Christian Ronse

LSIIT UPRES-A 7005
Université Louis Pasteur
Département d'Informatique
7 rue René Descartes
F-67000 Strasbourg, France

Tel. 33-3-88.41.66.38
Fax. 33-3-88.61.90.69
E-mail: ronse@dpt-info.u-strasbg.fr
URL: http://gr6.u-strasbg.fr/~ronse/

The phase congruence model for edge detection in two-dimensional pictures: a mathematical study

C. Ronse

September 1995

Abstract. Morrone and co-workers have proposed a model for edge detection in grey-level images, based on psychophysical experiments in human vision. Assuming a one-dimensional visual signal, edges correspond to points of maximal Fourier phase congruence of the signal, and are localized at peaks of an energy function obtained as the quadratic combination of convolutions of the signal with two filters in Fourier quadrature (i.e., forming a Hilbert transform pair); such properties were explained in the case where the one-dimensional visual signal was periodic. In this report we make an in-depth theoretical study of this phase congruence model for two-dimensional non-periodic images. We consider one-dimensional edges and features (in the sense that significant grey-level changes occur along a single direction), but whose orientation is not fixed. We extend previous work in several respects:

— The mathematical framework is improved: both the signal and the filters are two-dimensional and non-periodic; new mathematical properties are obtained.

— We consider edges having various orientations, and characterize mathematically the joint determination of orientation and position of edges.

— We show that usual models for edge detection involving a single filter are a particular case of this phase congruence approach.

— We introduce some further filter combinations for classifying types of edges.

We discuss also of types of unidirectional edges and local features that can be encountered in natural images (steps, lines, ramps, Mach bands, compound edges, etc.), and of the appropriateness of the model to their detection and localization. In order to make this report mathematically self-contained, we have included a detailed exposition the basic mathematical theory of the Fourier and Hilbert transforms.

Key words. Edge and feature detection, edge types, symmetry, integrable and square-integrable functions, Fourier integral, Fourier phase and amplitude, constant phase, phase congruence, quadrature, directional Hilbert transform, linear and quadratic filters, energy feature detector.

Résumé. Morrone et ses collaborateurs ont proposé un modèle pour la détection d'arêtes dans les images à niveaux de gris, basé sur des expériences psychophysiques concernant la vision humaine. Supposant un signal visuel unidimensionnel, les arêtes correspondent aux points de congruence maximale des phases de Fourier du signal, et sont localisées aux sommets d'une fonction d'énergie obtenue par une combinaison quadratique du signal avec deux filtres en quadrature de Fourier (c.a.d. formant une paire de la transformée d'Hilbert); de telles propriétés furent expliquées dans le cas où le signal visuel unidimensionnel était périodique. Dans ce rapport nous faisons une étude théorique en profondeur de ce modèle de congruence de phases pour des images bidimensionnelles non péridodiques. Nous considérons des arêtes et traits unidimensionnels (dans le sens que les changements significatifs de niveaux de gris ont lieu dans une seule direction), mais dont l'orientation n'est pas fixée. Nous étendons les travaux précédents sur plusieurs points:

— Le cadre mathématique est amélioré: tant le signal que les filtres sont bidimensionnels et non-périodiques; de nouvelles propriétés mathématiques sont obtenues.
— Nous considérons des arêtes ayant des orientations variées, et caractérisons mathématiquement la détermination conjointe de l'orientation et de la position des arêtes.
— Nous montrons que les modèles usuels pour la détection d'arêtes impliquant un seul filtre sont un cas particulier de cette approche par congruence de phases.
— Nous introduisons de nouvelles combinaisons des filtres pour classer les types d'arêtes.

Nous discutons aussi des types d'arêtes et traits locaux unidirectionnels qu'on peut rencontrer dans les images naturelles (pas, lignes, bandes de Mach, arêtes composites, etc.), et de l'adéquation de ce modèle pour leur détection et leur localisation. Afin de rendre ce rapport autonome du point de vue mathématique, nous avons inclu un exposé détaillé de la théorie mathématique de base des transformées de Fourier et de Hilbert.

Mots clés. Détection d'arêtes et de traits, types d'arêtes, symétrie, fonctions intégrables et carré-intégrables, intégrale de Fourier, phase et amplitude de Fourier, phase constante, congruence de phases, quadrature, transformée de Hilbert directionnelle, filtres linéaires et quadratiques, détecteur de traits d'énergie.

1. Introduction

Edges are visually salient features in a grey-level image, whose positions in the plane form a one-dimensional structure. They include boundaries between distinct regions of a picture, as well as discontinuities inside such regions, or line drawings in sketches [1]. They can have various grey-level profiles, such as steps, lines, roofs, and combinations of these [2]. Numerous papers have appeared on the subject of edge detection, and research carries on with this topic, for no method presented so far is completely satisfactory. One of the first problems faced by anyone attempting to devise a new edge detector is to define precisely what is an edge. We see three types of definitions:

— *Physical definition:* An image is formed as light coming from certain sources is reflected towards the viewer by surfaces in three-dimensional space. Thus edges are the optical projection of object discontinuities, such as: changes in surface reflectance or orientation, object termination (leading to boundary contours), occlusion, shadows, etc. Edge characteristics are thus determined by photometric properties of various materials. This approach has been championed by Horn [2], and recently illustrated in [3]. It can be characterized as *objective* and *materialistic*.

— *Physiological or psychophysical definition:* Edges correspond to what humans (or primates) perceive as such. This does not say what are the characteristics of an edge (such as the kind of grey-level profile in its neighbourhood), but only where an edge must be detected in a given image. Here the emphasis is put on the definition of the *edge detector* rather than that of the edge. Visual response to artificial images are measured by psychophysical or electrophysiological studies, and edge detection operators are then modeled after the behaviour of visual neurons. This approach traces back from the experiments by Hubel and Wiesel [4] on "edge and bar detectors" in the primary visual cortex of monkeys. It can be considered as *subjective*.

— *Mathematical definition:* The grey-level image is a mathematical function, and edges form the locus of points where that function satisfies certain mathematical properties, such as: step discontinuity of the function or of one of its derivatives [5], non-analyticity, zero-crossing of the Laplacian [6], peak in the absolute value of the gradient of the Laplacian [7], Fourier phase congruence [8,9], etc. Alternatively the edge detection process can be modeled as a mathematically ill-posed problem which must be regularized [10]; for example in [11] a filter detecting step edges was considered as the regularization of differentiation. This approach can be characterized as *objective* and *idealistic*; indeed it relies often on abstract models of "ideal" edges with added noise.

Following Marr [12] there is a tendency to combine these three approaches, and to define edge detection in terms of photometry, physiology, and mathematics. In practice it is impossible to demand that an edge map computed from an image according to a mathematical algorithm, the human perception of edges in that image, and the optical projection of discontinuities in the scene giving rise to that image, should all three coincide *exactly*. However it is reasonable to ask that these three edge maps should be close to each other.

The phase congruence model for edge detection [8,9] considers that edges in a one-dimensional visual signal correspond to points where the cosine curves composing the signal (in a Fourier decomposition) have their phases in conjunction; this is for example the case for an ideal step or triangular line, see Figure 1. As indicated in [9], such edges can be detected as peaks of an energy function obtained as the sum of squares of convolutions of the signal with two constant phase filters forming a Hilbert transform pair (we will explain this in detail in Section 4, where we investigate the model). This abstract mathematical model has been deepened in [13,14,15], and more general models using quadratic combinations of various pairs of filters have been studied [16,17,18,19]. Similar approaches have been tested [20,21] in order to model the human visual system. Authors agree that the quadratic approach allows the accurate detection of both step and line edges, something which is difficult to achieve in usual methods using a single filter, such as [7].

In fact this mathematical model arises from psychophysical experiments on human detection of edges or Mach bands in vertical gratings (images whose grey-level is constant in the vertical direction) [22,9,23]. Related psychophysical [24] and electrophysiological [25] studies suggest that the responses of simple cells in the visual cortex have Fourier phases corresponding to those of the two filters used in phase congruence model [9]. Hence this approach to edge detection has a sound foundation in both mathematics and physiology. However there has to our knowledge been no study of its relation to photometry: what are the Fourier phase characteristics of edges arising in natural images from the optical projection of object discontinuities?

In this paper we make a systematical study of this model for one-dimensional edges in two-dimensional images. By one-dimensional edges we mean local images features where the grey-level changes significantly in one direction and is relatively constant in the perpendicular direction, such as steps, lines, and roofs having a relatively constant orientation; however the orientation of the edge is not fixed, we deal also with the selection of edge orientation. We exclude from this study the analysis of bi-directional visual features, where there are significant variations of grey-levels in two directions, such as corners, end-stopped edges, or strongly curved edges; these have been sudied in [26,27,28,29,30,31,32]; the latter four references are related to the phase congruence approach.

Another extension of edge detection with which we do not deal here, is the detection of spatiotemporal edges in moving images. For example in [33,34] moving images constitute a signal on a three-dimensional space-time, and spatiotemporal edges are surfaces in space-time; in [33] they are detected through an energy function obtained as a quadratic combination of convolutions of the signal with spatiotemporal filters in phase quadrature; in [34] one uses instead the sum of two filters which are in phase quadrature both in space and in time.

This paper is organized as follows. In Section 2 we discuss briefly the various types of edges that can be found in natural images, from the three points of view of optics, psychophysics, and computation; in particular we recall some advantages of quadratic edge detectors, and especially of the phase congruence model.

Section 3 is devoted to the mathematical foundations of the model: the L^p spaces and convolution, the Fourier transform, constant phase signals, phase quadrature and the directional Hilbert transform, the analytical signal and its energy, etc. This makes our paper self-contained from a mathematical point of view.

