Exact solution for scalar diffraction between tilted and translated planes using impulse functions over a surface

Levent Onural

Department of Electrical and Electronics Engineering, Bilkent University,
TR-06800 Ankara, Turkey (onural@bilkent.edu.tr)

Received July 30, 2010; revised October 26, 2010; accepted December 2, 2010;
posted December 9, 2010 (Doc. ID 132567); published February 4, 2011

The diffraction relation between a plane and another plane that is both tilted and translated with respect to the first one is revisited. The derivation of the result becomes easier when the impulse function over a surface is used as a tool. Such an approach converts the original 2D problem to an intermediate 3D problem and thus allows utilization of easy-to-interpret Fourier transform properties due to rotation and translation. An exact solution for the scalar monochromatic propagating waves case when the propagation direction is restricted to be in the forward direction is presented. © 2011 Optical Society of America

OCIS codes: 050.1940, 050.1960, 070.7345.

1. INTRODUCTION

The problem of finding the scalar optical field over a plane when the location and properties of light sources and diffracting apertures are given has attracted the attention of scientists for centuries; the problem is well known in optics (see, for example [1], Chaps. 3 and 4, and [2], Chaps. 9 and 11). Fraunhofer, Fresnel, Rayleigh-Sommerfeld, and Kirchoff solutions are well known and applicable to many geometries and practical cases. The approximations related to these solutions, and the associated limitations, are also well studied and reported in the literature [1].

The approach in this paper is the plane wave decomposition approach; this approach has been used for solving various optics problems, including diffraction [1,2]. The method is elegant and exact for scalar waves; usually evanescent modes are not included, and thus only propagating wave components are considered. Furthermore, this method allows ease in utilization of various signal processing techniques and tools and therefore paves the way for the solution of more complicated diffraction problems. However, the author believes that the method is underutilized.

Generally speaking, it may not be possible to compute the field at every point in the 3D space (volume) from the observed field over a subset of this 3D volume. However, additional constraints imposed by a particular problem might allow full reconstruction of the 3D field from the field over a lower-dimensional manifold in 3D space. For example, knowing the field over a 2D plane and further imposing a monochromatic propagating wave constraint, together with restrictions on the direction of propagation, we can uniquely find the 3D field. Usually, what is sought is not the field over the entire space (the 3D field); instead a subset of it, like the field over a plane or a surface, is needed.

It is desirable to find analytic solutions to various diffraction problems; furthermore, the efficiency and speed of subsequent digital implementations are always an issue.

Solutions with desirable features for the two parallel plane case have been known for a long time [3–5]. As a summary,

$$\psi_{z=0}(x,y) = \mathcal{F}_{2D}^{-1} \left( \mathcal{F}_{2D} \psi_{z=0}(x,y) \right) e^{i \pi \sqrt{k_x^2 + k_y^2}} ,$$

(1)

where $\psi_{z=0}(x,y)$ is the known field (object) over the $z = 0$ plane, and $\psi_{z=z_0}(x,y)$ is the desired diffraction field over the $z = z_0$ plane; $k$ is the wavenumber $\frac{2 \pi}{\lambda}$ associated with the monochromatic wave whose wavelength is $\lambda$. $\mathcal{F}_{2D}$ represents the 2D Fourier transform from the $(x,y)$ domain to the $(k_x,k_y)$ domain as $\mathcal{F}_{2D} f(x,y) = \int_{-\infty}^{\infty} e^{-i (k_x x + k_y y)} dx dy$, and $\mathcal{F}_{2D}^{-1}$ is its inverse. A good overview of basics of diffraction formulation based on a signal processing terminology is also presented in [6]. For simplicity we will also use the notation $\psi_1(x,y)$ and $\psi_2(x,y)$ to represent $\psi_{z=0}(x,y)$ and $\psi_{z=z_0}(x,y)$, respectively.

A definition for the impulse functions over surfaces in 3D space is already proposed and it is shown that this signal processing tool is powerful in formulating, and solving, diffraction problems [7]. We will show in this paper that the same signal processing tool gives a complete and compact analytic solution also for the tilted plane case.

