.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>SNR = 1</td>
<td>1.5920</td>
<td>1.6488</td>
<td>1.5392</td>
<td>1.3719</td>
<td>1.4465</td>
<td>1.2967</td>
<td>1.3014</td>
<td>1.0782</td>
<td>1.2846</td>
<td>1.3038</td>
</tr>
<tr>
<td>SNR = 10</td>
<td>1.0300</td>
<td>1.1169</td>
<td>0.8108</td>
<td>0.5813</td>
<td>0.7190</td>
<td>0.7487</td>
<td>0.4576</td>
<td>0.5578</td>
<td>0.4804</td>
<td>0.4181</td>
</tr>
<tr>
<td>SNR = 100</td>
<td>0.4315</td>
<td>0.3019</td>
<td>0.2982</td>
<td>0.2863</td>
<td>0.2041</td>
<td>0.2480</td>
<td>0.2297</td>
<td>0.1387</td>
<td>0.1150</td>
<td>0.1520</td>
</tr>
<tr>
<td>SNR = 1000</td>
<td>0.2822</td>
<td>0.1574</td>
<td>0.0983</td>
<td>0.0715</td>
<td>0.0804</td>
<td>0.0547</td>
<td>0.0846</td>
<td>0.0585</td>
<td>0.0514</td>
<td>0.1095</td>
</tr>
<tr>
<td>SNR = 10,000</td>
<td>0.2653</td>
<td>0.1118</td>
<td>0.0783</td>
<td>0.0333</td>
<td>0.0352</td>
<td>0.0207</td>
<td>0.0174</td>
<td>0.0228</td>
<td>0.0255</td>
<td>0.0173</td>
</tr>
</tbody>
</table>

\[
\sqrt{\frac{1}{4} \int_{-2}^{2} |\phi(u) - \tilde{\phi}(u)|^2 \, du},
\]  

Fig. 3. Magnitude (a) and phase (b) of second example function.
comments as in the previous example. At 500 iterations, the difference between consecutive iterates decreases with order as 0.2, 0.3, 0.4, 0.1, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0, with only the 0.1th order spoiling the monotonic dependence. Table 3 shows the final error between the thousandth/ten-thousandth iterate and the original phase function (convergence is slower in this example than in the previous one). Notice that for the 0.2nd, 0.4th, 0.5th, 0.6th orders, the error has actually increased in the ten thousandth iteration, illustrating the fact that the algorithm does not always exhibit monotonic convergence. Generally speaking, the error is lower for larger orders and higher for smaller orders, although the ordering is far from being strict.

Table 4 shows the final errors between the ten thousandth iterate and the original phase function for different amounts of added noise. As expected, in general the final error increases as the noise increases, but there are exceptions; depending on unpredictable effects of noise on convergence, occasionally a higher error is obtained although the noise is lower. Again, as a general trend, the dependence of the final error on noise is greater for higher orders. For a given amount of noise the error tends to decrease as the order becomes closer to unity and the dependence on order is stronger for higher SNR values. However, these observations are more in the nature of a general pattern and far from a strict dependence. Unlike

<table>
<thead>
<tr>
<th>Order</th>
<th>0.1</th>
<th>0.2</th>
<th>0.3</th>
<th>0.4</th>
<th>0.5</th>
<th>0.6</th>
<th>0.7</th>
<th>0.8</th>
<th>0.9</th>
<th>1.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>1000 iterations</td>
<td>0.2214</td>
<td>0.1761</td>
<td>0.1331</td>
<td>0.1792</td>
<td>0.1218</td>
<td>0.0457</td>
<td>0.0419</td>
<td>0.0199</td>
<td>0.0225</td>
<td>0.0362</td>
</tr>
<tr>
<td>10,000 iterations</td>
<td>0.1661</td>
<td>0.1971</td>
<td>0.0258</td>
<td>0.1902</td>
<td>0.1564</td>
<td>0.0495</td>
<td>0.0089</td>
<td>0.0174</td>
<td>0.0028</td>
<td>0.0362</td>
</tr>
</tbody>
</table>
the previous example, the errors here are calculated over the whole interval $[-4,4]$.