The phase congruence model of edge detection in non-periodic two-dimensional images is described in Section 4. We specify the spatial and Fourier requirements on the two filters, and we give the mathematical properties of the energy function as well as of other quadratic operators obtained from the convolution of the image with these two filters; then we interpret traditional single-filter approaches to edge detection (such as Canny's operator [7]) in terms of energy and phase congruence. Next we give a mathematical justification to the traditional method of selecting the edge orientation by taking filters at various orientations and choosing at each point the orientation giving the greatest energy function.

Finally Section 5 discusses various questions related to this model: the digitization of the filters, possible extensions towards the detection of bi-directional features (corners, end-stopped edges, junctions, or strongly curved edges), applications to other vision tasks, etc. That section ends with the conclusion.

2. Types of edges and visual features in natural images

This section is somewhat complementary to the remainder of the paper, because we will discuss here photometry and psychophysics rather than mathematics. We will describe several types of luminance profiles that we call edges, and which types of scene events give rise to them. Finally we will examine briefly how they are perceived by the human visual system, and how this can be relevant for computer vision.

Horn [2] showed that it is possible to describe the form of luminance profiles of edges in a grey-level image, according to the corresponding types surface discontinuities in the scene giving rise to that image, provided one makes some simplifying assumptions, such as:

— The surface of the objects in the scene is piecewise smooth.

— The reflectance of objects is *Lambertian*, that is: every surface patch looks equally bright from any viewing direction, and its brightness is proportional to the amount of light it receives.

— The primary illumination is *coherent*, in other words there is a single light source spanning a small solid angle (e.g., the sun, or a light bulb).

In general, things are more complicated. First, the surface is not always smooth, it can be grainy. Second, the reflectance of a patch on the surface of an object has two components [3]: a matte *body reflectance*, which is not necessarily Lambertian, and a glossy *surface reflectance*, which is not necessarily specular (mirror-like); moreover these two reflectance components may have distinct chrominances (in a coloured object, the gloss is generally whiter). Third, mutual reflections between the objects must be taken into account.

However, a simplified model of scene illumination, geometry, and reflectance can lead to qualitative distinctions among the various profiles given in the literature as models of ideal edges. As a consequence, it is possible to infer three-dimensional scene events from luminance profiles in a single two-dimensional image.

Let us first describe several types of luminance profiles encountered in practice. Afterwards we will explain briefly which types of scene events give rise to them.

We show in Figure 1 seven types of one-dimensional luminance profiles, and the names we have chosen to distinguish them. These profiles can be considered as representatives of *primary* types of edges. It is possible to add to the list the grey-level inversion of a line, bar, or roof, which are not shown.

We discriminate clearly between what we call a *line* edge and a *bar* edge; the latter has a plateau between the sides. The line corresponds to what Horn [2] calls a *peak*, and it is considered by him as a fundamental type of edge. In some other studies (for example [35,36]), authors call a line edge what we classify as a bar. The confusion between the two comes from the fact that early line detectors were designed for the recognition one-pixel wide lines, and used convolutions with 3×3 masks for this purpose; now in a digital framework, there is no distinction between one pixel thick bars and lines. Note that the phase congruence model recognizes a symmetric triangular line as a feature at all scales, while it classifies a bar as a feature at coarser scales only; at finer scales, a bar is recognized as a pair of two steps.

There can be progressive or sharp steps, in the sense that the transition between the low to the high grey-level can be continuous or abrupt. The difference between the two is physically meaningful, as recognized in [2] (we will see this again later). The *convex* edge was recognized by Ling [35] as distinct from the step; in other studies the convex edge has been ignored (for example by [2]) or identified with the step (as in [36]).

Roofs are commonly acknowledged in computer vision. This is not the case for *Mach bands*, which have generally been discussed only in relation to human visual perception (see in particular [22,9,23,21]). However they occur naturally at the extremity of *extended edges*, where the grey-level changes gradually over a relatively long distance (extended edges arise for example at the border of cast shadows). Marr [37] acknowledged Mach bands and extended edges, distinguishing the latter from the featureless gradual luminance profile due to the shading of a curved surface.

As shown in Figure 2, the above-mentioned primary edge profiles can be combined in several ways in order to produce more complex edge profiles, that we call *secondary*. First, the grey-levels of several profiles can be arithmetically added; we label this combination by giving the names of the primary profiles separated by + signs. We illustrate this operation in Figure 2 (a) with the step, line, and roof from Figure 1; the two additions shown there were recognized by Horn [2] as physically significant. Second, one can put the primary edge profiles in succession; this combination is labelled by giving the names of the primary profiles separated by commas. This is shown in Figure 2 (b) with a sharp step followed by a convex edge, and with two opposite convex edges; these two similar new profiles emulate an inverted line edge, and they were recognized by Ling [35], who called them *valley* edges.

In their 1970 study on edge detection [38], Herskovitz and Binford noticed that steps, lines, and roofs were the most frequent edge profiles in natural images [2]. Ling [35] examined edge profiles in images of surface mounted devices, and found these three profiles, but also convex edges and the compound edges of Figure 2 (b) that she called valley edges. In order to justify these experimental findings, let us now explain briefly how events in a scene geometry lead to the various types of luminance edges described above. For a more detailed exposition, the reader should refer to the literature (e.g., [2,39,35,40]); otherwise she can verify our assertions by a simple geometric reasoning, or by observation of natural scenes.

We make the simplifying assumptions that there is a single source of light, the scene geometry consists of smooth surfaces having the same reflectance, and that reflectance approximates the Lambertian rule as follows: the apparent brightness of a surface patch increases as the direction of the surface normal approaches that of incoming light.

When two faces of an object meet at a convex angle, since that angle is not really sharp, the corresponding luminance profile will be a gradual step or bright line (see Figure 1), or an addition of the two (see Figure 2 (a)). When these faces meet at a concave angle, the same happens, except that mutual reflections between the two faces lead to the addition of a positive roof in the luminance profile of the edge; we can thus get a compound step plus line plus roof edge (see Figure 2 (a)).

When a surface occludes part of another, the border between the two in the image is very sharp (since the effect of light diffraction is negligible); hence the luminance profile of the edge will generally involve a sharp step. When the occluding surface makes a convex angle at that border, this sharp step can be flanked by a gradual line or step (the luminance profile corresponding to the convex angle); on the other hand when that occluding surface is curved along that border, the sharp step is combined with a convex edge. When the occluded surface is also round, an opposite convex edge appears at the other side of the step. Thus quite complex luminance profiles can arise in such situations, but one observes frequently [2,35] a step or convex edge (see Figure 1), or one of Ling's valley edges shown in Figure 2 (b).

When two objects are juxtaposed, their surfaces meet in a groove, and the corresponding luminance profile is generally a negative (dark) line, sometimes combined with a positive (bright) one. Finally cast shadows have generally elongated edges: the boundary between the shadowed part of the surface and the illuminated one is fuzzy, because the light source is not punctual. At the extremities of such extended edges, the luminance profiles are similar to those of Mach bands (the junction between a plateau and a ramp, see Figure 1).

We have briefly explained how various surface discontinuities lead to different types of edges in the luminance profile. Note that changes in the orientation of a surface (w.r.t. illumination or viewpoint) have in general a weak influence on the chrominance of the resulting image: the surface will appear lighter or darker, but the hue and saturation of its reflected colour will not change, except at specular highlights [3]. Thus a change in image chrominance is often an indication of a change in the chromaticity of the surface reflectance, or a transition between two surfaces having different colours [3]; such chromatic edges are generally simple steps. This justifies the restriction to grey-level images for the analysis of complex edges.

Let us now relate edges to human vision. The human visual system perceives upward and downward steps, as well as light and dark lines, as distinct events [22,9,23]. Other types of edges are generally perceived as a step, a line, or a mixture of both; for examples roofs are perceived as lines [22]. We give an illustration of this fact in Figure 3. It represents a vertical grating which is horizontally periodic; the grey-level profile is a triangular wave at the top row, a square wave at the bottom row, and at intermediate rows it is a convex linear combination of the two, evolving gradually from triangular to square wave. One perceives light and dark lines at the top, but upward and downward steps at the bottom. At the middle the grey-level profile is as in Figure 4, and there is a mixed perception of a step flanked by a line on its left; this line is similar to a Mach band. Globally, the compound feature is seen slightly to the left of its true position, and this shift in location is consistent wih the phase congruence model (as we will explain with more technical details in Section 4). Other examples of mixed visual features can be found in [9].

The name of *Mach bands* refers to the discovery by the physicist Ernst Mach that the junction between a ramp and a plateau in the luminance leads to the perception of a narrow band (or line) at the end of the plateau; that band is light when the plateau is at the top of the ramp, and dark when the plateau is at the bottom of the ramp. As remarked in [22], such lines are also perceived at positive and negative roofs, for example in a triangular wave. We can thus consider roofs and Mach bands as similar types of features, which are equivalent up to the addition of a linear ramp signal; positive roofs or Mach bands lead to the perception of a light line, while negative ones lead to the perception of a dark line. There is not a complete agreement as to the exact position of the perceived line [23,21]: exactly at the junction, or on the side of the plateau, or on the side of the ramp? This indicates that the edge detectors of the human visual system may have a non-zero response to the underlying linear ramp (this problem will be discussed more precisely in Section 4).

A common interpretation of Mach bands is that they are visual illusions produced by the mechanism of *lateral inhibition*. As we must recognize the albedo of objects under various illumination intensities, the visual system does not measure the absolute luminance of each point, but luminance contrasts between neighbouring areas. Thus the grey-level of each point is compared to those in its neighbourhood. Then at the location of a positive roof or Mach band, the grey-level is higher than the average in its neighbourhood, leading to the perception of a light line, and conversely for the dark line seen at a negative one. This has led many authors to consider Mach bands as a "visual illusion", and not as true edges.

