Journal of Modern Optics
Vol. 57, No. 1, 10 January 2010, 23–36

Taylor & Francis
Taylor & Francis Group

Mathematical moiré models and their limitations

Isaac Amidror* and Roger D. Hersch

Ecole Polytechnique Fédérale de Lausanne (EPFL), 1015 Lausanne, Switzerland

(*Received 25 September 2009; final version received 1 October 2009*)

In a previous recent paper we discussed the limitations of the Fourier-based model in the prediction of visible moiré effects. Here, we extend the discussion to other mathematical models that are also being used in the moiré theory: the classical indicial equations method for modelling the moiré fringes between periodic or curvilinear gratings, and two methods that are used to model different facets of the moiré effects between correlated aperiodic layers. We discuss the limitations of each model, explain their significance, and suggest possible ways to overcome them.

Keywords: moiré; line gratings; dot screens; periodic layers; aperiodic layers; superposition

1. Introduction

Moiré theory deals with the mathematical modelling of the phenomena that occur in the superposition of two or more structures (line gratings, dot screens, etc.), either periodic or not.

Although modern moiré theory is largely based on the Fourier approach, the Fourier-based model is not always the best tool for dealing with questions related to moiré effects. For example, the Fourier theory is very well suited to dealing with global or macrostructure effects in the layer superposition, but it is not adapted to the explanation of local, microstructure effects such as rosette shapes in the periodic case ([1], Chapter 8) or dot trajectories in the aperiodic case ([2], Chapter 4).[1] The detailed study of microstructures requires therefore, other mathematical tools.

Moreover, in many real-world applications the Fourier approach may prove to be impractical even for the study of macroscopic moiré effects, due to the complexity of the calculations involved. In such cases, moiré analysis or synthesis can be done using simplified models that operate directly in the image domain, such as the indicial equations method. Indeed, moiré theory is not limited to the Fourier-based model alone, and it offers several possible alternatives, some of which are complementary to the Fourier-based model, while others come to simplify it for practical use in applications.

However, when using a mathematical model to explain physical or experimental phenomena, it often happens that although the model being used is well adapted to the situation at hand, in some particular circumstances it gives unexpected results, or simply fails. In some cases this may indeed indicate that we have reached the limits of the present mathematical model, and that a new refinement of the model or even a new theoretical breakthrough may be needed. In other cases, the reason may be more prosaic, such as some misunderstood or disregarded conditions, an unadapted hypothesis, or simply an error.

In the following, we discuss this question in the context of the moiré theory. We describe points in which the mathematical models being used seem to deviate from our real observations, and we try to provide some plausible explanations. Issues that are related to the Fourier theory have already been treated separately in our previous paper [3], so in the present contribution we will focus on other models being used in the moiré theory. We discuss in Section 2 issues that are related to the indicial equations model for moiré effects between periodic or repetitive, curvilinear layers; then, in Section 3 we proceed to the fixed points model for Glass patterns between correlated aperiodic layers; and finally, in Section 4 we discuss issues related to the vector field model for the microstructures (dot trajectories) that are generated between correlated aperiodic layers.

It is interesting to note that the moiré theory is, indeed, a good test case for studying the limitations of a mathematical model, since in this case no approximations (such as least square fits or statistical considerations) are explicitly being used and, in principle, the model should be able to describe precisely the phenomena in question, without any approximation errors or other inherent error sources. Of course, the use of any mathematical model to describe the reality

*Corresponding author.

ISSN 0950–0340 print/ISSN 1362–3044 online

DOI: 10.1080/09500340903494629
http://www.informaworld.com

implicitly implies some assumptions, simplifications or approximations. For example, we may simplify things a lot by assuming that our structures in the superposed layers (gratings etc.) are perfect, or that they extend to infinity in all directions (in order not to be distracted by border considerations). But this 'idealisation' of the real world does not introduce into the model explicit error considerations or uncertainties as in models that are based on best fit or statistical approaches.

2. The indicial equations model

The indicial (or parametric) equations model (see, for example, [4], ([5], pp. 64–78) or ([6], pp. 14–91)) is probably the simplest and the most widely used method for predicting the geometric shape of moiré patterns in the superposition of curvilinear gratings. The origins of this model can be traced back to the beginning of the twentieth century with pioneering works such as [7,8]. This model, which involves only the image domain, is based on the curve equations of the original curvilinear gratings: if each of the original layers is regarded as an indexed family of lines (or curves), the moiré pattern that results from their interaction forms a new indexed family of lines (or curves), whose equations can be deduced from the equations of the original layers.

For example, assume that we are given a periodic grating of vertical lines with period T_1, and a second periodic line grating with period T_2 which is rotated by

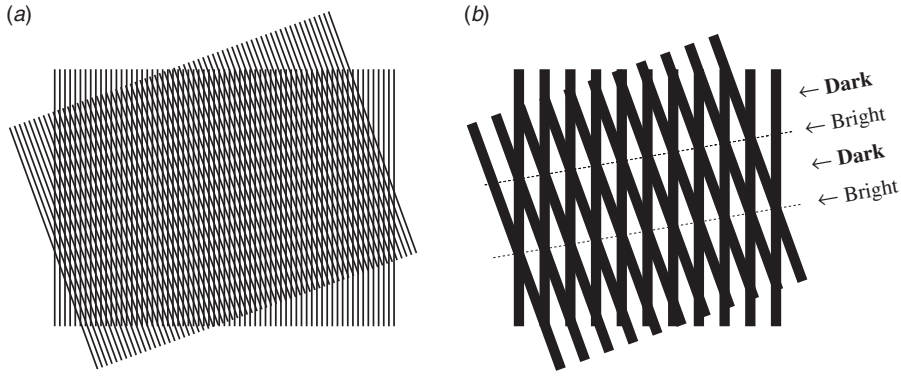

Figure 1. (*a*) Alternating dark and bright areas which form a $(1,-1)$-moiré effect in the superposition of two identical, mutually rotated line-gratings. (*b*) Enlarged view.

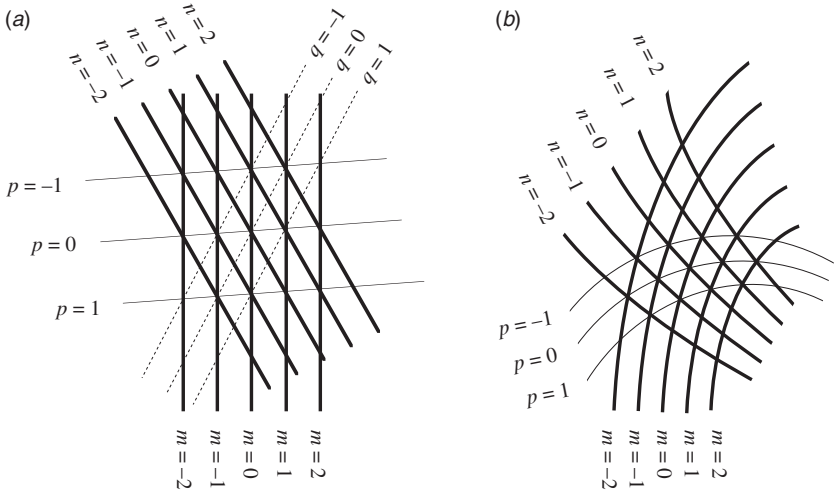

Figure 2. Schematic illustration of a $(1,-1)$-moiré (*a*) between two straight periodic gratings; and (*b*) between two curvilinear gratings. The two line families enumerated by the indices m and n, respectively, represent the centrelines of the superposed gratings, and the family of thin lines enumerated by the index p indicates the centrelines of the bright bands of the resulting $(1,-1)$-moiré. Similarly, the family of dotted lines enumerated by the index q indicates the centrelines of the bright bands of the $(2,-1)$-moiré. (For the sake of clarity the respective dotted lines have not been drawn in case (*b*).)

angle θ (as in Figures 1 and 2(*a*)). For the sake of simplicity we may assume that both gratings are centred on the origin. We consider each of the gratings as a family of lines, and we focus our attention on their centrelines, ignoring their real linewidths or their intensity profiles. If we enumerate the lines of the first grating by $m = \ldots, -2, -1, 0, 1, 2, \ldots$ then the equations of their centrelines in the x,y plane are given by:

$$x = mT_1, \quad m \in \mathbf{Z}. \tag{1}$$

Similarly, the equations of the centrelines of the rotated grating are:

$$x\cos\theta + y\sin\theta = nT_2, \quad n \in \mathbf{Z}. \tag{2}$$

As shown in Figure 1, moiré bands occur in a grating superposition since areas where black lines of the two gratings cross each other contain less black than areas where the grating lines fall between each other. Therefore, the bright bands of the most visible moiré run along the lines that connect closest crossing points in the superposition. This is illustrated in Figure 2 by the thin lines, which correspond here to the $(1, -1)$-moiré shown in Figure 1.[2] Note that, in general, the eye automatically selects as the most prominent moiré in the superposition the locus of intersection points in which the density of crossing points is the greatest; in the case of Figure 1 this corresponds to the $(1, -1)$-moiré, while in Figure 3 this corresponds to the $(2, -1)$-moiré.