It should be noted that in the examples presented (as well as others we have considered), the difference between two consecutive iterates quickly reaches very small values. Examining the final errors in tables 1 and 3, we observe that in some cases, usually corresponding to the higher orders, the final error has been reduced to very small values as well. On the other hand for other cases, usually corresponding to the lower orders, the final error does not decrease much with further iterations. This corresponds to stagnation of the algorithm or slow progress in a "tunnel" [1, p. 304], phenomena which are familiar from the conventional GS algorithm. In general it may be said that the likelihood of stagnation is greater for smaller orders. In the conventional GS algorithm involving the ordinary Fourier transform, the error can only decrease (or stay the same) at each iteration [1,16]. This also holds true when the ordinary Fourier transform is replaced with the FRT; since the FRT is also a unitary transform, the same proof is valid here as well. However, it must be noted that we use a different definition of error in this paper.

In this paper, we have followed earlier work on the GS algorithm and the FRT in considering one-dimensional examples to discuss the dependence of final error on order and SNR in the simplest and most transparent manner possible. However, it should be noted that while the solution to the same problem in two dimensions is in general nearly unique, the one-dimensional problem does not have a unique solution [45]. In practice, the non-uniqueness intrinsic to the one-dimensional problem must be resolved through the use of additional assumptions or information. However, since we did not wish to complicate our presentation with the introduction of a variety of such additional assumptions, we based our conclusions on examples converging to the original function.

### 4. Discussion and conclusions

The major observations of this paper may be summarized as follows: Applying the GS algorithm between two functions related through the FRT rather than the ordinary Fourier transform in general leads to similar qualitative behavior. A general degradation as the fractional order is decreased from unity towards zero is observed. However this degradation does not suddenly occur when we depart from the unity order (ordinary) Fourier transform, but rather gradually as the order changes from unity to zero. There is no specific threshold beyond which the results suddenly become worse. The degradation with decreasing order has several faces. Convergence is less satisfactory and stagnation more common for smaller orders. Furthermore, larger final errors are generally obtained with smaller orders. The effect of noise is to equalize the dependence on order. When the noise is small, the final errors depend strongly on order and SNR in the simplest and most transparent manner possible. However, it should be noted that while the solution to the same problem in two dimensions is in general nearly unique, the one-dimensional problem does not have a unique solution [45]. In practice, the non-uniqueness intrinsic to the one-dimensional problem must be resolved through the use of additional assumptions or information. However, since we did not
decrease greater for higher values of SNR. Again, we note that the degree of conformity with these general observations varies from example to example; thus they should be viewed more in the nature of a general pattern and not as a strict dependence.

That the results generally, if not always, improve as the order gets closer to unity, and degrades as the order gets closer to zero is understandable, since if the order was zero, the function and its FRT become identical and we do not have sufficient information to retrieve the phase. Therefore, when the order is close to zero, it is natural to expect the problem to be more difficult. In a sense, it is best if the two known magnitudes are “orthogonal” to each other, in the sense of belonging to two literally orthogonal fractional Fourier domains in the space-frequency plane. When the orders are close to each other, this makes the problem more ill-posed. Knowledge of the magnitude at two orders close to each other is less desirable while knowledge of the magnitude at two orders separated by a number as close to unity (or other odd integer) as possible is more desirable. Distinct methods for retrieving the phase from two closely spaced orders are presented in [24,25]. We also note that although it does not share precisely the same concerns, reference [44] deals with the importance of phase and amplitude information in different fractional domains.

It follows from the discussion of the previous paragraph that if two intensity measurements are to be made in an optical system and it is desired to retrieve the phase from these measurements, it would be best to select the measurement planes such that they correspond to two FRTs separated by orders close or equal to unity; that is, to two “orthogonal” domains. This conclusion, which follows immediately from our FRT based formulation, might have been less evident in a Fresnel transform based approach.

It may also be interesting to investigate the dependence of the results on the fractional order through the notion of the “phase contrast transfer function” [46]. This concept relates the signal strength and thus SNR to the fractional order (or propagation distance) and therefore can be used to gain further insight into the dependence of the results on the order, particularly regarding the noise-independent and noise-dependent contributions to this dependence.

Acknowledgement

H.M. Ozaktas acknowledges partial support of the Turkish Academy of Sciences.

A preliminary version of this paper appeared in [47] (copyright IEEE).

References