The formulation and efficient computation of diffraction between two tilted planes have attracted the attention of various researchers. Leseberg and Frère discussed the problem and provided analytic expressions and computational procedures under the Fresnel approximation [8,9], Tommasi and Bianco based their analysis to plane wave decomposition for the continuous case, and then discussed discretization issues and also presented numerical results [10,11]. An analysis based on the plane wave decomposition and an efficient discrete computational procedure to find the diffraction over a tilted and translated plane, using the fast fourier transform have been described [12]. The same problem is also examined in [13,14], and details of numerical issues and the consequences of implicit periodicity assumption associated with
the utilization of discrete fourier transform during the discrete computation are presented. The solution for the continuous case is given in [10–12] under the restriction that no plane wave component of the 3D volume diffraction field propagates backwards with respect to each one of the two planes. The same problem is further investigated in [15] and the shift in the 2D frequencies of the diffraction patterns over the tilted planes is demonstrated analytically and by 1D simulations. A thorough analysis, together with some sampling issues, are further discussed in [16]. The results are also applied to computation of diffraction patterns due to 3D objects composed of planar patches (triangular meshes) [17–20]. Such a superposition is a consequence of an implicit assumption that the surfaces are optical sources, and therefore, it is an approximation when fields, instead of sources, are superposed as in most holographic reconstruction applications. However, this is still a good approximation if the segments are joined rather with small angles and have larger sizes; the approximation vanishes and the approach becomes exact when the segments are coplanar [21]. An approach based on superposition of diffraction patterns due to triangular patches that make up a 3D object is also utilized in [6]. They also use the plane wave decomposition (angular spectrum) to deal with orientations and positions of patches.

The resultant solution provided in this paper is essentially the same as the solutions already given in the references mentioned above, as expected; however, the analysis approach used in this paper is different: instead of formulating the problem as a 2D problem (2D pattern on one plane and another 2D pattern on another plane), we chose to approach the problem as a 3D problem. To be more specific, we first write straightforward and compact analytical expressions for the 3D field generated from a 2D mask. The diffraction field over the tilted plane is then just an appropriate 2D cross section of the computed 3D field. The adopted approach of going through the full 3D field also paves the way to utilize well-known high-dimensional signal processing techniques; this in turn opens the door for elegant solutions to many related problems. We utilize the impulse functions defined over a surface in 3D space [7] to get an easy-to-understand and compact representation. Furthermore, the solution provided in this paper also includes the plane wave components which may propagate backwards with respect to one of the tilted planes.

2. PLANE WAVE DECOMPOSITION FORMULATION OF DIFFRACTION

It is shown in [7] that the 3D monochromatic field can be written as a superposition of plane waves as

\[ \psi(x, y, z) = \psi(x) = \int \delta(k)A(k)e^{ik\cdot x}dk. \]  

Here \( \delta(k)A(k)dk \) is the complex amplitude of the plane wave propagating along the \( k \) direction, where \( k \) is the 3D wave vector whose components \( k_x, k_y, k_z \) indicate the spatial frequency of the propagating wave along the \( x, y, \) and \( z \) axes, respectively. Therefore, \( \delta(k)A(k)e^{ik\cdot x}dk \) is the plane wave component of the 3D field along the \( k \) direction and defined for all \( k \in \mathbb{R}^3 \). The key difference of the superposition given in Eq. (2) compared to those available in the literature is the range of \( k \); here the integral runs over all \( k = (k_x, k_y, k_z) \) over the 3D space. In other words, the integral in Eq. (2) is a triple integral and all three integrals run from \( -\infty \) to \( \infty \). Therefore, \( \delta(k)A(k) \) is the angular spectrum of the 3D field \( \psi(x, y, z) \).

Here in the diffraction formulation, for the monochromatic light case, \( S \) is a sphere whose center is at the origin, and whose radius is \( k \). As expected, due to monochromaticity, the spectrum is impulsive over the Ewald sphere \( S \) [22]. Furthermore, if the propagation is restricted only for those waves propagating along the positive \( z \) direction, then \( S \) becomes the corresponding hemisphere, \( |k| = k, \) and \( k_z > 0 \). Full mathematical definition and the associated properties of impulse functions over lower-dimensional manifolds in space are given in [7]. The impulse functions over surfaces were used before to describe 3D wave fields [22]. However, their definition must be carefully made, and associated properties must be carefully examined to utilize them properly and in a useful manner.