Let us find the line equations of the most prominent moiré shown in Figure 2(*a*), i.e. the subtractive, $(1, -1)$-moiré.[3] In this case, the 0th line of the moiré line family (i.e. the centreline of the 0th bright band of the moiré) joins all the intersection points where $m - n = 0$, namely, the intersection points $(m, n) = \ldots, (-1, -1), (0, 0), (1, 1), \ldots$. But since the moiré bands are continuous, the 0th line of the moiré also contains all the intermediate points between these intersection points; clearly, it contains all the (x, y) points in the plane for which Equations (1) and (2) satisfy $m - n = 0$.

Similarly, the pth line of the $(1, -1)$-moiré consists of all the (x, y) points in the plane for which Equations (1) and (2) satisfy the condition:

$$m - n = p, \quad \text{where } p \in \mathbf{Z}. \tag{3}$$

In order to find the equation of the pth line of the moiré, i.e. the locus of all the (x, y) points that satisfy Equations (1)–(3), we have to solve these three simultaneous equations for x, y and p. This can be done by solving for m in Equation (1) and for n in Equation (2), and inserting these results into the indicial Equation (3). We obtain, therefore:

$$\frac{x}{T_1} - \frac{x\cos\theta + y\sin\theta}{T_2} = p$$

or after rearrangement:

$$x(T_2 - T_1\cos\theta) - yT_1\sin\theta = T_1T_2p.$$

This is the equation of the centreline of the pth bright moiré band. If we let the index p vary through all integers, $p = \ldots, -2, -1, 0, 1, 2, \ldots$, this equation represents the line family of the centrelines of the subtractive $(1, -1)$-moiré bands.[4]

More generally, the line equations of any (k_1, k_2)-moiré between two superposed gratings can be obtained in the same way, but this time instead of using condition (3) the more general indicial equation is used:

$$k_1m + k_2n = p, \quad p \in \mathbf{Z} \tag{4}$$

where k_1 and k_2 are constant integers, and m, n and p are the indexing parameters of the three line families (the two original gratings and the (k_1, k_2)-moiré bands).[5]

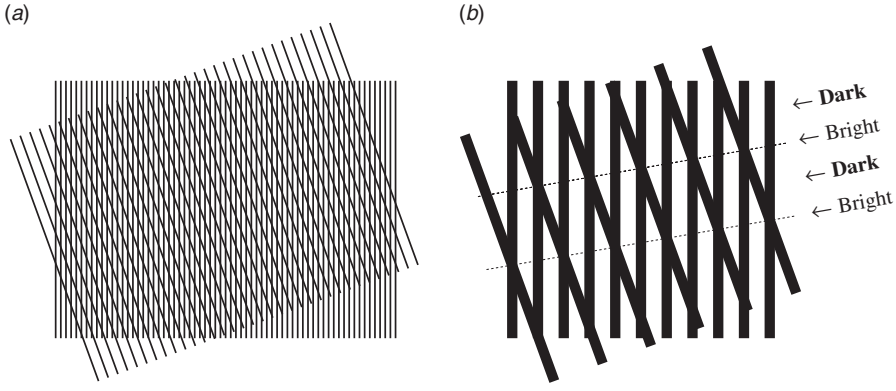

Figure 3. Same as Figure 1, but with the second grating having a double period. In this case the alternating dark and bright areas form a $(2, -1)$-moiré effect. This second order moiré has the same angle and period as the $(1, -1)$-moiré of Figure 1, and only its intensity is weaker.

In the most general case of a $(k_1, \ldots, k_s)$-moiré between s superposed gratings we will have s equations, one for each layer, plus the condition formulated by the general indicial equation:

$$k_1 n_1 + \cdots + k_s n_s = p, \quad p \in \mathbf{Z}. \tag{5}$$

The line equations of the $(k_1, \ldots, k_s)$-moiré bands can be found again, by replacing the indices $n_1, \ldots, n_s$ in Equation (5) with the expressions deduced from the line equations of the s original gratings, i.e. by solving the set of $s+1$ equations for x, y and p.

These considerations can also be used when the superposed layers are curvilinear, as illustrated in Figure 2(*b*). For example, assume we want to find the curve equations of the moiré between two identical shifted circular gratings (see Figure 4). In this case, derivations in the spectral domain become quite complicated (see Remark 10.7 in ([1], Sec. 10.7.6)). However, the derivation of the moiré shapes using the indicial equations method remains straightforward: the curve equations of the two horizontally shifted circular gratings in the x,y plane are given by

$$(x - x_0)^2 + y^2 = (mT)^2, \quad m = 1, 2, \ldots$$
$$(x + x_0)^2 + y^2 = (nT)^2, \quad n = 1, 2, \ldots$$

where x_0 and $-x_0$ are the respective horizontal shifts of the two circular gratings, and T is their radial period (i.e. the radial spacing between the centrelines of two consecutive circles in each circular grating). The equations of the curve families of the $(1, -1)$-moiré and of the (1,1)-moiré are obtained by solving the above equations for m and n and inserting the resulting expressions into the indicial equation $m \pm n = p$ in order to eliminate m and n. After some rearrangements one obtains ([6], p. 17):

$$\frac{x^2}{(pT/2)^2} \pm \frac{y^2}{x_0^2 - (pT/2)^2} = 1, \quad p = 1, 2 \ldots. \tag{6}$$

This means that the curves of the additive (1,1)-moiré form a family of ellipses, while the curves of the subtractive $(1, -1)$-moiré form a family of hyperbolas.

Further examples using the indicial equations method can be found, for instance, in the references mentioned in the beginning of the section. Interesting examples can be also found in [9] and [10], which analyse in depth the various (k_1, k_2)-moirés obtained between circular zone gratings or between a circular zone grating and a periodic straight grating; their indicial equations even include the phases of the different gratings and of the resulting moirés.

The indicial equations method can be summarised, therefore, as follows: if the centrelines of the s superposed curvilinear gratings are given by the curve families[6]

$$\begin{aligned} g_1(x, y) &= n_1, \quad n_1 \in \mathbf{Z} \\ &\vdots \\ g_s(x, y) &= n_s, \quad n_s \in \mathbf{Z} \end{aligned} \tag{7}$$

(where the line spacing T_i of each layer is already incorporated into $g_i(x, y)$) then the centrelines of the $(k_1, \ldots, k_s)$-moiré curves can be obtained from the indicial Equation (5) by eliminating the indices $n_1, \ldots, n_s$ using Equations (7). We obtain, therefore:

$$k_1 g_1(x, y) + \cdots + k_s g_s(x, y) = p, \quad p \in \mathbf{Z}. \tag{8}$$

This is the relationship between x and y that describes the curve family of the $(k_1, \ldots, k_s)$-moiré:

$$g_{k_1, \ldots, k_s}(x, y) = p, \quad p \in \mathbf{Z}.$$

We see therefore, that the geometric layout of the $(k_1, \ldots, k_s)$-moiré is determined by:[7]

$$g_{k_1, \ldots, k_s}(x, y) = k_1 g_1(x, y) + \cdots + k_s g_s(x, y). \tag{9}$$

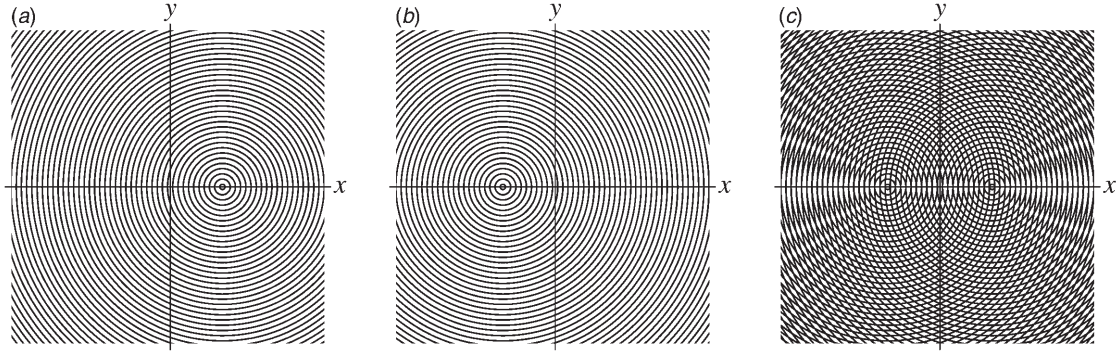

Figure 4. Two identical circular gratings, which have been horizontally shifted from the origin to the points (*a*) $x = x_0$ and (*b*) $x = -x_0$, and their superposition (*c*). The resulting hyperbolic fringes in the superposition (*c*) correspond to the subtractive $(1, -1)$-moiré, while the elliptical fringes correspond to the additive (1,1)-moiré.