There are several arguments against this interpretation. First, experiments by Morrone et al. [22] on periodic vertical gratings (images whose grey-level is constant in the vertical direction and periodic in the horizontal one) have shown that the sharpness of the perceived Mach band is not related to the sharpness of the angle between the ramp and the plateau in the grey-level profile; it depends rather on Fourier phase characteristics of the image. Second, lateral inhibition should also apply to colours, since the spectrum of sunlight varies a great deal between morning and evening, and we can still perceive the intrinsic colour of objects and distinguish hues which differ by much less than the daily variation of the sunlight spectrum [41]. However in *isoluminant chromatic* images (made with colours having all the same lightness), Mach bands are not seen. This was found by coworkers of Koffka (see [42], pp. 170-171), and has been verified in recent experiments (D. Burr, private communication). Now it is well-known that the perception of the structure of an image (figure and ground, perspective, etc.) depends on luminance changes, and vanishes for isoluminant chromatic images (see [42], pp. 126-128, where it is called the "Liebmann effect"). This has been verified in neurophysiological studies of Livingstone and Hubel (see [43,44,45]), and it is justified by the above-mentioned fact that changes in surface orientation lead to variations in the image luminance, but without changes in its chrominance. We can thus agree with Koffka that Mach bands are not side-effects of visual mechanisms such as lateral inhibition, but rather a form of perceptual organization in the image: Koffka classifies correctly a discontinuity in the second derivative of luminance as an edge!

One property of human vision which must be taken into account for artificial vision, is that each visual attribute (for example a feature) corresponds to a size scale. For example a black cat seen from afar looks like a black blob; at moderate distance, limbs are distinguished; closer yet, the fur appears as a texture; within hand reach, individual hair are perceived. This was seriously recognized by Marr [12], who linked this fact to the existence of banks of visual filters tuned to different size scales. Since then, it is customary to analyse images with filters at several size scales (usually 3 or 4), with a scale factor of 2. One should note that the nature of an edge can be scale-dependent. We illustrate this in Figure 5, where an edge profile is shown with various spatial magnification factors; whenever it is enlarged, what we see correponds to what will be detected in the original profile at a smaller scale. Here it appears that a bar edge at coarse scale becomes a pair of step edges at a finer scale, and then a set of four Mach bands at a still finer scale. Thus, whenever one speaks of an edge, one must specify its scale; in practice, it corresponds to the size scale of the filters used to detect it: filters with wide grey-level profile detect features at coarse scale, and those with narrow profile detect features at fine scale.

Marr and Hildreth [6] required from significant features that they appear at all scales used.

One can even envisage "ideal" edges which should be detected and localized at the same position for all possible scales, arbitrarily small or large. In the case of the edges of Figure 1, an ideal line would be a Dirac delta, an ideal step a Heaviside step function, an ideal roof or Mach band would have its ramps on the two sides extending to infinity. Such ideal edges do not occur in natural images, but they can be used as a mathematical check for an edge detector: verify that when applying the detector to the ideal edge, that edge is properly detected and localized, and no additional feature is detected. We will indeed use such a condition in Section 4 when we will specify our filters. As we will also see there, a usual edge detector localizing edges at maxima of the absolute value of the filtered image has in general an underlying ideal edge, for which it gives an optimal response at all scales. We can mention in particular that for step detectors based on Gaussian-smoothed gradients, such as Canny's [7], a signal with constant Fourier phase $\pi/2$ is a perfect downward step (the phase is $-\pi/2$ for a perfect upward step); on the other hand for the phase congruence model, a perfect edge is any signal having constant Fourier phase, and this is an example of the fact that the phase congruence model generalizes some previous methods by allowing a wider family of edges to be detected.

Besides considerations in the Fourier domain and related experimental results on human vision, one of the rationales behind using a pair of filters (as in the phase congruence model), instead of a single one (as in older methods), is that two filters can respond to distinct types of features; for example an odd-symmetric filter would respond maximally to steps, and an even-symmetric one would respond maximally to lines and roofs [18,15]. Indeed, it is known [13] that traditional step detectors based on odd-symmetric filters detect two neighbouring steps in a line, and lead thus to edge duplication. Even then, the justification for combining them quadratically is not evident: one could for example apply separately a step detector and a line/roof detector, and combine together the two edge maps obtained separately. We will show in Section 4 that this leads again to edge duplication: for a combined step + line edge (see Figure 2 (a)), the two detectors will localize the step and the line at opposite sides of the true edge location.

This argument does not exclude the use of quadratic combinations of three or more filters, as suggested in [18]. We can justify our choice of only two filters on the following grounds:

— In human visual perception, it seems that every edge is seen as a mixture of a line and a step; this suggests a combination of two filters, an even-symmetric one leading to the perception of a line (a "line detector"), and an odd-symmetric one leading to the perception of a step (a "step detector"). This correlates with classical neurophysiological findings concerning the behaviour of simple cells in the primary visual cortex of monkeys [4].

— We can classify edge profiles in two groups: the first one contains odd-symmetric signals superposed on a dc level (in their Fourier decomposition, all nonzero frequencies have phase $\pm\pi/2$); the second one contains even-symmetric signals (in their Fourier decomposition, all frequencies have phase 0 or π). Thus the first group contains the Heaviside step, while the second one contains Dirac's delta and any symmetric roof. Every edge would thus be decomposed into the sum of two edges from the two groups.

— We have not yet seen any theoretical or practical argument showing the advantage of taking more than two filters for unidirectional features. On the other hand, with two constant-phase filters in Fourier quadrature (that is, their respective constant Fourier phases differ by $\pi/2$, and their Fourier amplitudes are equal), we have an elegant mathematical theory of phase congruence developed in Sections 3 and 4.

We do not exclude the possibility that future studies may suggest the need for more than two filters.

Note that some authors [13,16,17,18,19,29,30,31,33] have applied the quadratic approach (taking the sum of squares of convolutions of the image with two filters), in the case where the two filters have indeed constant Fourier phase (respectively 0 and $\pi/2$), but have different Fourier amplitudes, and so do not satisfy the requirement of Fourier quadrature of the phase congruence model. Serious reason should be given for omitting this constraint on the filters, because we will see in Section 4 that it leads to many interesting results. One possible justification would be that line and step edges in natural images have different Fourier amplitude spectra, so that the two filters in the edge detector should be adapted to this fact. There can also be mathematical justifications; see for example [17], where the pair of filters consisting of a n-th derivative of a Gaussian and of its derivative is shown to satisfy the causality requirement in scale-space.

There remains one question: does the phase congruence model accurately detect and localize edges in natural images? We can give here a partial theoretical answer. The ideal steps, lines, and roofs shown in Figure 1 have constant phase at the edge location (the phase is zero for lines and roofs, but $-\pi/2$ for positive-going steps); as we will see in Section 4, the energy function obtained as the sum of squares of convolutions of the signal with two constant phase filters in phase quadrature, has an absolute maximum at this edge position. Ramp edges and Mach bands are a linear combination of a symmetric roof and a linear ramp signal; by choosing the two filters in such a way that they have a zero response on linear signals, the energy function will be the same as for a symmetric roof, and it will have an absolute maximum at the edge position. The compound step + line or step + line + roof edges shown in Figure 2 (a) are the sum of a signal with constant zero phase and another one with constant $-\pi/2$ phase; hence they have all Fourier phases comprised between 0 and $-\pi/2$ at the edge location; it is thus likely that these phases will be maximally congruent at a point near the edge location, where the energy function will reach a local maximum. For example in the grating of Figure 3, whose grey-level profile in the middle rows is shown in Figure 4, maximum phase congruence is achieved slightly to the left of the true edge location, and this is consistent with visual perception.

From a practical point of view, experiments made among others by Morrone, Owens and, Venkatesh [1,8,9,13,14] indicate that this model gives satisfactory results in natural images.

A related question is whether the phase congruence model can give false edges, in other words if one can get local maxima of the energy function which do not correspond to true edges. For example Kube and Perona [17] have shown that for many choices of the pair of filters (in particular, Hilbert transform pairs), edges can appear at a coarse scale, which do not correspond to edges at finer scales, and there are some reasons to suspect that they might correspond to spurious local maxima of the energy function. They illustrated this fact with the pair consisting of the second derivative of a Gaussian and its Hilbert transform. This problem is rather delicate. For example in the edge shown in Figure 5, it is legitimate to postulate that there is a single bar edge at coarse scale, two step edges at medium scale, and four Mach bands at fine scale, and it is not obvious if the evolution in scale space through these three edge maps will be continuous. We feel that filter specification in the Fourier domain is unsufficient, and we give in Section 4 some requirements in the spatial domain that the filters must satisfy in order to avoid the detection of false edges in ideal steps, lines and roofs. More research should be conducted on non-Fourier requirements for quadratic edge detectors. The problem of choosing a good pair of filters is not yet solved.

Another possible answer to this question is to restrict edges to *regional* maxima of the energy function, that is points where that function has a maximum within a certain radius r corresponding to the scale of analysis (in fact, to the spatial extent of the filters), so that purely local maxima are eliminated. Indeed, for ideal edges, the energy function reaches an *absolute* maximum at the edge location, and we might plausibly deduce that for several ideal edges distant from each other by at least $2r$, each one may give a regional maximum of the energy function within radius r. We have taken this approach in Section 4. Further research should elucidate criteria for selecting "good" maxima of the energy function, or equivalently, "good" maxima of phase congruence.