Such impulse functions provide a useful tool to describe many physical and mathematical conditions in a concise manner and pave the way for elegant solutions to otherwise difficult problems. This paper is also an example for such a case. Equation (2) is an important intermediate step in solving diffraction problems: it carries the problem to a higher dimension (3D) but provides a neat Fourier transform relation. Therefore, it allows us to use well-known Fourier transform properties. Clearly, \( \delta(k)A(k) \) is the Fourier transform of a 3D field, within a constant multiplier. The 3D approach, with the aid of the impulse function, as given by Eq. (2), allows us to use the well-known rotation and translation property of the Fourier transform to easily solve this seemingly complicated problem, as shown in the next section.

3. ROTATION OF THE 3D FIELD

Let us define a 3D Cartesian coordinate system \( x \), and another Cartesian coordinate system \( x' \), where \( x' = Rx + b \). Here \( R \) is a rotation matrix [i.e., a matrix whose rows (or columns) are orthonormal and \( \det(R) = 1 \)], and \( b \) represents the translation. Therefore, \( R^{-1} = R^T \). We define a new 3D function \( \psi_{Rb}(x) \) which represents the field as seen by the new coordinates \( x' \) as

\[ \psi_{Rb}(x) = \psi(x') = \psi(Rx + b). \]  

It is easy to show that the Fourier transforms, \( \Psi_{Rb}(k) = F_3D(\psi_{Rb}(x)) \) and \( \Psi(k) = F_3D(\psi(x)) \), of these functions are related as

\[ \Psi_{Rb}(k) = \Psi(Rk)e^{i(Rk)b}. \]  

But from Eq. (2), we know that

\[ F_3D(\psi(x)) = \Psi(k) = 8\pi^3\delta(\mathbf{k})A(k). \]  

Therefore, we can immediately write that
Using the rotation property in [7], we know that $\delta_S(Rk) = \delta_{SR}(k)$, where $SR$ is the rotated version of $S$. ($S$ and $SR$ are related to each other such that if $k \in SR$ then $Rk \in S$.) So,

$$\Psi_{R,b}(k) = 8\pi^3 \delta_{SR}(k) A(Rk) e^{iRk \cdot b}.$$  \hfill (7)

4. FIELD RELATION BETWEEN TILTED AND TRANSLATED PLANES

In this section we will show that the tools and the results outlined in previous sections can be used to find the diffraction relation between two tilted and translated planes. We start with the 3D relations already obtained in Sec. 3. Knowing the 3D Fourier transform $\Psi_{R,b}(k)$ of $\psi_{R,b}(x)$ as given by Eq. (8), we can find $\psi_{R,b}(x)$ via an inverse Fourier transform as

$$\psi_{R,b}(x) = \int_k \delta_{SR}(k) A(Rk) e^{iRk \cdot b} e^{iRk \cdot x} dk.$$  \hfill (9)

and using the properties of the impulse functions over surfaces [7], we write,

$$\psi_{R,b}(x) = \int_k \frac{(A(Rk)) e^{iRk \cdot x}}{S_R} dk = \sum_i \int_k \frac{(A(Rk)) e^{iRk \cdot x}}{B_i} dk,$$

where $\frac{dS_R}{dkx,dky} = \frac{k}{\sqrt{k^2 - k_x^2 - k_y^2}}$ is the Jacobian due to change of variables. Derivations in the above equation all follow from the properties given in [7]. $B_i$’s are the orthogonal projections of segments of $SR$ onto the $(k_x, k_y)$ plane; see Fig. 1. The summation is needed to take care of multiple projections. Equations (9) and (10) provide the link between the adopted 3D approach in this paper and the conventional 2D superposition.