For example, in the case of the $(1,-1)$-moiré between two line gratings the geometric layout of the moiré is determined by:

$$g_{1,-1}(x,y) = g_1(x,y) - g_2(x,y). \tag{10}$$

It is interesting to note that the indicial equations method can also be given a more visual interpretation, by regarding the indexed family of curves that describes a given layer as the level lines of a curved surface that are perpendicularly projected onto the x,y plane, as in a topographic map. According to this interpretation, if the ith indexed family of curves represents level lines of the surface $z = g_i(x,y)$, then the indexed curve family of the s-layers $(k_1,\ldots,k_s)$-moiré consists of the level lines of the surface:

$$z = k_1 g_1(x,y) + \cdots + k_s g_s(x,y). \tag{11}$$

In particular, the indexed curve family of the $(1,-1)$-moiré between two curvilinear gratings consists of the level lines of the difference surface $z = g_1(x,y) - g_2(x,y)$; the level line $z = p$ corresponds therefore to the projection onto the x,y plane of the intersection curve between the two surfaces $z = g_1(x,y)$ and $z = g_2(x,y) + p$, namely: $g_1(x,y) = g_2(x,y) + p$ (see [11], p. 25).

Finally, the indicial equations model can also be used in a different $(1,-1)$-moiré variant, in which one of the two superposed gratings, called the revealing layer, consists of thin line slits on a dark background, and samples the first line grating (see Figure 5). This moiré variant is very useful in applications of the moiré effect in the field of document security [12].

2.1. Limitations of the indicial equations model

In the indicial equations model the original gratings, as well as the resulting moiré bands in the superposition, are expressed by indexed families of curves that represent the centrelines of the curvilinear gratings and of the moiré bands. Therefore, unlike the Fourier-based model, the indicial equations model takes into account only the geometric layout of the centrelines of the curvilinear gratings and of the resulting moirés, but it totally ignores their intensity profiles (i.e. the real linewidths of the original gratings and the intensity or grey-level variations of the moiré). In fact, as shown in ([1], Section 11.2.2), the indicial equations that represent the curve families of the original layers and of the resulting moirés are already incorporated within their respective Fourier series representations. This shows that the indicial equations model is indeed fully encompassed by the Fourier-based model; in fact, Equation (9) corresponds to the second part of the main Fourier-based result, the fundamental moiré theorem for curvilinear gratings (see ([1], Section 10.9.1)). Clearly, since it only uses a part of the full information included in the Fourier expressions (the geometric layout of the layers, but not their intensity profiles), it is not surprising that the indicial equations model can only provide partial information about the moirés: it only gives the geometric layout of the resulting moiré, but not its intensity profile. Note, however, that in the particular moiré case mentioned above, where the revealing layer samples the other line grating, the intensity profile of the resulting moiré is essentially a larger version of the intensity profile of the grating being sampled (see Figure 5). This means that,

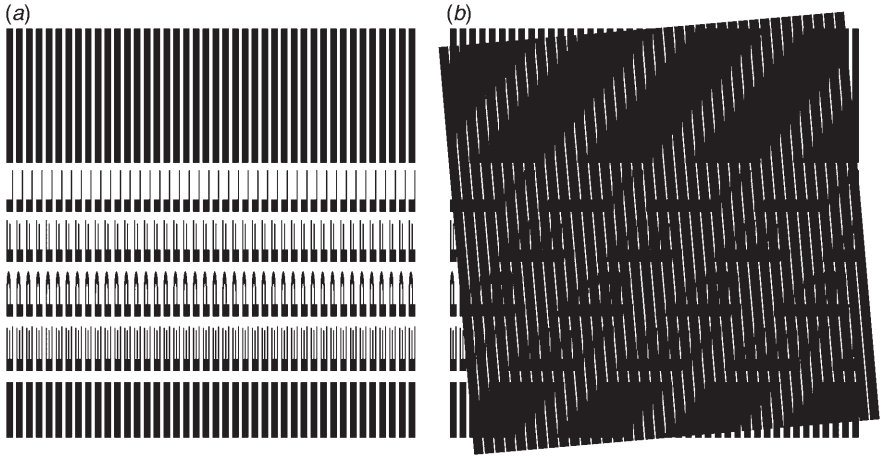

Figure 5. An example illustrating the particular case of $(1,-1)$-moiré in which one of the two superposed layers (called the revealing layer) consists of narrow slits on a black background and samples the other line grating (the base layer). The superposition (*b*) gives moiré bands whose intensity profile is essentially a larger version of the intensity profile of the base layer (*a*).

in this particular case, the intensity profile of the moiré *is* known when using the indicial equations model, although this knowledge is not directly provided by the indicial equations themselves but rather by Fourier-based considerations such as profile convolutions (see [1], Section C.14).

A second limitation of the indicial equations method resides in the ambiguity concerning the phase of the resulting moiré curves. As we have seen in the examples above and in Figure 2, the equations of the original gratings represent the centreline of the black curves, but the resulting equations of the moiré curves represent the centrelines of the bright moiré bands (which correspond to the intersection points between the black lines of the two original gratings). This kind of phase ambiguity is inherent in the indicial equations model. Although one can include in the indicial equations the phase of the original gratings and of the resulting moiré (for example by allowing the addition of a fractional part to each of the integer indices m, n, p, etc.), it turns out that unlike in the Fourier-based approach, the analysis of the phase in the indicial equations method is rather limited; for example, it fails to discriminate between black and white zones of a zone grating moiré (see [9], p. 40 and [10], p. 596).

This limitation is even more evident in the superposition of three or more gratings. In such cases, the geometric interpretation of intersection points between the different gratings becomes much more complex than in the case of two gratings shown in Figure 2, and the geometric connection with the phase of the resulting moiré bands is no longer obvious (see, for example, Figure 6 and compare with Figures 1 or 2).

A further limitation of the indicial equations model is that although it can provide the equation family for any desired $(k_1, \ldots, k_s)$-moiré in the given superposition (depending on the values of $k_1, \ldots, k_s$ that we insert in Equation (5) or (8)), it does not tell us which of them is indeed visible in the given superposition, or which of them is the most prominent. This depends, of course, on the grating periods, on their superposition angles, and also on the grating profiles.[8] This kind of information can be obtained in the Fourier-based model from the locations and the amplitudes of the elements (impulses etc.) in the spectrum of the superposition. Note, however, that this limitation of the indicial equations model can sometimes be overcome using the following rule of thumb, which may be helpful for the case of two gratings: in general, the human eye automatically selects as the most prominent moiré in the superposition the locus of intersection points in which the density of crossing points is the greatest. In the case of Figure 1 this corresponds to the $(1, -1)$-moiré; but in the case of Figure 3 this corresponds to the $(2, -1)$-moiré, so we should model here the visible moiré using,

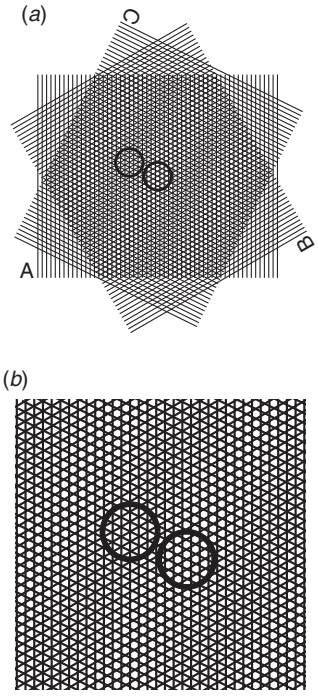

Figure 6. (*a*) A superposition of three line gratings that gives a visible moiré effect. (*b*) Enlarged view, showing in detail the microstructure of the superposition. Note that the geometric interpretation of the moiré bands in terms of intersection points between the original gratings is no longer obvious as in Figures 1–3.

in Equation (4), $k_1 = 2$ and $k_2 = -1$. Using instead $k_1 = 1$ and $k_2 = -1$ (namely, Equation (3)) would simply give the equations of the $(1, -1)$-moiré, which is not the most prominent moiré that we see in our superposition.