3. Mathematical foundations of the phase congruence model

In order to make this paper self-contained, we recall the mathematics underlying the phase congruence model; this will essentially be Fourier analysis in L^1 and L^2, and some elementary facts derived from it. We introduce first our notation:

Write $\mathbb{R}$ for the set of real numbers and $\mathbb{C}$ for the set of complex numbers. Let $\mathcal{E} = \mathbb{R}^d$ for some integer $d \geq 1$; we will consider all signals in the spatial or Fourier domain as functions $\mathcal{E} \to \mathbb{C}$ (or $\mathcal{E} \to \mathbb{R}$). Although the phase congruence model will be studied in the next section for $d = 2$ or $d = 1$, we make no such assumption here, in order to allow the extension of this model to volumetric or moving images (cfr. the spatiotemporal edge model of [33,34]).

We write: x, y, z, etc. for real or complex variables; $\mathbf{x}$, $\mathbf{y}$, $\mathbf{z}$, etc. for vectors in $\mathcal{E}$; f, g, etc. for functions $\mathbb{R} \to \mathbb{C}$; F, G, etc. for functions $\mathcal{E} \to \mathbb{C}$. We write $\mathbf{x} \cdot \mathbf{y}$ for the scalar product of $\mathbf{x}$ and $\mathbf{y} \in \mathcal{E}$, and $|\mathbf{x}|$ for the Euclidean norm of $\mathbf{x}$, that is $|\mathbf{x}| = (\mathbf{x} \cdot \mathbf{x})^{1/2}$. For $x \in \mathbb{C}$, let $\overline{x}$ be its complex conjugate and $|x|$ its absolute value. For a function $F : \mathcal{E} \to \mathbb{C}$, we define $\overline{F}$ and $|F|$ by $\overline{F}(\mathbf{x}) = \overline{F(\mathbf{x})}$ and $|F|(\mathbf{x}) = |F(\mathbf{x})|$. We define the *signum* function sgn on $\mathbb{C}$ by

$$\operatorname{sgn}(x) = \begin{cases} x/|x| & \text{if } x \neq 0; \\ 0 & \text{if } x = 0. \end{cases}$$

In particular for $x \in \mathbb{R}$, we have $\operatorname{sgn}(x) = 1$ if $x > 0$ and $\operatorname{sgn}(x) = -1$ if $x < 0$.

The *reflection* F_ρ of a function F is given by $F_\rho(\mathbf{x}) = F(-\mathbf{x})$. We say that F is *even-symmetric* if $F_\rho = F$, *odd-symmetric* if $F_\rho = -F$, and *conjugate-symmetric* if $F_\rho = \overline{F}$. Every function F can be decomposed in a unique way as the sum of an even-symmetric function and an odd-symmetric one, namely $(F + F_\rho)/2$ and $(F - F_\rho)/2$. We say that the function F *vanishes at infinity* if $\lim_{|\mathbf{x}| \to \infty} F(\mathbf{x}) = 0$.

For any $\mathbf{u} \in \mathcal{E}$, let $\tau_\mathbf{u}$ be the translation by $\mathbf{u}$, which moves horizontally by $\mathbf{u}$ the graph of a function F, that is $\tau_\mathbf{u}(F)$ is defined by $\tau_\mathbf{u}(F)(\mathbf{x}) = F(\mathbf{x} - \mathbf{u})$. We define $\text{cis}_\mathbf{u}$, the "cisoid" function of frequency $\mathbf{u}$, by setting for $\mathbf{x} \in \mathcal{E}$:

$$\text{cis}_\mathbf{u}(\mathbf{x}) = \exp(2\pi i\, \mathbf{u} \cdot \mathbf{x}) = \cos(2\pi\, \mathbf{u} \cdot \mathbf{x}) + i \sin(2\pi\, \mathbf{u} \cdot \mathbf{x}). \tag{3.1}$$

Note that we consider frequency in cycles per unit, and not angular frequency in radians per unit; in this we follow [46].

We will write $\mathbf{n}$ for a unit vector in $\mathcal{E}$ ($\mathbf{n} \cdot \mathbf{n} = 1$); in the next section, it will be interpreted as the unit vector normal to the edge. It partitions the space $\mathcal{E}$ into the three sets

$$\mathcal{E}_\mathbf{n}^+ = \{\mathbf{x} \in \mathcal{E} \mid \mathbf{n} \cdot \mathbf{x} > 0\},$$
$$\mathcal{E}_\mathbf{n}^- = \{\mathbf{x} \in \mathcal{E} \mid \mathbf{n} \cdot \mathbf{x} < 0\},$$
$$\mathcal{E}_\mathbf{n}^0 = \{\mathbf{x} \in \mathcal{E} \mid \mathbf{n} \cdot \mathbf{x} = 0\}.$$

We define also the three functions $\text{pos}_{\mathbf{n}}$, $\text{neg}_{\mathbf{n}}$, and $\text{sgn}_{\mathbf{n}}$ on $\mathcal{E}$ by setting for $\mathbf{x} \in \mathcal{E}$:

$$\text{pos}_{\mathbf{n}}(\mathbf{x}) = \begin{cases} 1 & \text{if } \mathbf{n} \cdot \mathbf{x} > 0, \\ 0 & \text{if } \mathbf{n} \cdot \mathbf{x} \leq 0; \end{cases}$$

$$\text{neg}_{\mathbf{n}}(\mathbf{x}) = \begin{cases} 1 & \text{if } \mathbf{n} \cdot \mathbf{x} < 0, \\ 0 & \text{if } \mathbf{n} \cdot \mathbf{x} \geq 0; \end{cases}$$

$$\text{sgn}_{\mathbf{n}}(\mathbf{x}) = \begin{cases} +1 & \text{if } \mathbf{n} \cdot \mathbf{x} > 0, \\ 0 & \text{if } \mathbf{n} \cdot \mathbf{x} = 0, \\ -1 & \text{if } \mathbf{n} \cdot \mathbf{x} < 0. \end{cases}$$

(Cfr. the definition of the signum function sgn above). Note that $\text{sgn}_{\mathbf{n}} = \text{pos}_{\mathbf{n}} - \text{neg}_{\mathbf{n}}$.

Suppose that $\mathcal{E} = \mathcal{E}_1 \times \mathcal{E}_2$, the product of two orthogonal subspaces $\mathcal{E}_1$ and $\mathcal{E}_2$ of respective dimensions d_1 and d_2 (where $d_1, d_2 \geq 1$ and $d = d_1 + d_2 \geq 2$); every $\mathbf{x} \in \mathcal{E}$ can be written as an ordered pair $(\mathbf{x}_1, \mathbf{x}_2)$, where $\mathbf{x}_1 \in \mathcal{E}_1$ and $\mathbf{x}_2 \in \mathcal{E}_2$. Given a function $F : \mathcal{E} \to \mathbb{C}$, for any $\mathbf{x}_1 \in \mathcal{E}_1$ and $\mathbf{x}_2 \in \mathcal{E}_2$ we define the *$\mathcal{E}_1$-section* $F_{\mathbf{x}_1} : \mathcal{E}_2 \to \mathbb{C} : \mathbf{x}_2 \mapsto F(\mathbf{x}_1, \mathbf{x}_2)$ and the *$\mathcal{E}_2$-section* $F^{\mathbf{x}_2} : \mathcal{E}_1 \to \mathbb{C} : \mathbf{x}_1 \mapsto F(\mathbf{x}_1, \mathbf{x}_2)$, in other words $F_{\mathbf{x}_1}(\mathbf{x}_2) = F^{\mathbf{x}_2}(\mathbf{x}_1) = F(\mathbf{x}_1, \mathbf{x}_2)$ (see [48], p. 63).

We use on the space $\mathcal{E}$ the Lebesgue measure and integral as in [47] (see also [48,49]). We write $d\mathbf{x}$ for $dx_1 \cdots dx_d$, the d-dimensional Lebesgue measure element in an integral (cfr. [46]). Thus we write $\int_{\mathcal{S}} F(\mathbf{x})\, d\mathbf{x}$ or simply $\int_{\mathcal{S}} F$ for the integral of F over a Lebesgue measurable set $\mathcal{S} \subseteq \mathcal{E}$; the notation $\int_{\mathcal{S}} d\mathbf{x}\, F(\mathbf{x})$ is even more convenient for multiple integrals (cfr. [50]). Note that in measure theory, "integrable" means "integrable in absolute value": a measurable function F is integrable over $\mathcal{S}$ if and only if $|F|$ is integrable over $\mathcal{S}$, in other words $\int_{\mathcal{S}} |F(\mathbf{x})\, d\mathbf{x}| < +\infty$. From now on all subsets of $\mathcal{E}$ and all functions on $\mathcal{E}$ are implicitly assumed to be Lebesgue measurable.

A *negligible* set is a set whose Lebesgue measure equals zero. As customary, we say "almost all ..." for "all ..., except in a negligible set", and "almost everywhere" means "for almost all points"; this can be abbreviated by "a.e.". For example, we say that function F is *a.e. even-symmetric* if $F(\mathbf{x}) = F(-\mathbf{x})$ almost everywhere.

A function F is called *locally integrable* if for any compact subset $\mathcal{S}$ of $\mathcal{E}$, the restriction of F to $\mathcal{S}$ is integrable, in other words $\int_{\mathcal{S}} |F| < +\infty$. The *Lebesgue set* of a locally integrable function F on $\mathcal{E}$ is the set of all points $\mathbf{x} \in \mathcal{E}$ such that

$$\lim_{r \to 0} \frac{1}{r^d} \int_{|\mathbf{t}|<r} |F(\mathbf{x} - \mathbf{t}) - F(\mathbf{x})| \; d\mathbf{t} = 0. \tag{3.2}$$

It contains in particular all points where F is continuous. Moreover, allmost all points of $\mathcal{E}$ belong to the Lebesgue set of F (see [46], pp. 12 and 13, or [48], pp. 92 and 93).