Since $S$ is the hemisphere whose pole is the vector $p = (0, 0, 1)^T$, $SR$ is the rotated hemisphere whose pole is $p' = R^{-1}p = R^Tp$. Projection of $S$ onto the $(k_x, k_y)$ plane is the complete disc $k_x^2 + k_y^2 \leq k$, and there are no multiple segments during this projection; see Fig. 1(a). However, the projection of $SR$ onto the $(k_x, k_y)$ plane is no longer a full disc, and there are two overlapping projections. Let us partition $SR$ into two nonoverlapping parts, $S_1$ and $S_2$, such that all points on $S_1$ have a positive $k_z$ component (and therefore all points on $S_2$ have a negative $k_z$); see Fig. 1(b). Let $B_1$ and $B_2$ be the projections of $S_1$ and $S_2$ onto the $(k_x, k_y)$ plane, respectively. Note that either $B_1 \subset B_2$ or $B_2 \subset B_1$ and therefore, $B$, which is the projection of overall $SR$ is equal to the larger of $B_1$ or $B_2$. Therefore, Eq. (10) becomes

$$\psi_{R,b}(x) = \int_{B_1} \frac{A(Rk) e^{iRk \cdot x}}{S_R} \frac{k}{\sqrt{k^2 - k_x^2 - k_y^2}} dkx dk_y$$

+ $\int_{B_2} \frac{A(Rk) e^{iRk \cdot x}}{S_R} \frac{k}{\sqrt{k^2 - k_x^2 - k_y^2}} dkx dk_y.$  \hfill (11)

(Note that $\sqrt{k^2 - k_x^2 - k_y^2} = |k_z|$.) We know how to find $A(k)$ from the 2D field over the $z = 0$ plane [7]:

$$4\pi^2 A(k) \frac{k}{|k_z|} = F_{2D}(\psi_0(x,y)) = \Psi_0(k_x, k_y),$$

where $k_z$ in the argument of $A(k)$ is $\sqrt{k^2 - k_x^2 - k_y^2}$. So we can find $A(k)$ in terms of the Fourier transform of the diffraction pattern over the $z = 0$ plane as

$$A(k) = A(k_x, k_y, \sqrt{k^2 - k_x^2 - k_y^2} = \frac{1}{4\pi^2} \frac{|k_z|}{k} \Psi_0(k_x, k_y).$$  \hfill (13)

Therefore,

$$A(Rk) = A(k_x', k_y', k_z') = \frac{1}{4\pi^2} \frac{|k_z'|}{k} \Psi_0(k_x', k_y').$$  \hfill (14)

(Note that $k' = Rk$; therefore, $k_x', k_y'$ and $k_z'$ are all functions of $k_x, k_y, k_z$ through this matrix relation.) Furthermore, $k_z = \sqrt{k^2 - k_x^2 - k_y^2}$ for the integral over $B_1$, and $k_z = -\sqrt{k^2 - k_x^2 - k_y^2}$ for the integral over $B_2$. So, Eq. (11) becomes

$$\psi_{R,b}(x) = \frac{1}{4\pi^2} \int_{B_1} \Psi_0(k_x', k_y') \frac{|k_z'|}{|k_z|} e^{iRk \cdot x} dkx dk_y$$

+ $\frac{1}{4\pi^2} \int_{B_2} \Psi_0(k_x', k_y') \frac{|k_z'|}{|k_z|} e^{iRk \cdot x} dkx dk_y.$  \hfill (15)
Therefore, the 2D field pattern \( \psi_f(x', y') \) over the \( z' = 0 \) (tilted) plane is

\[
\psi_f(x', y') = \psi_{Rb}(x)|_{z=0} = \frac{1}{4\pi^2} \int_{B_1} \psi_0(k_x', k_y') \left| \frac{k_z'}{|k_z|} \right| e^{ik_z'(R^7b)\theta}(k_x'x + k_y'y) dk_x' dk_y' + \frac{1}{4\pi^2} \int_{B_2} \psi_0(k_x', k_y') \left| \frac{k_z'}{|k_z|} \right| e^{ik_z'(R^7b)\theta}(k_x'x + k_y'y) dk_x' dk_y' = F^{-1}_{2D}\{U(k_x, k_y)\},
\]

where,

\[
U(k_x, k_y) \triangleq \left| \frac{k_x'}{|k_z|} \right| \psi_0(k_x', k_y')[I(B_1) + I(B_2)],
\]

\[
I(B) \triangleq \begin{cases} 
1 & \text{if } (k_x, k_y) \in B \\
0 & \text{else} 
\end{cases}.
\]

Here \( F^{-1}_{2D} \) is the 2D inverse Fourier transform from the \( (k_x, k_y) \) domain to the \( (x, y) \) domain.