3. The fixed points model for Glass patterns

In this section we discuss the fixed points model that is used to explain the macroscopic phenomena (grey level variations) that occur in the superposition of aperiodic, correlated layers; the microstructure phenomena (dot trajectories) that appear in such superpositions will be considered later, in Section 4.

Suppose we are given an aperiodic layer such as a random dot screen, and that we superpose on top of it a second copy of the same structure that has undergone a small rotation, scaling, or both. As shown in Figure 7, we obtain in the superposition a new structure resembling a top-viewed funnel or a distant galaxy in the night sky, consisting of a usually brighter area that is surrounded by a microstructure of circular, radial or spiral dot trajectories. This phenomenon is known in literature as a *Glass pattern*, after Leon Glass who described it in the late 1960s [13,14].

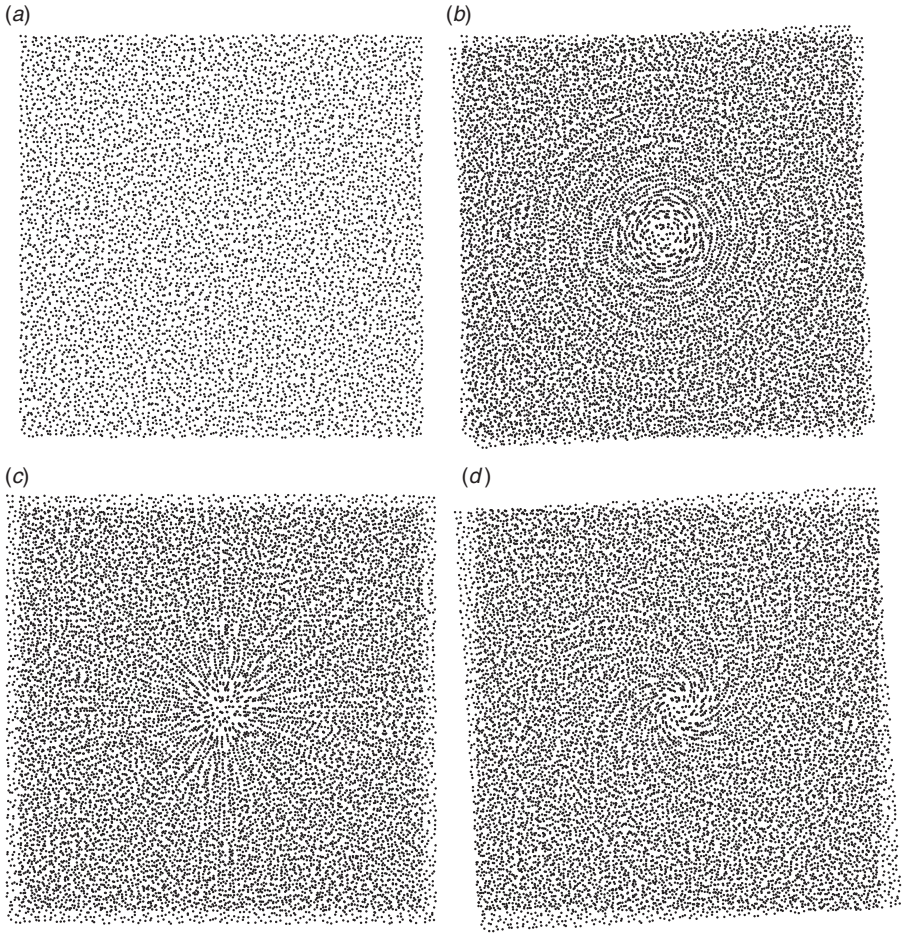

Figure 7. Glass patterns between aperiodic dot screens. (*a*) The original aperiodic dot screen being used for generating all the superpositions shown in this figure. (*b*) The superposition of two identical copies of aperiodic dot screen (*a*) with a small angle difference gives a Glass pattern about the centre of rotation. Note its typical microstructure consisting of concentric circular dot trajectories. (*c*) The superposition of two identical copies of aperiodic dot screen (*a*), one of which is slightly scaled up, gives a Glass pattern about the centre, which is surrounded by radial dot trajectories. (*d*) The superposition of two identical copies of aperiodic dot screen (*a*) one of which is slightly scaled up and rotated, gives a Glass pattern about the centre, which is surrounded by spiral dot trajectories.

Depending on the geometric transformations being applied to one or both of the originally identical aperiodic layers, we may obtain in the superposition Glass patterns of various different shapes (see Figure 8). Obviously, however, no Glass patterns can be expected in the superposition unless the superposed layers are sufficiently correlated.

Glass patterns are, in fact, the aperiodic counterpart of the moiré patterns, and indeed, as shown in ([2], Chapter 7), both of these phenomena are fully explained by the same Fourier-based model. But here, too, just as in the periodic case, there also exists a simplified image-domain model, which is partial to the Fourier-based model and is therefore more limited, but has the advantage of being much simpler to use. This model, known as the fixed-points model, has been reported for the first time in 1995 in the context of an application to stereo matching [15]. This model is based on the fact that the brighter zone in the centre of a Glass pattern is precisely the area where individual elements (dots) of the two superposed layers coincide (or almost coincide) on top of each other. This happens around the points (x, y) for which the transformations $\mathbf{g}_1(x, y)$ and $\mathbf{g}_2(x, y)$, which have been applied to the two originally identical layers, satisfy:[9]

$$\mathbf{g}_1(x, y) - \mathbf{g}_2(x, y) = (0, 0). \tag{12}$$

The points that satisfy this condition are precisely the mutual fixed points of the transformations $\mathbf{g}_1(x, y)$ and $\mathbf{g}_2(x, y)$.[10] Note that if only one of the two layers (say, the first one) has been transformed this condition becomes:

$$\mathbf{g}_1(x, y) - (x, y) = (0, 0). \tag{13}$$

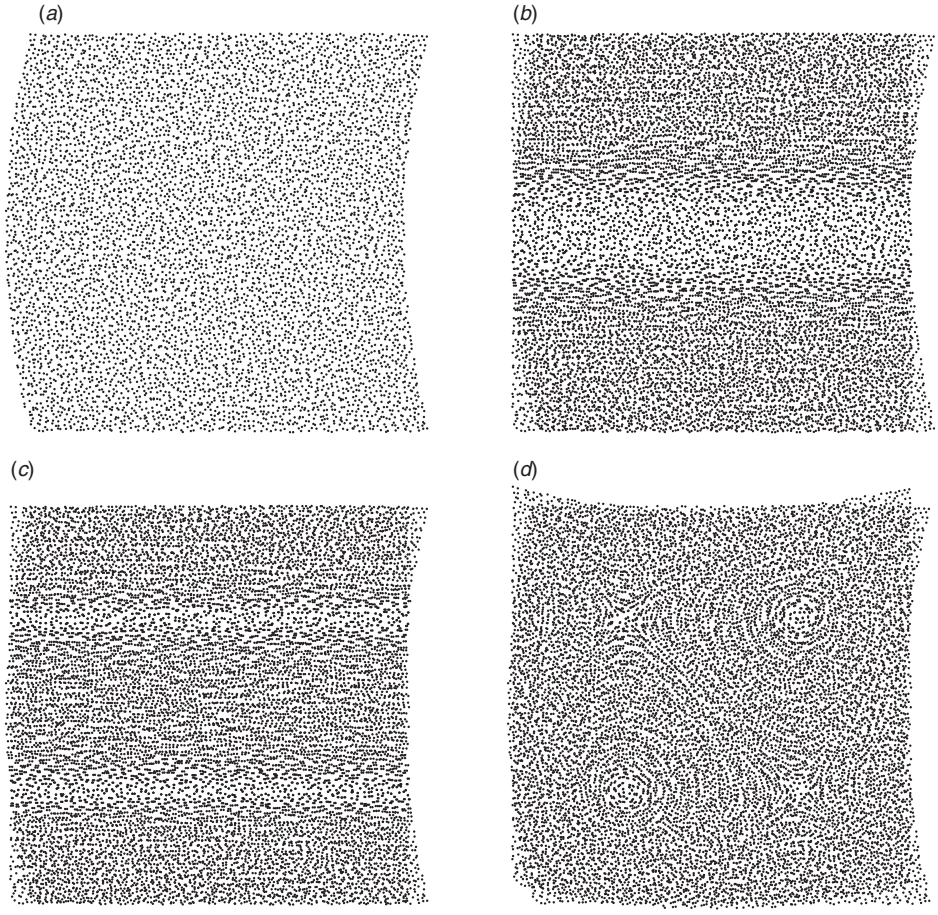

Figure 8. Glass patterns between aperiodic dot screens. (*a*) The aperiodic dot screen of Figure 7(*a*) after having undergone the parabolic transformation $\mathbf{g}(x, y) = (x - ay^2, y)$. (*b*) The superposition of two identical aperiodic dot screens, one of which has undergone the parabolic transformation $\mathbf{g}(x, y)$. Since this transformation does not involve layer shifts, the two layers clearly coincide along the x axis. (*c*) Same as in (*b*), but here the untransformed screen has been slightly shifted by $x_0 = T$ to the right. (*d*) An example with four mutual fixed points: the superposition of two identical aperiodic dot screens, one of which has undergone a vertical parabolic transformation plus a slight vertical shift of $y_0 = -T$, while the other has undergone a horizontal parabolic transformation plus a slight horizontal shift of $x_0 = -T$. Note the two circular and the two hyperbolic Glass patterns, which are clearly visible about the fixed points in the superposition.

The points that satisfy this condition are the fixed points of the transformation $\mathbf{g}_1(x, y)$.

Thus, in order to find the location of the Glass pattern, it is sufficient to solve Equations (12) or (13), depending on the case. And indeed, the fixed points model explains the various Glass patterns that appear in superpositions such as in Figures 7 and 8; the mathematical derivations of such cases can be found, for example, in ([2], Chapter 3). However, this model, too, suffers from some limitations, as will be shown below.

3.1. Limitations of the fixed points model

Just like the indicial equations model, the fixed points model operates in the image domain alone, and it only gives the 'skeleton' (centre point or centreline) of the resulting Glass patterns, i.e. the locus in the x,y plane (point, curve, etc.) where Equations (12) or (13) are satisfied. But it says nothing on the intensity profile of the Glass pattern, not even whether it is brighter or darker than the surrounding areas in the superposition.[11]

This is explained, once again, by the fact that the fixed points model is only partial to the Fourier-based model; in fact, Equation (12), which represents the skeleton of the resulting Glass pattern, follows from the second part of the main Fourier-based result, the fundamental moiré theorem for aperiodic screens (see ([2], p. 269)).[12] Hence, since it only uses a part of the full information included in the Fourier expressions (the geometric layout of the layers, but not their intensity profiles), it is not surprising that the fixed points method can only provide partial information about the resulting Glass patterns: it only gives the geometric layout of their skeleton, but not their intensity profile.

A second limitation of the fixed points model resides in the fact that although fixed points or mutual fixed points exist in a wide range of transformations, Glass patterns are only visible in the superposition when the transformations being applied are sufficiently weak. The reason is that Glass patterns are only visible between highly correlated layers, but when applying strong transformations the correlation between the layers becomes too low, and no Glass patterns can be seen in the superposition, even when mathematically fixed points do exist.

Note, however, that this limitation of the fixed points model can often be overcome by 'softening' the layer transformations being used. For example, instead of using transformations of the form $\mathbf{g}(x,y)=(x,y)+\mathbf{k}(x,y)$, one may try to use their softened versions $\mathbf{g}(x,y)=(x,y)+\varepsilon\mathbf{k}(x,y)$, where ε is a small positive fraction, as they are closer to the identity transformation $\mathbf{g}(x,y)=(x,y)$ and thus softer.

A third limitation of the fixed points model can be considered, in a way, as the converse of the previous limitation: it turns out that Glass patterns may still be visible in the superposition even if the transformations being used have no fixed points at all (namely, if there exist no points (x,y) that satisfy Equations (12) or (13)). This situation may occur if the transformations being used have *almost fixed points* rather than fixed points, i.e. points where $\mathbf{g}_1(x,y)$ is not fully identical to $\mathbf{g}_2(x,y)$ but only very close to it. If several such points exist, they form together an *almost fixed locus*. Although in almost fixed points there is no perfect coincidence between the two superposed layers, the elements of both layers around such points still fall very close to each other, while farther away the correlation gradually decreases. This generates around the almost fixed point (or locus) a visible Glass pattern whose centre is just slightly darker than in the case of perfect fixed point. A few such examples can be found in ([2], Section 3.7).

Mathematically speaking, although in this case no points (x,y) satisfy $\mathbf{g}_1(x,y)-\mathbf{g}_2(x,y)=(0,0)$, we can still extend the fixed points model to find those points for which the difference $\mathbf{k}(x,y)=\mathbf{g}_1(x,y)-\mathbf{g}_2(x,y)$ is almost zero, or, more precisely, the points (x,y) for which we obtain the minimum of this difference. Hence, if we denote the components of the difference $\mathbf{k}(x,y)$ by $k_1(x,y)$ and $k_2(x,y)$:

$$\mathbf{k}(x,y)=(k_1(x,y),\ k_2(x,y))$$

then what we are looking for is the loci of the minima of the function

$$k(x,y)=\sqrt{k_1(x,y)^2+k_2(x,y)^2} \tag{14}$$

which gives the length of the vector $\mathbf{k}(x,y)$ (or, in other words, the distance of $\mathbf{k}(x,y)$ from (0,0)). Alternatively, for the sake of simplicity, we may prefer to find the loci of the minima of its squared version, $n(x,y)$:

$$n(x,y)=k_1(x,y)^2+k_2(x,y)^2.$$

This kind of reasoning may help one overcome the third limitation of the fixed points model.

4. The vector field model for dot trajectories

The last model we discuss here concerns the dot trajectories, i.e. the microstructure dot arrangements that are usually visible around the Glass pattern in the superposition of aperiodic layers. These dot trajectories may have various different geometric shapes depending on the transformations that have been applied to the initially identical layers (see Figures 7 and 8).

As already mentioned, such microstructure effects cannot be taken care of by the Fourier-based model; and indeed, the model used to describe these phenomena, the vector field model [16], is not encompassed by the Fourier-based model, and the results it provides cannot be obtained by the Fourier approach. This model is therefore complementary to the Fourier-based model. Let us now briefly review the vector field model.

Suppose we are given an aperiodic layer such as a random dot screen, and that we superpose on top of it a second copy of the same structure, which has undergone a direct transformation $(x,y)\mapsto\bar{\mathbf{g}}(x,y)$.[13] Thus, the dot trajectories in our superposition consist of pairs of dots, which represent the location of a screen dot before and after the layer transformation $\bar{\mathbf{g}}(x,y)$ has been applied. These dot pairs can be represented, therefore, as a vector field, which assigns to each point $(x,y)\in\mathbf{R}^2$ a vector that connects (x,y) to its new location $\bar{\mathbf{g}}(x,y)\in\mathbf{R}^2$ under the transformation $\bar{\mathbf{g}}$.[14] It is important to note, however, that the vector field of the transformation $\bar{\mathbf{g}}(x,y)$ itself does not have this property; that is, the vector it assigns to (x,y) does not connect (x,y) to its destination $\bar{\mathbf{g}}(x,y)$, but rather to the point $(x,y)+\bar{\mathbf{g}}(x,y)$. For instance, if we consider the identity transformation $\bar{\mathbf{g}}(x,y)=(x,y)$, it is clear that in this case the vector attached to each point (x,y) is the vector (x,y) itself, which points, therefore, to the point $(2x,2y)$ and not to the destination point under $\bar{\mathbf{g}}$, which is the point (x,y).