We define the equivalence relation $\equiv$ on functions by setting $F \equiv G$ for two functions F, G if $F(\mathbf{x}) = G(\mathbf{x})$ a.e.; in particular $F(\mathbf{x}) = G(\mathbf{x})$ for all points $\mathbf{x}$ at which both F and G are continuous; furthermore, if F and G are locally integrable, then $F(\mathbf{x}) = G(\mathbf{x})$ for all points $\mathbf{x}$ in the Lebesgue set of both F and G. Clearly the equivalence $\equiv$ is compatible with algebraic operations, and for integrable functions, $F \equiv G$ implies that $\int F = \int G$.

3.1. L^p spaces and convolution

Let p be such that $1 \leq p \leq \infty$. For $p < \infty$, the *L^p norm* $\| \; \|_p$ of is defined by $\|F\|_p = \left(\int_{\mathcal{E}} |F(\mathbf{x})|^p \, d\mathbf{x}\right)^{1/p}$ for every function $F : \mathcal{E} \to \mathbb{C}$. For $p = \infty$, the L^∞ norm $\|F\|_\infty$ of F is the *essential supremum* of all $|F(\mathbf{x})|$, in other words the least $m \in [0, \infty]$ such that $|F(\mathbf{x})| \leq m$ almost everywhere.

Let L^p be the set of functions F such that $\|F\|_p < \infty$; it is a vector space in which $\| \ \|_p$ is a norm in the classical sense [47,48,49]. For example L^1 is the space of integrable functions; functions in L^2 are called *square-integrable*; functions in L^∞ are called *essentially bounded*. For any F, G in L^p, $F \equiv G$ if and only if $\|F - G\|_p = 0$. Let $L^p/\!\equiv$ be the set of equivalence classes of $\equiv$ over L^p. Then the L^p norm makes $L^p/\!\equiv$ into a metric space, where the distance between two functions F and G is $\|F - G\|_p$. A well-known theorem states that the metric space $L^p/\!\equiv$ is *complete* (i.e., every Cauchy sequence converges).

For $p < \infty$, functions in L^p are locally integrable; hence allmost all points of $\mathcal{E}$ belong to their Lebesgue set. Furthermore these functions satisfy the property of *L^p-continuity*, namely that for F in L^p,

$$\lim_{\mathbf{u}\to\mathbf{0}} \|F - \tau_{\mathbf{u}}(F)\|_p = 0. \tag{3.3}$$

A proof can be found in Proposition 8.5 of [48]. Note that (3.3) does not necessarily hold for $p = \infty$.

In the sequel, we will use the following integrability criterion:

LEMMA 3.1. *Let the function $F : \mathbb{R}^d \to \mathbb{C}$ be such that for every subset I of $\{1, \ldots, d\}$ there is some $p(I)$ with $1 \leq p(I) < \infty$ for which the function $\mathbb{R}^d \to \mathbb{C} : (x_1, \ldots, x_d) \mapsto \prod_{i\in I} x_i \cdot F(x_1, \ldots, x_d)$ is in $L^{p(I)}$ (in particular F is in $L^{p(\emptyset)}$). Then F is integrable.*

PROOF. Let $S = [-1, 1]$ and $T = \mathbb{R} \setminus S = \{x \in \mathbb{R} \mid |x| > 1\}$. Take $I \subseteq \{1, \ldots, d\}$ of size n ($0 \leq n \leq d$); let G be the function $\mathbb{R}^d \to \mathbb{C} : (x_1, \ldots, x_d) \mapsto \prod_{i\in I} x_i$, let $q = p(I)$, and let $\mathcal{V}$ be the set of points $(x_1, \ldots, x_d)$ such that $|x_i| > 1$ for $i \in I$ and $|x_i| \leq 1$ for $i \notin I$. If $q = 1$, then FG is integrable, and as $|F| \leq |FG|$ on $\mathcal{V}$, F is integrable on $\mathcal{V}$. Otherwise let $q' = q/(q - 1)$; thus $1 < q' < \infty$. We have

$$\int_{\mathcal{V}} |1/G|^{q'} = \Big(\int_S 1^{q'} \, dx\Big)^{d-n} \Big(\int_T 1/|x^{q'}| \, dx\Big)^n = 2^{d-n}\big(2/(q' - 1)\big)^n = 2^d(q - 1)^n.$$

As FG is in L^q, Hölder's inequality gives then

$$\int_{\mathcal{V}} |F| = \int_{\mathcal{V}} |FG \cdot 1/G| \leq \Big(\int_{\mathcal{V}} |FG|^q\Big)^{\frac{1}{q}} \cdot \Big(\int_{\mathcal{V}} |1/G|^{q'}\Big)^{\frac{1}{q'}} = \|FG\|_q \cdot \big(2^d(q - 1)^n\big)^{\frac{q-1}{q}} < \infty.$$

Hence F is integrable over $\mathcal{V}$. As $\mathbb{R}^d$ is the union of all such sets $\mathcal{V}$ for I ranging over the set of parts of $\{1, \ldots, d\}$, F is integrable over $\mathbb{R}^d$. ∎

For any two functions F and G, their *convolution* $F * G$ is given by

$$[F * G](\mathbf{x}) = \int_{\mathcal{E}} F(\mathbf{x} - \mathbf{t})G(\mathbf{t}) \, d\mathbf{t} \tag{3.4}$$

(whenever the integral exists). This operation is bilinear and commutative. Moreover, it commutes with translation, that is $F * \tau_{\mathbf{u}}(G) = \tau_{\mathbf{u}}(F * G) = \tau_{\mathbf{u}}(F) * G$ for all $\mathbf{u} \in \mathcal{E}$. The convolution by a function in L^1 is a stable operation w.r.t. the L^p norm (see Theorem 1.3 of [46] and Theorem 8.7 of [48]):

— **Young's inequality:** *Let F be in L^1 and G in L^p ($1 \leq p \leq \infty$). Then $(F * G)(x)$ is defined almost everywhere, $F * G$ belongs to L^p, and $\|F * G\|_p \leq \|F\|_1 \|G\|_p$.*

In particular the convolution operation is a bilinear product operation in $L^1/\!\equiv$, which is associative in the sense that for F, G, H in L^1, $F * (G * H) \equiv (F * G) * H$.

A slightly stronger result holds for the convolution of a function in L^p with one in $L^{p/p-1}$ (see Theorem 8.8 of [48]):

— **p-p' convolution property:** *Let F be in L^p and G in $L^{p'}$, where $1 \leq p, p' \leq \infty$ and $(1/) + (1/p') = 1$. Then $(F * G)(x)$ is defined for all $\mathbf{x} \in \mathcal{E}$, $F * G$ is uniformly continuous and bounded: for all $x \in \mathcal{E}$, $\big|[F * G](x)\big| \leq \|F\|_p \|G\|_{p'}$. Furthermore, $F * G$ vanishes at infinity, provided that one of the following is satisfied:*

(*a*) *$p, p' < \infty$;*

(*b*) *$p = \infty$ and F vanishes at infinity;*

(*c*) *$p' = \infty$ and G vanishes at infinity.*

A particular case is $p = 1$ and $p' = \infty$: the convolution of a function in L^1 and one in L^∞ is uniformly continuous and bounded. Another one is given by $p = p' = 2$.

The above two properties are interesting enough to justify the choice of functions in L^1 for the filters applied to the image. Besides, grey-levels of pictures usually belong to a bounded range; so in view of the p-p' convolution property, if we want to have the picture convolved with a mask to get its grey-levels in the same range, we must require the convolution function to be in L^1. Furthermore, the uniform continuity of the result of the convolution is interesting in view of digitization, as we will explain in Section 5.

In the sequel, we will generally restrict ourselves to the spaces L^1, L^2, and L^∞. For example in Section 4, all visual signals will be the sum of three components in L^1, L^2, and L^∞ respectively.

3.2. The Fourier transform in L^1 and L^2

As said above, we follow [46] in considering that frequency is in cycles per unit, and not angular frequency in radians per unit. Thus the factor 2π precedes the frequency in the argument of an imaginary exponential (cfr. (3.1)). As in [46], we write $\widehat{F}$ for the Fourier transform of a function F; the Fourier transform of an expression $(\ldots)$ will be written $(\ldots)^\wedge$.

The *Fourier transform* $\widehat{F}$ of an integrable function F is defined pointwise by the *Fourier integral*:

$$\widehat{F}(\mathbf{u}) = \int_{\mathcal{E}} F(\mathbf{x}) \exp[-2\pi i\, \mathbf{u} \cdot \mathbf{x}]\, d\mathbf{x}. \tag{3.5}$$

In particular $\int_{\mathcal{E}} F = \widehat{F}(0)$.

For a non-integrable function, the above formula does not apply. Since the Fourier transform preserves the L^2 norm of functions in $L^1 \cap L^2$, and $L^1 \cap L^2$ is dense in L^2, we can extend it to an isometry of the complete metric space $L^2/\!\equiv$. In other words the Fourier transform of a square-integrable function F is the function F^* in L^2, unique up to equivalence by $\equiv$, such that for every sequence of functions F_n in $L^1 \cap L^2$ satisfying $\lim_{n\to\infty} \|F_n - F\|_2 = 0$, we have $\lim_{n\to\infty} \|\widehat{F_n} - F^*\|_2 = 0$. Thus the Fourier transform is defined on $L^2/\!\equiv$ rather than L^2, in other words the Fourier transform of a square-integrable function is not defined pointwise, but up to a negligible set. For a function F which is both integrable and square-integrable, the two definitions of $\widehat{F}$ in L^1 and L^2 coincide up to equivalence by $\equiv$.