A simple drawing given in Fig. 2 shows the relation between the propagation direction of a single propagating plane wave and its cross section over two tilted planes (a 2D propagation with 1D cross sections are shown in the figure for the sake of clarity). As indicated in the literature, and in this paper, the corresponding fields over the tilted planes due to a single propagating wave will have different frequencies. A weighted superposition of 3D plane waves over a continuum of propagation angles for the monochromatic case, in the form of integrals as presented, yields the 3D field; the corresponding superposition of 2D cross sections over the tilted planes, with additional amplitude modifications as a consequence of associated change of variables, gives the fields over the tilted planes. The propagation direction given in Fig. 2(b) yields the maximum possible frequency over the first line; this maximum frequency is equal to \( k \).

5. COMMENTS AND CONCLUSIONS

Equation (15), plus the rotation relation which links \( k_x' \) and \( k_y' \) with \( k_x \) and \( k_y \), and the monochromaticity constraint which makes \( k_z \) and \( k_z' \) functions of \( (k_x, k_y) \) and \( (k_x', k_y') \), respectively, form the complete solution to the problem of finding the 2D scalar diffraction field over a tilted plane from a scalar field over another plane. The reference plane (object plane) is the \( z = 0 \) plane which is represented by the two parameters \( (x, y) \), where as the tilted plane is given by the parametric equation which describes the 3D coordinates, \( x \) of the plane as \( x = (R^{-1} \mathbf{b}) + x' \mathbf{r}_1 + y' \mathbf{r}_2 \), where the scalar parameters \( (x', y') \in (-\infty, \infty) \), and the vectors \( \mathbf{r}_1 \) and \( \mathbf{r}_2 \) are the first two column vectors of \( R^{-1} = [\mathbf{r}_1 \ \mathbf{r}_2] \). If needed explicitly, the relationship between \( (k_x', k_y') \) and \( (k_x, k_y) \) is

\[
k_x' = k_x r_{11} + k_y r_{12} + \sqrt{k_x^2 - k_x^2 - k_y^2 r_{13}},
\]

\[
k_y' = k_x r_{21} + k_y r_{22} + \sqrt{k_x^2 - k_x^2 - k_y^2 r_{23}},
\]

\[
|k_z'| = \sqrt{k_x^2 - k_x^2 - k_y^2}.
\]

The solution is exact, in the sense that, if the scalar field over a plane (infinite extent) is known, then the exact scalar field over the tilted plane can be computed. However, it should be noted that not all patterns are admissible field patterns over the \( z = 0 \) plane; only those patterns which can be generated by monochromatic propagating light waves are
allowed. An equivalent statement is to say that the field patterns over the \( z = 0 \) plane are restricted to low-frequency-content spatial patterns whose bandwidth is restricted to a circular region whose radius is \( k = \frac{\lambda}{2\pi} \) in the \((k_x, k_y)\) domain.

The equations provide a simple recipe for the computation which involves one forward and one backward Fourier transform, together with computation of some frequency shifts as implied by the relation between \( k_x', k_y' \) and \( k_x, k_y \); the amplitude modifications as implied by Eqs. (16) and (17) are important, but associated computational burden is insignificant.

The term \( \frac{\lambda}{2\pi} \) also deserves some attention: both \( k_x' \) and \( k_y' \) are cosines of some associated angles. Therefore, as \( k_z \) gets closer to zero, the amplitude of \( U(x, y) \) tends to infinity. \( k_z = 0 \) corresponds to the highest 2D spatial frequencies of the diffraction pattern \( \psi_0(x, y) \), which are then related to the 3D propagating waves whose propagation direction is parallel to the \( z = 0 \) plane [the highest 2D spatial frequency is \( k \) with arbitrary \((x, y)\) orientation]. Therefore, although the solution is analytically correct and fully justifiable physically, there could be numerical problems around high frequencies of the 2D diffraction pattern. By the way, it is already mentioned that this term is related to the Jacobian due to change of variables as indicated in Eq. (10).