Therefore, in order to obtain a vector field that correctly represents our dot trajectories, we must consider, instead of the transformation $\bar{\mathbf{g}}(x,y)$ itself (the transformation that has been applied to one of the

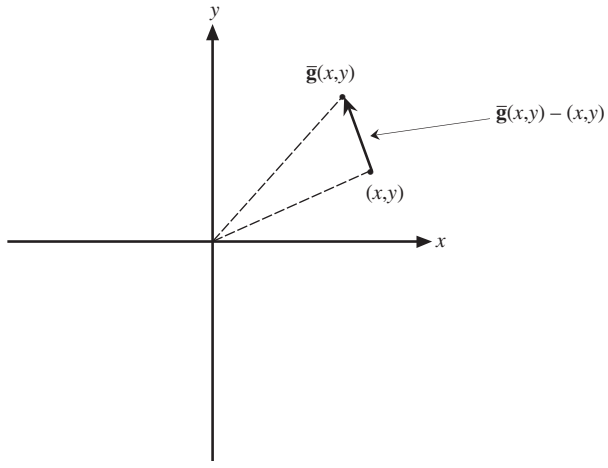

Figure 9. Point (x, y) in the x, y plane and its image $\bar{\mathbf{g}}(x, y)$ under the transformation $\bar{\mathbf{g}}$: $\mathbf{R}^2 \rightarrow \mathbf{R}^2$. The vector connecting the original point (x, y) to its destination $\bar{\mathbf{g}}(x, y)$ under transformation $\bar{\mathbf{g}}$ is given by $\bar{\mathbf{g}}(x, y) - (x, y)$.

superposed layers), the relative transformation between the two layers, which is given by:

$$\bar{\mathbf{h}}(x, y) = \bar{\mathbf{g}}(x, y) - (x, y). \tag{15}$$

If we draw the vector field representation of this transformation, we obtain exactly what we have desired: the vector field of $\bar{\mathbf{h}}(x, y)$ assigns to each point (x, y) the vector $\bar{\mathbf{g}}(x, y) - (x, y)$ which connects the original point (x, y) to its destination under the layer transformation $\bar{\mathbf{g}}$, the point $\bar{\mathbf{g}}(x, y)$ (see Figure 9). Indeed, as we can see in Figures 10(*a*)–(*c*), the vector fields obtained by Equation (15) agree with the dot trajectories in the corresponding superpositions (Figures 7(*b*)–(*d*)).

It should be noted, however, that the dot trajectories in our layer superposition can be represented by either of the vector fields $\bar{\mathbf{h}}(x, y)$ or $-\bar{\mathbf{h}}(x, y)$. This is because the two dots that compose each dot pair in the layer superposition are identical, so that the dot pairs (and hence the dot trajectories in the superposition) remain unchanged when we interchange the two layers. This means that the dot trajectories in the superposition do not show the direction (the positive or negative sense) of the difference vector.

Suppose now that we take one step further and allow both of the superposed layers to be transformed, one by a mapping $\bar{\mathbf{g}}_1(x, y)$ and the other by a mapping $\bar{\mathbf{g}}_2(x, y)$. As a straightforward generalisation of Equation (15), one would expect the dot trajectories in this case to be represented by the vector field of the relative transformation between the two layers, namely:

$$\bar{\mathbf{h}}(x, y) = \bar{\mathbf{g}}_1(x, y) - \bar{\mathbf{g}}_2(x, y). \tag{16}$$

Note, however, that in this case the dot pairs that make up the dot trajectories in the superposition no longer represent a dot's location before and after the layer transformation has been applied, but rather the new locations of the same original dot under the transformation $\bar{\mathbf{g}}_1(x, y)$ and under the transformation $\bar{\mathbf{g}}_2(x, y)$. Indeed, unlike the vector field (15), which perfectly corresponds to the dot trajectories that are obtained when one of the superposed layers is transformed, it turns out that in cases where both layers are transformed the vector field (16) only provides an approximation to the dot trajectories that are generated in the superposition. This is explained as follows.

Suppose we are given two identical aperiodic dot screens that are superposed on top of each other in full coincidence, dot on dot. Clearly, if we apply to both layers the same transformation $\bar{\mathbf{f}}(x, y)$, we still remain with two identical aperiodic screens that are superposed in full coincidence. Therefore, if we now apply transformations $\bar{\mathbf{g}}_1(x, y)$ and $\bar{\mathbf{g}}_2(x, y)$ to the two transformed layers, the resulting dot trajectories will have the same shapes as the dot trajectories that would be obtained by applying $\bar{\mathbf{g}}_1(x, y)$ and $\bar{\mathbf{g}}_2(x, y)$ to the original, untransformed layers. In other words, we have the following result.

Proposition: *The dot trajectories obtained by applying the transformations* $\bar{\mathbf{g}}_1(x, y)$ *and* $\bar{\mathbf{g}}_2(x, y)$ *to two identical aperiodic dot screens are equivalent to the dot trajectories that are obtained by applying to the same original dot screens the transformations* $\bar{\mathbf{g}}'_1(x, y) = \bar{\mathbf{g}}_1(\bar{\mathbf{f}}(x, y))$ *and* $\bar{\mathbf{g}}'_2(x, y) = \bar{\mathbf{g}}_2(\bar{\mathbf{f}}(x, y))$, *where* $\bar{\mathbf{f}}(x, y)$ *is any arbitrary transformation.*

Now, since $\bar{\mathbf{f}}(x, y)$ stands here for any arbitrary transformation, it is clear that this proposition also remains true in the particular case where $\bar{\mathbf{f}}$ is the inverse of the transformation $\bar{\mathbf{g}}_2(x, y)$, namely, $\bar{\mathbf{f}}(x, y) = \mathbf{g}_2(x, y)$. This means that the dot trajectories obtained by applying the transformations $\bar{\mathbf{g}}_1(x, y)$ and $\bar{\mathbf{g}}_2(x, y)$ to our original screens are equivalent to the dot trajectories that are obtained by applying to our original screens the transformations $\bar{\mathbf{g}}'_1(x, y) = \bar{\mathbf{g}}_1(\mathbf{g}_2(x, y))$ and $\bar{\mathbf{g}}'_2(x, y) = (x, y)$, where $\mathbf{g}_2$ is the inverse of the direct transformation $\bar{\mathbf{g}}_2$. Now, this last superposition has the particularity that only one of its two layers has been transformed. Therefore, we see by virtue of Equation (15) that the vector field that accurately represents the dot trajectories of this superposition, and hence also the dot trajectories obtained by applying the transformations $\bar{\mathbf{g}}_1(x, y)$ and $\bar{\mathbf{g}}_2(x, y)$ to our original screens, is given by:

$$\bar{\mathbf{h}}_1(x, y) = \bar{\mathbf{g}}_1(\mathbf{g}_2(x, y)) - (x, y) \tag{17}$$