Given an integrable or square-integrable function F, write $F^{\mathcal{A}}$ for the *Fourier amplitude* of F, that is $F^{\mathcal{A}}(\mathbf{u}) = \big|\widehat{F}(\mathbf{u})\big|$, and F^{Φ} for the *Fourier phase* of F, that is $\widehat{F}(\mathbf{u}) = F^{\mathcal{A}}(\mathbf{u}) \exp\big[iF^{\Phi}(\mathbf{u})\big]$.

Let us recall the main properties of the Fourier transform (see Chapter 1 of [46]). By default, all functions are assumed in $L^1 \cup L^2$. First the *uniqueness property*: for F, G in $L^1 \cup L^2$, $\widehat{F} \equiv \widehat{G}$ if and only if $F \equiv G$; furthermore, for F, G in L^1, $\widehat{F} \equiv \widehat{G}$ implies the pointwise equality $\widehat{F} = \widehat{G}$. Next, we have several elementary formulas. First, the commutation with reflexion:

$$\left(F_\rho\right)^\wedge = \left(\widehat{F}\right)_\rho. \tag{3.6}$$

Let us write $F^\vee$ for $(F_\rho)^\wedge = (\widehat{F})_\rho$. When F is integrable we have

$$F^\vee(\mathbf{u}) = \int_{\mathcal{E}} F(\mathbf{x}) \exp[2\pi i\, \mathbf{u}\cdot\mathbf{x}]\, d\mathbf{x}.$$

By (3.6) and the uniqueness property, F is a.e. even-symmetric (resp., a.e. odd-symmetric) if and only if $\widehat{F}$ is a.e. even-symmetric (resp., a.e. odd-symmetric). Next:

$$\widehat{\overline{F}}(\mathbf{u}) = \overline{\widehat{F}(-\mathbf{u})}. \tag{3.7}$$

From (3.6) and (3.7) we get:

$$\widehat{\overline{F}} = \overline{F^\vee} \qquad \text{and} \qquad \overline{F}^\vee = \overline{\widehat{F}}. \tag{3.8}$$

For $\mathbf{h} \in \mathcal{E}$ we have:

$$[\tau_{\mathbf{h}}(F)]^\wedge = \mathrm{cis}_{-\mathbf{h}} \cdot \widehat{F}; \tag{3.9}$$

$$[\mathrm{cis}_{\mathbf{h}} \cdot F]^\wedge = \tau_{\mathbf{h}}(\widehat{F}). \tag{3.10}$$

Besides the definition, there are important differences between the Fourier transform in L^1 and L^2:

— **Riemann-Lebesgue theorem:** *For F in L^1, $\widehat{F}$ is uniformly continuous and bounded on $\mathcal{E}$: $\left|\widehat{F}(\mathbf{u})\right| \leq \|F\|_1$ for all $\mathbf{u} \in \mathcal{E}$; moreover $\widehat{F}$ vanishes at infinity.*

— **Plancherel theorem:** *For F in L^2, $\widehat{F}$ is in L^2, $\|\widehat{F}\|_2 = \|F\|_2$, and $F_\rho \equiv \widehat{\widehat{F}}$, in other words $F \equiv \widehat{F}^\vee$.*

Thus the Fourier transform is an invertible isometry on $L^2/\equiv$. On L^1 we have the following result concerning the inversion of the Fourier transform:

— $\mathbf{L^1}$ **Fourier inversion theorem:** *Given F in L^1 such that $\widehat{F}$ is in $L^1 \cup L^2$, then F is in L^2 and $F_\rho \equiv \widehat{\widehat{F}}$, in other words $F \equiv \widehat{F}^\vee$.*

There remain two fundamental formulas:

— **Convolution formula:** *Let F be in L^1 and G in $L^1 \cup L^2$. Then $[F * G]^\wedge = \widehat{F}\widehat{G}$ (a.e. for G not integrable).*

— **Multiplication formula:** *Let F, G be either both in L^1 or both in L^2. Then $F\widehat{G}$ and $\widehat{F}G$ are integrable and $\int_{\mathcal{E}} F\widehat{G} = \int_{\mathcal{E}} \widehat{F}G$.*

For F, G both in L^2, by (3.9) and the Plancherel theorem, the Fourier transform of $\tau_{\mathbf{x}}(F^\vee) : \mathbf{t} \mapsto \widehat{F}(\mathbf{x}-\mathbf{t})$ is $\mathrm{cis}_{-\mathbf{x}}(F^\vee)^\wedge = \mathrm{cis}_{-\mathbf{x}} \cdot F : \mathbf{t} \mapsto \exp[-2\pi i\, \mathbf{t}\cdot\mathbf{x}]F(\mathbf{t})$; the multiplication formula gives then

$$\int_{\mathcal{E}} \widehat{F}(\mathbf{x}-\mathbf{t})\widehat{G}(\mathbf{t})\, d\mathbf{t} = \int_{\mathcal{E}} \exp[-2\pi i\, \mathbf{t}\cdot\mathbf{x}]F(\mathbf{t})G(\mathbf{t})\, d\mathbf{t}.$$

We get thus the following:

— **Dual convolution formula:** *Let F, G be both in L^2. Then $[FG]^\wedge = \widehat{F} * \widehat{G}$.*

Given a bijective linear transform α of $\mathcal{E}$, the function G defined by $G(\mathbf{x}) = F(\alpha(\mathbf{x}))$ has its Fourier transform given by $\widehat{G}(\mathbf{u}) = |\det(\alpha)|^{-1}\widehat{F}(\alpha^{-T}(\mathbf{u}))$, where $\det(\alpha)$ is the determinant of α and α^{-T} is the inverse transpose of α. In particular if α is an isometry (α is its own inverse transpose), then $\widehat{G}(\mathbf{u}) = \widehat{F}(\alpha(\mathbf{u}))$, in other words the Fourier transform commutes with α. For example reflection commutes with the Fourier transform (cfr. (3.6)); the Fourier transform commutes also with rotations of $\mathcal{E}$.

The following property is fundamental for multidimensional Fourier analysis: a multidimensional Fourier transform on a product space can be decomposed into a sequence of Fourier transforms on each of the subspaces in the product. Suppose that $\mathcal{E} = \mathcal{E}_1 \times \mathcal{E}_2$, the product of two orthogonal subspaces $\mathcal{E}_1$ and $\mathcal{E}_2$ of respective dimensions d_1 and d_2 (where $d_1, d_2 \geq 1$ and $d = d_1 + d_2 \geq 2$). We define the *Fourier transform inside* $\mathcal{E}_1$ as the transformation $\mathcal{F}_{\mathcal{E}_1}$ of functions $F : \mathcal{E} \to \mathbb{C}$ which applies to every $\mathcal{E}_2$-section $F^{\mathbf{x}_2}$ the Fourier transform for functions $\mathcal{E}_1 \to \mathbb{C}$; thus for every $\mathbf{x}_2 \in \mathcal{E}_2$ we have $\left[\mathcal{F}_{\mathcal{E}_1}(F)\right]^{\mathbf{x}_2} = \left[F^{\mathbf{x}_2}\right]^\wedge$, in other words, the function $\mathcal{E}_1 \to \mathbb{C} : \mathbf{x}_1 \mapsto \mathcal{F}_{\mathcal{E}_1}(F)(\mathbf{x}_1, \mathbf{x}_2)$ is the Fourier transform of the function $\mathcal{E}_1 \to \mathbb{C} : \mathbf{x}_1 \mapsto F(\mathbf{x}_1, \mathbf{x}_2)$. For example if F is integrable, for every $\mathbf{u_1} \in \mathcal{E}_1$ and $\mathbf{x}_2 \in \mathcal{E}_2$ we have

$$\mathcal{F}_{\mathcal{E}_1}(F)(\mathbf{u}_1, \mathbf{x}_2) = \int_{\mathcal{E}_1} F(\mathbf{x_1}, \mathbf{x}_2) \exp[-2\pi i\, \mathbf{u}_1 \cdot \mathbf{x_1}]\, d\mathbf{x}_1.$$

We define similarly $\mathcal{F}_{\mathcal{E}_2}$, the *Fourier transform inside* $\mathcal{E}_2$, by $\left[\mathcal{F}_{\mathcal{E}_2}(F)\right]_{\mathbf{x}_1} = \left[F_{\mathbf{x}_1}\right]^\wedge$ for all $\mathbf{x}_1 \in \mathcal{E}_1$.