The term \[ I[B_1] + I[B_2] \] also deserves some comments. Two monochromatic propagating waves, with the same \( k_x, k_y \), but with different \( k_z \), where \( k_z = \pm \sqrt{k^2 - k_x^2 - k_y^2} \) or \( k_z = -\sqrt{k^2 - k_x^2 - k_y^2} \), would generate the same 2D pattern on the \( z = 0 \) plane; however, one of the waves propagates in the positive \( k_z \) direction whereas the other one along the negative \( k_z \) direction. We restrict the direction of propagation to be all in the positive direction with respect to the \( z = 0 \) plane at the beginning of the problem; obviously, this restriction reflects a choice which may be related to some practical conditions, but the problem can also be solved for other cases where this restriction is removed. However, even under this restriction, another plane which is tilted with respect to the original plane will have waves propagating towards the "front" or "back" of it. Therefore, a single 2D spatial frequency component, of the diffraction pattern over the tilted plane, might have contributions from both forward and backward waves. \( I[B_1] \) and \( I[B_2] \) indicate the contributions of the forward and backward waves, respectively. In case both of them are 1, the inverse solution is not unique. In other words, the problem is not reciprocal; we cannot solve the diffraction pattern over the \( z = 0 \) plane given the field over the tilted plane, unless the constraints related to direction of propagation are revisited and modified.

Usually the solutions given in the literature [10–12, 16] assume that there are no back-propagating fields with respect to either planes. Such restrictions may come up depending on the specific nature of the physical implementations and associated limitations. For example, if the second plane is an opaque planar screen which receives the diffracted field from the first plane, there will be surely no back-propagating waves with respect to such a plane. However, when the problem is posed as a 3D propagating wave field in 3D free space with just two hypothetical planes intersecting it (as done in this paper), there is no reason to impose propagation direction restriction based on orientations of both of the planes. The restriction is necessary with respect to only one plane to remove the ambiguity of the direction of propagation due to the diffraction pattern given on that plane; this in turn assures a unique 3D field. Additional restrictions, like no backward propagation with respect to either planes, further remove some of the plane wave components which would otherwise contribute to the richness and quality of the 3D field and its 2D cross sections. Such removals will result in a loss in the image quality. Higher incidence angles with respect to plane normals generate higher-frequency 2D components over the cross sections; when these components are removed, the resultant fringe patterns will inevitably be blurred. This is significant especially when there is a large angle between the tilted planes. Therefore, such restrictions might be reasonable only for small tilt angles between the planes and for those 2D fringe patterns over the object plane which are originally rather low resolution compared to the highest possible optical resolution of \( \frac{1}{2} \) cycles per unit length. But, actually, unless there are physical limitations as outlined above, there is no reason to exclude back-propagating components with respect to one of the tilted planes; the solution given in this paper includes those components, as well, and therefore, the associated restrictions are not needed.

The results obtained in this paper are essentially the same as those given in the literature [10–12, 16], especially when the "no back propagation" constraint with respect to either plane is imposed, as expected. The significance of this work is in the procedure to obtain those results: here we used a technique which (i) obtains a 3D diffraction field from the 2D diffraction pattern given over a plane Eq. (2) and (13), (ii) rotates and translates the field Eqs. (4) and (9), and (iii) finds the planar cross section of the field to get the result, i.e., the field over a tilted and translated 2D plane Eq. (16)–(18).

**Author’s Remark:** The main concepts of this paper were developed in 2002, and the paper was essentially written in 2002 and 2003. However, during the writing process, a need for a better definition of the impulse function over manifolds had arisen. Therefore, the publication was delayed until such work was completed. As the impulse-function-related concepts were developed, and a paper was finally published in that field [7], this paper was revised; however, the paper was not submitted for publication until 2010. It was briefly revised once more in 2010 to include citations to related publications that had appeared since 2003 together with related comments.

**ACKNOWLEDGMENTS**

The author thanks Erdem Şahin for plotting Figs. 1 and 2.

**REFERENCES**