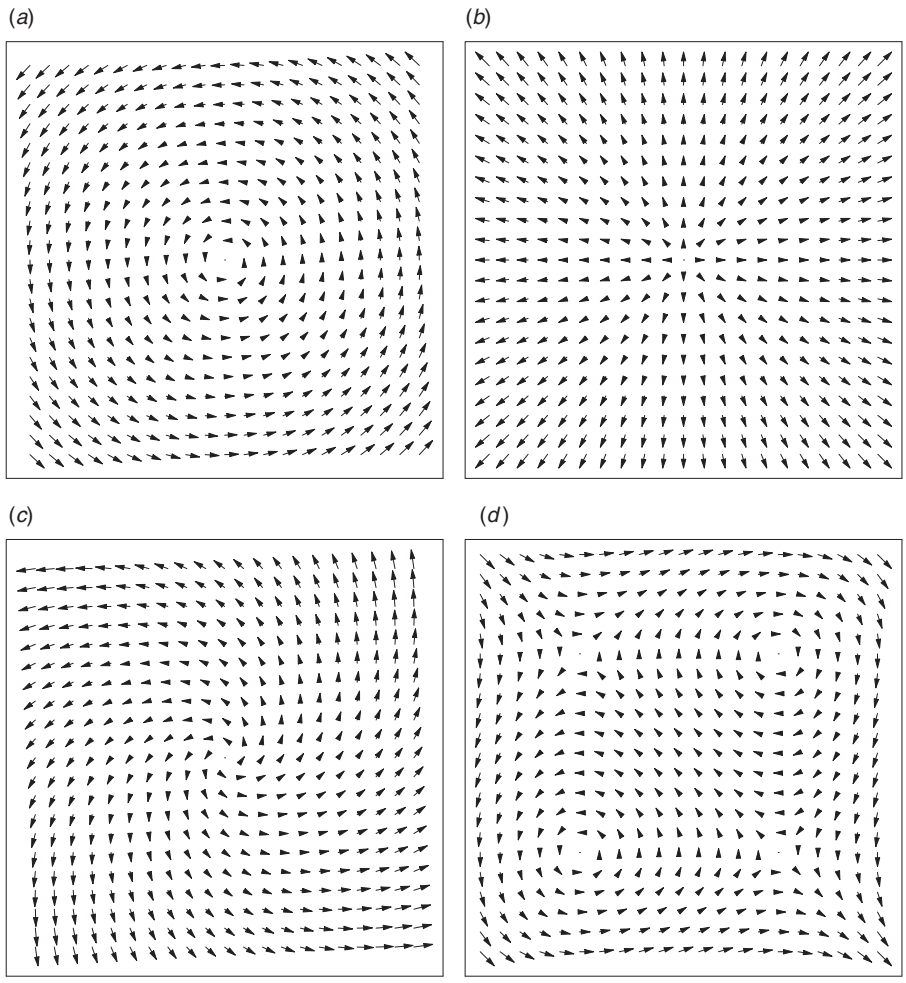

Figure 10. Vector field representation of the relative transformation $\bar{\mathbf{h}}(x,y) = \bar{\mathbf{g}}(x,y) - (x,y)$ where $\bar{\mathbf{g}}(x,y)$ (the transformation undergone by one of the layers) is: (*a*) a small rotation; (*b*) a small expansion; (*c*) both a small rotation and a small expansion. (*d*) Vector field representation of the relative transformation $\bar{\mathbf{h}}(x,y) = \bar{\mathbf{g}}_1(x,y) - \bar{\mathbf{g}}_2(x,y)$ where one layer has undergone a vertical parabolic transformation plus a slight vertical shift of $y_0 = -T$, and the other layer has undergone a horizontal parabolic transformation plus a slight horizontal shift of $x_0 = -T$. Compare with Figures 7(*b*)–(*d*) and 8(*d*).

(or equivalently, by $-\bar{\mathbf{h}}_1(x,y)$) rather than by Equation (16).

It should be noted, however, that although the vector field (16) only approximates our dot trajectories, it often turns out to be more practical to use than the accurate vector field (17). The reason is that the explicit form of vector field (17) may be quite complex, because it includes transformation compositions; furthermore, in many cases it may not even be available, since we do not always have the explicit forms of both of the direct and inverse transformations required. Therefore, in cases where the use of the precise vector field (17) is not practical, one can often use instead the approximated vector field (16). This approximation is valid under the assumption that both $\bar{\mathbf{g}}_1$ and $\bar{\mathbf{g}}_2$ are weak transformations (see Proposition D.11 in ([2], p. 410)). But this assumption is fully justified here since in any case, as explained in the following subsection, dot trajectories are only visible in the superposition if the layer transformations $\bar{\mathbf{g}}_1(x,y)$ and $\bar{\mathbf{g}}_2(x,y)$ are not too 'violent'. See, for example, Figure 10(*d*) which models the dot trajectories of Figure 8(*d*) using the approximation (16).

4.1. Limitations of the vector field model

Just like the fixed points model, the vector field model only works well when the transformations being applied are sufficiently weak. Although, mathematically speaking, the vector field $\bar{\mathbf{h}}(x,y)$ connecting the departure and destination points of each screen element can always be plotted for any layer transformation, dot trajectories can only be visible in the superposition (and correspond to the vector field $\bar{\mathbf{h}}(x,y)$) if the layer transformations being used are

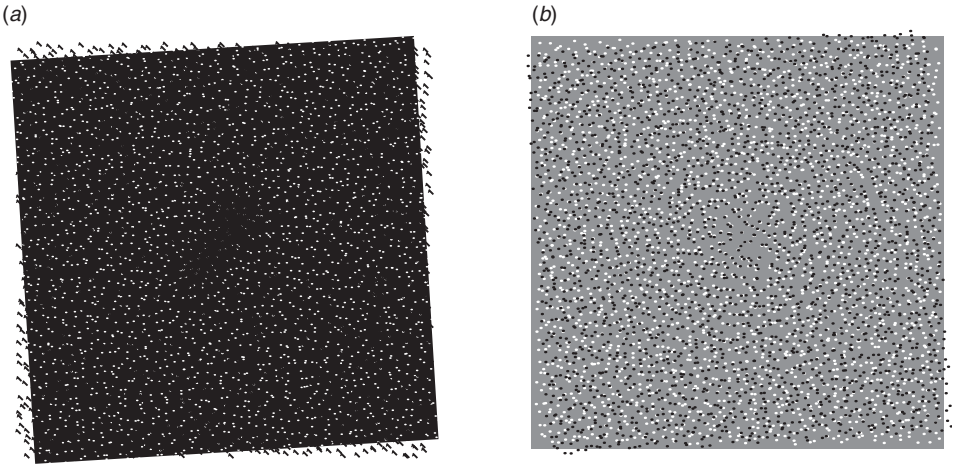

Figure 11. Cases where the vector field model does not predict correctly the dot trajectories. (*a*) Same as in Figure 7(*b*), but here one layer consists of tiny '1'-shaped dots while the second layer consists of tiny holes on a black background. (*b*) Same as in Figure 7(*b*), but here the unrotated layer has been replaced by its own negative, the black background of the resulting negative layer being replaced by an intermediate grey level. In both (*a*) and (*b*) the vector field model would predict circular dot trajectories as in Figure 7(*b*), since it only takes into account the geometric transformations undergone by the layers (in this case: rotation in one of the layers).

not too 'violent'. Otherwise, the correlation between the superposed layers is strongly reduced, and our eyes can no longer trace the departure/destination pairs in the layer superposition, so that the visual effect of the dot trajectories may be lost and no longer agree with the vector field plot.

Note, however, that here, too, this limitation can often be overcome by 'softening' the layer transformations being used. Thus, instead of using transformations of the form $\bar{\mathbf{g}}(x,y) = (x,y) + \bar{\mathbf{h}}(x,y)$, one may try to use their softened versions $\bar{\mathbf{g}}(x,y) = (x,y) + \varepsilon\bar{\mathbf{h}}(x,y)$ – where ε is a small positive fraction – which are closer to the identity transformation $\mathbf{g}(x,y) = (x,y)$ and thus softer. For example, if the transformations $\bar{\mathbf{g}}_1(x,y) = (x,y) + \frac{1}{2}\bar{\mathbf{h}}(x,y)$ and $\bar{\mathbf{g}}_2(x,y) = (x,y) - \frac{1}{2}\bar{\mathbf{h}}(x,y)$ that one wishes to apply by virtue of Equation (16) to synthesise dot trajectories having a desired shape $\bar{\mathbf{h}}(x,y)$ are not sufficiently weak for generating visible dot trajectories, one can try to use, instead, their softened versions $\bar{\mathbf{g}}_1(x,y) = (x,y) + \frac{1}{2}\varepsilon\bar{\mathbf{h}}(x,y)$ and $\bar{\mathbf{g}}_2(x,y) = (x,y) - \frac{1}{2}\varepsilon\bar{\mathbf{h}}(x,y)$.

A second, closely related limitation of the present model is that even when the layer transformations do provide visible dot trajectories, the vector field $\bar{\mathbf{h}}(x,y)$ only matches them in areas of the superposition where the corresponding dots of both layers remain close to each other (i.e. in areas where the arrow lengths in $\bar{\mathbf{h}}(x,y)$ are small). In areas where the dots get farther apart (and the arrow lengths in $\bar{\mathbf{h}}(x,y)$ increase) the correlation between the layers is reduced, and the dot trajectories are no longer visible. Note, however, that in such areas field lines still do exist in the vector field – in fact, the vectors in these areas are even longer, since the distance between the corresponding dots in the two layers is bigger. Thus, a visual agreement between the dot trajectories and the vector field is only possible in areas where the correlation between the superposed layers is sufficiently high (meaning that the vectors in the vector field are not too long). For example, by comparing Figures 7(*b*) and 10(*a*) we see that the vector field model corresponds here to the visible dot trajectories only up to a certain distance from the centre of the Glass pattern.