— **Decomposability:** *Let $\mathcal{E} = \mathcal{E}_1 \times \mathcal{E}_2$, the product of two orthogonal subspaces $\mathcal{E}_1$ and $\mathcal{E}_2$. For any function $F : \mathcal{E} \to \mathbb{C}$, let $\mathcal{F}_{\mathcal{E}_1}(F)$ be its Fourier transform inside $\mathcal{E}_1$ defined by $\left[\mathcal{F}_{\mathcal{E}_1}(F)\right]^{\mathbf{x}_2} = \left[F^{\mathbf{x}_2}\right]^\wedge$ for all $\mathbf{x}_2 \in \mathcal{E}_2$; define similarly $\mathcal{F}_{\mathcal{E}_2}(F)$, its Fourier transform inside $\mathcal{E}_2$, by $\left[\mathcal{F}_{\mathcal{E}_2}(F)\right]_{\mathbf{x}_1} = \left[F_{\mathbf{x}_1}\right]^\wedge$ for all $\mathbf{x}_1 \in \mathcal{E}_1$. Then $\widehat{F} = \mathcal{F}_{\mathcal{E}_1}(\mathcal{F}_{\mathcal{E}_2}(F)) = \mathcal{F}_{\mathcal{E}_2}(\mathcal{F}_{\mathcal{E}_1}(F))$. In particular, if $F(\mathbf{x}_1, \mathbf{x}_2) = F_1(\mathbf{x}_1)F_2(\mathbf{x}_2)$ for two functions $F_1 : \mathcal{E}_1 \to \mathbb{C}$ and $F_2 : \mathcal{E}_2 \to \mathbb{C}$, then $\widehat{F}(\mathbf{u}_1, \mathbf{u}_2) = \widehat{F}_1(\mathbf{u}_1)\widehat{F}_2(\mathbf{u}_2)$.*

For F integrable, this is shown by a straightforward application of Fubini's theorem. For F square-integrable, the result follows because $L^1 \cap L^2$ is dense in L^2, and both $\mathcal{F}_{\mathcal{E}_1}$ and $\mathcal{F}_{\mathcal{E}_2}$ preserve the L^2 norm; alternately, one can use the multiplication formula with $G(\mathbf{x}_1, \mathbf{x}_2) = G_1(\mathbf{x}_1)G_2(\mathbf{x}_2)$, where G_1 and G_2 are in $L^1 \cap L^2$. A consequence of this property is that a multidimensional Fourier transform can be decomposed into a series of one-dimensional Fourier transforms.

Finally, write x_n for the coordinate of the vector $\mathbf{x}$ in the direction of a unit vector $\mathbf{n}$ (thus $x_n = \mathbf{x} \cdot \mathbf{n}$). The following will be used in the sequel:

— **$\mathbf{L^1}$ Fourier derivative formula:** *Let G be given by $G(\mathbf{x}) = x_n F(\mathbf{x})$; if F and G are both in L^1, then $\widehat{F}$ is derivable in x_n, and $\partial/\partial x_n \widehat{F} = -2\pi i \widehat{G}$.*

We deduce the first consequence of the Riemann-Lebesgue theorem and the Fourier inversion formula:

Lemma 3.2. *Let F be in $L^1 \cup L^2$ and such that $\widehat{F}$ is integrable. Define F_u by*

$$F_u(\mathbf{x}) = \int_{\mathcal{E}} \widehat{F}(\mathbf{u}) \exp[2\pi i\, \mathbf{u} \cdot \mathbf{x}]\, d\mathbf{u} \tag{3.11}$$

for $\mathbf{x} \in \mathcal{E}$. Then $F \equiv F_u$, F_u is uniformly continuous and bounded on $\mathcal{E}$, and it vanishes at infinity.

In the remainder of this paper, we will use the notation F_u introduced here.

Let us consider in particular *real-valued* functions. When F is real-valued, (3.7) gives $\overline{\widehat{F}(\mathbf{u})} = \widehat{F}(-\mathbf{u})$ for any $\mathbf{u} \in \mathcal{E}$, in other words $\widehat{F}$ is conjugate-symmetric. In particular, $\widehat{F}$ is even-symmetric (resp., odd-symmetric), if and only if $\widehat{F}$ is real (resp., imaginary). As $\widehat{F}$ is conjugate-symmetric, $F^{\mathcal{A}}$ is even-symmetric and F^{Φ} is odd-symmetric. Moreover, for every unit vector $\mathbf{n}$, $\widehat{F}$ is (up to equivalence by $\equiv$) determined by its restriction $\mathrm{pos}_{\mathbf{n}} \cdot \widehat{F}$ to $\mathcal{E}^+_{\mathbf{n}}$. When $\widehat{F}$ is in $L^1 \cup L^2$, combining (3.6), (3.8), and the fact that $\big(\mathrm{neg}_{\mathbf{n}} \cdot \widehat{F}\big)_\rho = \overline{\mathrm{pos}_{\mathbf{n}} \cdot \widehat{F}}$, we get

$$\big(\mathrm{neg}_{\mathbf{n}} \cdot \widehat{F}\big)^\vee = \overline{\big(\mathrm{pos}_{\mathbf{n}} \cdot \widehat{F}\big)^\vee},$$

while the Fourier inversion formula gives

$$\big(\mathrm{neg}_{\mathbf{n}} \cdot \widehat{F}\big)^\vee + \big(\mathrm{pos}_{\mathbf{n}} \cdot \widehat{F}\big)^\vee \equiv \widehat{F}^\vee \equiv F;$$

as F is real-valued, both equations combined give:

$$F \equiv 2\Re\big[\big(\mathrm{neg}_{\mathbf{n}} \cdot \widehat{F}\big)^\vee\big] \equiv 2\Re\big[\big(\mathrm{pos}_{\mathbf{n}} \cdot \widehat{F}\big)^\vee\big]. \tag{3.12}$$

When $\widehat{F}$ is integrable, (3.11) and (3.12) give

$$\begin{aligned} F_u(\mathbf{x}) &= \int_{\mathcal{E}} \widehat{F}(\mathbf{u}) \exp[2\pi i\, \mathbf{u} \cdot \mathbf{x}]\, d\mathbf{u} = \int_{\mathcal{E}} F^{\mathcal{A}}(\mathbf{u}) \cos\big[2\pi \mathbf{u} \cdot \mathbf{x} + F^{\Phi}(\mathbf{u})\big]\, d\mathbf{u} \\ &= 2\Re\Big[\int_{\mathcal{E}^+_{\mathbf{n}}} \widehat{F}(\mathbf{u}) \exp[2\pi i\, \mathbf{u} \cdot \mathbf{x}]\, d\mathbf{u}\Big] = 2 \int_{\mathcal{E}^+_{\mathbf{n}}} F^{\mathcal{A}}(\mathbf{u}) \cos\big[2\pi \mathbf{u} \cdot \mathbf{x} + F^{\Phi}(\mathbf{u})\big]\, d\mathbf{u} \end{aligned} \tag{3.13}$$

for any half-space $\mathcal{E}^+_{\mathbf{n}}$.

Given a function F and a point $\mathbf{p} \in \mathcal{E}$, we call the *Fourier transform of F at $\mathbf{p}$* the Fourier transform of the function resulting from F when the origin of $\mathcal{E}$ is shifted to $\mathbf{p}$; in other words it is

$$[\tau_{-\mathbf{p}}(F)]^\wedge = \mathrm{cis}_{\mathbf{p}} \cdot \widehat{F}.$$

Here, the Fourier amplitudes are those of F, but the Fourier phases are advanced proportionally to $2\pi\, \mathbf{p}$:

$$\big[\tau_{-\mathbf{p}}(F)\big]^\Phi(\mathbf{u}) = F^\Phi(\mathbf{u}) + 2\pi\, \mathbf{p} \cdot \mathbf{u}.$$

Let us write $F^\Phi(\mathbf{u}, \mathbf{p})$ for the Fourier phase of F at $\mathbf{p}$ for the frequency $\mathbf{u}$, in other words

$$F^\Phi(\mathbf{u}, \mathbf{p}) = F^\Phi(\mathbf{u}) + 2\pi\, \mathbf{p} \cdot \mathbf{u}. \tag{3.14}$$

Then, for F real-valued and $\widehat{F}$ integrable, (3.13) can be written:

$$F_u(\mathbf{x}) = \int_{\mathcal{E}} F^{\mathcal{A}}(\mathbf{u}) \cos\big[F^\Phi(\mathbf{u}, \mathbf{p})\big]\, d\mathbf{u} = 2 \int_{\mathcal{E}^+_{\mathbf{n}}} F^{\mathcal{A}}(\mathbf{u}) \cos\big[F^\Phi(\mathbf{u}, \mathbf{p})\big]\, d\mathbf{u}. \tag{3.15}$$

Let us briefly recall the definition of the Fourier transform for tempered distributions. We refer to [46], Section 1.3, or to [48], Sections 8.1 and 8.5, for further details. A *Schwartz function* is a C^∞ function such that itself and all its derivatives, multiplied by any polynomial, remain bounded. A *tempered distribution* is a continuous linear functional on the space of Schwartz functions; the tempered distribution ψ associates to a Schwartz function S the value $\langle \psi, S \rangle$. A *tempered function* is a function F such that there is some $N \geq 0$ for which the function $\mathbf{x} \mapsto (1+|\mathbf{x}|)^{-N} F(\mathbf{x})$ is integrable; for examples functions in L^p $(1 \leq p \leq \infty)$ are tempered. A tempered function F induces a tempered distribution F^{dist} given by $\langle F^{\mathrm{dist}}, S\rangle = \int_{\mathcal{E}} FS$; we generally identify F with F^{dist}, and set thus $\langle F, S\rangle = \int_{\mathcal{E}} FS$.

The Fourier transform $\widehat{\psi}$ of a tempered distribution ψ is the tempered distribution given by $\langle \widehat{\psi}, S\rangle = \langle \psi, \widehat{S}\rangle$. For a function F in $L^1 \cup L^2$, the definition of $\widehat{F}$ in the sense of tempered distributions coincides with that given above, thanks to the multiplication formula $\int_{\mathcal{E}} \widehat{F} S = \int_{\mathcal{E}} F \widehat{S}$ which holds for any Schwartz function S.

The above properties given for the Fourier transform in L^1 or L^2 have a counterpart for tempered distributions.

3.3. *Constant zero Fourier phase*

When we want to obtain results concerning the pointwise values of a function F, this can be achieved by considering $\widehat{F}$ if the latter is integrable. This was for example the case in Lemma 3.2, and we will continue in this way in order to show that functions with constant zero Fourier phase have generally an absolute maximum at the origin.