A further limitation of the vector field model is that it only can describe the dot trajectories that are obtained in the superposition of two aperiodic layers having the same element type (such as black dots on a white background). If the two superposed layers consist of elements of different shapes, or if the superposition rule being used is different from the classical one, this model may fail to describe correctly the shapes of the resulting dot trajectories. This limitation is illustrated in Figure 11.

5. Conclusions

Because the Fourier-based approach for modelling the moiré phenomenon is not always well adapted for real-world applications, other simpler models that only involve image-domain considerations are sometimes used instead. We briefly review three of these alternative models: the indicial equations model for moirés between curvilinear gratings; the fixed points model for Glass patterns in the superposition of correlated

aperiodic layers; and the vector field approach for the modelling of the dot trajectories in such aperiodic layer superpositions. We describe the limitations of each of these models, explain their significance, and suggest possible ways to overcome them. This should help one in choosing the most appropriate mathematical model for his needs, while fully understanding its advantages and its limitations.

Notes

1. Since the Fourier transform is a global operation that is applied to the entire spatial image domain, local microstructure effects are averaged together and buried in the global spectrum of the entire image. And even if we apply the Fourier transform (or a localised version thereof such as a wavelet transform) to different local areas of the entire image, it will only help us to distinguish between the different local microstructures and to identify and analyse their particular spectral properties; but this will not yet explain the various geometric shapes of the microstructure elements (rosettes etc.).
2. The $(1,-1)$-moiré between two superposed gratings is the moiré effect that is generated by the difference between the two original grating frequencies, f_1-f_2. More generally, the (k_1,k_2)-moiré between two superposed gratings is generated by the sum of the k_1 and k_2 harmonics of the original grating frequencies, namely, $k_1f_1+k_2f_2$. For a more detailed discussion see, for example ([1], Chapter 2).
3. The classical terms *subtractive moiré* and *additive moiré* designate moirés which correspond, respectively, to frequency differences or frequency sums. For example, the $(1,-1)$-moiré that is generated by the frequency difference f_1-f_2 is subtractive, while the $(1,1)$-moiré that is generated by the frequency sum f_1+f_2 is additive.
4. It can be shown that this equation leads indeed, to the classical formulas that give the period and the angle of the $(1,-1)$-moiré between two superposed line gratings (see [4], p. 170).
5. Note that a moiré of order >1, too, is the locus of points of intersection (see, for example, the dotted lines of the $(2,-1)$-moiré in Figure 2(*a*) or in Figure 3). Usually the density of points of intersection along such loci is lower than in a first-order moiré, so higher-order moirés are usually less clearly visible. But when this density is higher than the others, as in Figure 3, a higher-order moiré becomes dominant.
6. As shown in the examples above, in some cases $\mathbf{Z}$ should be replaced by $\mathbf{Z}^+$ (non-negative integers), etc.
7. Although the indicial equations model only considers integer values of p, we know from Fourier theory that this equality is not only limited to $p\in\mathbf{Z}$, but actually holds for all real values of p.
8. Note that in the first example in Section 2 we tacitly used $k_1=1$ and $k_2=-1$ (see paragraph starting from Equation (3)), while in the second example we tacitly used $k_1=1$ and $k_2=\pm1$ (see before Equation (6)). These values give, indeed, the most visible moirés in these examples – but this knowledge is not provided by the indicial equations model itself.
9. Note that $\mathbf{g}_i(x,y)$ are mappings of $\mathbf{R}^2$ onto $\mathbf{R}^2$; we denote them by a boldface letter since the values they return, $(u,v)=\mathbf{g}_i(x,y)$, are vectors.
10. We use the term *mutual fixed point* of $\mathbf{g}_1$ and $\mathbf{g}_2$ to designate a point (x_F,y_F) for which $\mathbf{g}_1(x_F,y_F)=\mathbf{g}_2(x_F,y_F)$. Similarly, a *mutual fixed locus* of $\mathbf{g}_1$ and $\mathbf{g}_2$ is a locus in the plane that consists of all the mutual fixed points of $\mathbf{g}_1$ and $\mathbf{g}_2$. Note that the term *common fixed point* of $\mathbf{g}_1$ and $\mathbf{g}_2$ is already used in the mathematical literature for a point (x_F,y_F) that satisfies $\mathbf{g}_1(x_F,y_F)=(x_F,y_F)=\mathbf{g}_2(x_F,y_F)$, but this definition is too restrictive for our needs.
11. Depending on the aperiodic layers being superposed, Glass patterns can sometimes be darker in their centre; see, for example, Figure 11. The explanation is provided by the Fourier-based model ([2], Chapter 7).
12. The first part of this theorem provides the intensity profile of the Glass pattern, while its second part provides the geometric transformation undergone by the glass pattern: $\mathbf{g}(x,y)=\mathbf{g}_1(x,y)-\mathbf{g}_2(x,y)$.
13. We use the upper bar notation to clearly indicate that the transformation in question is used here as a *direct* transformation. Note that in the previous sections the transformations $\mathbf{g}_i(x,y)$ were always applied to the given layers $p_i(x,y)$ as *domain* transformations giving $p_i(\mathbf{g}_i(x,y))$, and thus their effect was inversed. For example, when the transformation $(u,v)=(2x,2y)$ is used as a direct transformation, $(x,y)\mapsto(2x,2y)$, its effect is a two-fold expansion of the affected layer; but when it is applied to the same layer $p_i(x,y)$ as a domain transformation, as was the case in the previous sections, the result is $p_i(2x,2y)$, which is a two-fold *shrinked* version of the original layer. Note that the transformations $\bar{\mathbf{g}}$ and $\mathbf{g}$ are indeed the inverse of each other.
14. Any 2D transformation $\bar{\mathbf{g}}(x,y)$ can be also interpreted as a *vector field* that assigns to each point (x,y) in the x,y plane the vector $\bar{\mathbf{g}}(x,y)$. This vector field can be illustrated visually by drawing, starting from each point (x,y), an arrow having the length and the orientation of the vector $\bar{\mathbf{g}}(x,y)$.

References

[1] Amidror, I. *The Theory of the Moiré Phenomenon, Volume I: Periodic Layers*, 2nd ed.; Springer: New York, 2009.

[2] Amidror, I. *The Theory of the Moiré Phenomenon, Volume II: Aperiodic Layers*; Springer: Dordrecht, 2007.

[3] Amidror, I.; Hersch, R.D. *J. Mod. Opt.* **2009,** *56,* 1103–1118.

[4] Oster, G.; Wasserman, M.; Zwerling, C. *J. Opt. Soc. Am.* **1964,** *54,* 169–175.

[5] Durelli, A.J.; Parks, V.J. *Moiré Analysis of Strain*; Prentice-Hall: Englewood Cliffs, New Jersey, 1970.

[6] Patorski, K. *Handbook of the Moiré Fringe Technique*; Elsevier: Amsterdam, 1993.

[7] Raman, C.V.; Datta, S.K. *Trans. Opt. Soc.* **1926,** *27,* 51–55.

[8] Raman, C.V.; Datta, S.K. *Trans. Opt. Soc.* **1927,** *28,* 214–217.
[9] Leifer, I.; Walls, J.M.; Southworth, H.N. *Optica Acta* **1973,** *20,* 33–47.
[10] Walls, J.M.; Southworth, H.N. *Optica Acta* **1975,** *22,* 591–601.
[11] Oster, G. *The Science of Moiré Patterns*, 2nd ed.; Edmund Scientific: Barrington, NJ, 1969.
[12] Hersch, R.D.; Chosson, S. *Proc. of SIGGRAPH (2004), ACM Trans. Graphics* **2004,** *23,* 239–248.
[13] Glass, L. *Nature* **1969,** *223,* 578–580.
[14] Glass, L.; Pérez, R. *Nature* **1973,** *246,* 360–362.
[15] Pocheć, P. *Proceedings of the IEEE International Conference on Image Processing*, Washington DC, October 23–26, 1995.
[16] Glass, L. *Math. Intell.* **2002,** *24,* 37–43.