LEMMA 3.3. *Let $F \not\equiv 0$ be integrable and having real non-negative values. Then:*

(*a*) *For $\mathbf{t} \neq \mathbf{0}$, $|\widehat{F}(\mathbf{t})| < \widehat{F}(\mathbf{0})$.*

(*b*) *Given G such that $|G(\mathbf{x})| \leq F(\mathbf{x})$ for all $\mathbf{x} \in \mathcal{E}$, then for $\mathbf{t} \neq \mathbf{0}$, either $|\widehat{G}(\mathbf{t})| < \widehat{F}(\mathbf{0})$, or there is some φ such that $G \equiv e^{i\varphi} \mathrm{cis}_{\mathbf{t}} \cdot F$, that is $\widehat{G} = e^{i\varphi} \tau_{\mathbf{t}}(\widehat{F})$.*

PROOF. For $\mathbf{t} \neq \mathbf{0}$ we have

$$|\widehat{G}(\mathbf{t})| = \Big|\int_{\mathcal{E}} \mathrm{cis}_{-\mathbf{t}} \cdot G\Big| \leq \int_{\mathcal{E}} |\mathrm{cis}_{-\mathbf{t}} \cdot G| = \int_{\mathcal{E}} |G| \leq \int_{E} F = \widehat{F}(\mathbf{0}).$$

Thus $|\widehat{G}(\mathbf{t})| \leq \widehat{F}(\mathbf{0})$, and the equality holds if and only if $|G| \equiv F$ and $\mathrm{cis}_{-\mathbf{t}} \cdot G$ has constant complex argument a.e. on $\mathcal{E}$, in other words $\mathrm{cis}_{-\mathbf{t}} \cdot G \equiv e^{i\varphi} F$, that is $G \equiv e^{i\varphi} \mathrm{cis}_{\mathbf{t}} \cdot F$. By the uniqueness property in L^1 and (3.10), the latter is equivalent to $\widehat{G} = [e^{i\varphi} \mathrm{cis}_{\mathbf{t}} \cdot F]^{\wedge} = e^{i\varphi} \tau_{\mathbf{t}}(\widehat{F})$. Thus ($b$) holds.

Taking $G = F$, as $e^{i\varphi} \mathrm{cis}_{\mathbf{t}}(\mathbf{x}) \neq 1$ almost everywhere, the equality $F \equiv e^{i\varphi} \mathrm{cis}_{\mathbf{t}} \cdot F$ implies that $F \equiv 0$. Thus $F \not\equiv 0$ gives $|\widehat{F}(\mathbf{t})| < \widehat{F}(\mathbf{0})$, and ($a$) holds. ∎

COROLLARY 3.4. *Let F be in $L^1 \cup L^2$ such that $\widehat{F}$ is integrable and has real non-negative values. Then:*

(*a*) *For $\mathbf{t} \neq \mathbf{0}$, $|F_u(\mathbf{t})| < F_u(\mathbf{0})$.*

(*b*) *Given G such that $|\widehat{G}(\mathbf{x})| \leq \widehat{F}(\mathbf{x})$ for all $\mathbf{x} \in \mathcal{E}$, then for $\mathbf{t} \neq \mathbf{0}$, either $|G_u(\mathbf{t})| < F_u(\mathbf{0})$, or there is some φ such that $G_u = e^{i\varphi} \tau_{\mathbf{t}}(F_u)$.*

PROOF. By Lemma 3.2, $\widehat{\widehat{F}} = (F_u)_\rho$ and $\widehat{\widehat{G}} = (G_u)_\rho$. By Lemma 3.3, the result follows for $(F_u)_\rho$ and $(G_u)_\rho$, and it is then easily seen that this means that it holds for F_u and G_u. ∎

Thus a real-valued function F with constant zero Fourier phase has an absolute maximum at the origin, provided that its Fourier transform $\widehat{F}$ is integrable. We show below that this happens if F is bounded in a neighbourhood of the origin.

PROPOSITION 3.5. *Let F be in $L^1 \cup L^2$ such that F is a.e. bounded in a neighbourhood of the origin and $\widehat{F}$ has real non-negative values. Then $\widehat{F}$ is integrable.*

Proof. For $r > 0$, let V_r be the set of all $\mathbf{x} \in \mathcal{E}$ having $|\mathbf{x}| < r$. There are $m, r > 0$ such that for almost all $\mathbf{x} \in V_r$ we have $|F(\mathbf{x})| < m$. Let the function G be given by $G(\mathbf{x}) = \exp[-\pi\,|\mathbf{x}|^2]$; for every integer n, define the two functions H_n and K_n as follows:

$$H_n(\mathbf{x}) = G(n^{-1}\,\mathbf{x}) = \exp\big[-\pi\,|\mathbf{x}|^2/n^2\big],$$
$$K_n(\mathbf{x}) = n^d G(n\,\mathbf{x}) = n^d \exp\big[-\pi n^2\,|\mathbf{x}|^2\big].$$

It is well-known (see [46]) that $G = \widehat{G}$, and so that $K_n = \widehat{H}_n$ and $H_n = \widehat{K}_n$. As H_n is in $L^1 \cap L^2$, the multiplication formula implies that FK_n and $\widehat{F}H_n$ are integrable, and $\int_{\mathcal{E}} FK_n = \int_{\mathcal{E}} \widehat{F}H_n$. As $|F(\mathbf{x})| < m$ a.e. in V_r, we get

$$\Big|\int_{V_r} FK_n\Big| \leq \int_{V_r} |FK_n| \leq m \cdot \int_{V_r} |K_n| \leq m \cdot \int_{\mathcal{E}} |K_n| = m \cdot \int_{\mathcal{E}} K_n = m \cdot \widehat{K}_n(\mathbf{0}) = m \cdot H_n(\mathbf{0}) = m.$$

Take an integer n_0 such that $2\pi r^2 n_0^2 > d$; then the calculation of $\partial K_n/\partial n$ shows that for $n \geq n_0$ and $|\mathbf{x}| \geq r$ we have $0 \leq K_n(\mathbf{x}) \leq K_{n_0}(\mathbf{x})$. Hence:

$$\Big|\int_{\mathcal{E}\setminus V_r} FK_n\Big| \leq \int_{\mathcal{E}\setminus V_r} |FK_n| \leq \int_{\mathcal{E}\setminus V_r} |F|K_{n_0} \leq \int_{\mathcal{E}} |F|K_{n_0}.$$

Combining both inequations above, for $n \geq n_0$ we have

$$\Big|\int_{\mathcal{E}} FK_n\Big| \leq m + \int_{\mathcal{E}} |F|K_{n_0}. \tag{3.16}$$

Hence $\int_{\mathcal{E}} FK_n$ remains bounded for $n \to \infty$. The functions H_n are positive and increase with n. As $\widehat{F}$ has non-negative real values, the functions $\widehat{F}H_n$ are non-negative and increase with n; moreover for $n \to \infty$, $H_n(\mathbf{x}) \to 1$, and so $FH_n \to F$. Hence by the Lebesgue monotone convergence theorem, the multiplication formula, and (3.16), we obtain

$$\int_{\mathcal{E}} \widehat{F} = \int_{\mathcal{E}} \lim_{n\to\infty} \widehat{F}H_n = \lim_{n\to\infty} \int_{\mathcal{E}} F\widehat{H}_n = \lim_{n\to\infty} \int_{\mathcal{E}} FK_n < \infty,$$

that is $\widehat{F}$ is integrable. ∎

Note that the fact that F belongs to $L^1 \cup L^2$ is not crucial in our proof; the latter can be extended to the case where F is a tempered function whose Fourier transform (in the tempered distribution sense) is also a tempered function. On the other hand the requirement that F is a.e. bounded in a neighbourhood of the origin is necessary (in fact F_u is bounded everywhere by the Riemann-Lebesgue Lemma).

We conclude that for a real-valued function F in $L^1 \cup L^2$ such that F^Φ is constant zero and F is a.e. bounded in a neighbourhood of the origin, $\widehat{F}$ is integrable and F_u has an absolute maximum at the origin.

3.4. The Hilbert transform and phase quadrature

The Hilbert transform is defined for functions $\mathbb{R} \to \mathbb{C}$, and it plays an important role in the one-dimensional phase congruence model. In the earliest version of the model proposed by Morrone and Owens [8], the energy function was defined as the sum of squares of the one-dimensional signal and its Hilbert transform. In the second version proposed afterwards by Morrone and Burr [9], this energy function was defined as the sum of squares of the convolutions of the one-dimensional signal with two filters forming a Hilbert transform pair. In [15] we have highlighted the hidden mathematical asumptions underlying both versions of the phase congruence model. As we will deal with oriented edges in two-dimensional images, a muldimensional model of oriented energy functions must be built. Hence we will extend the Hilbert transform in an anisotropic way to functions $\mathbb{R}^d \to \mathbb{C}$; this will be the *directional Hilbert transform*, which will be defined w.r.t. a unit vector $\mathbf{n}$.

Two functions $f, g : \mathbb{R} \to \mathbb{C}$ are said to be *in quadrature*, or to form a *quadrature pair* if we have $\widehat{g}(\nu) = -i\,\mathrm{sgn}(\nu)\widehat{f}(\nu)$ for almost all $\nu \in \mathbb{R}$; when f and g are real-valued, this means that they have the same Fourier amplitude for positive frequencies, but that the phase of g is shifted by $-\pi/2$. Note that the order of f and g in this relation does not really matter, since the ordered quadrature pair (f, g) implies the ordered quadrature pair $(g, -f)$, and that f and g are generally squared in an energy function.

Let $f : \mathbb{R} \to \mathbb{C}$ be in L^p, where $1 \leq p < \infty$; its *Hilbert transform* $\mathcal{H}[f]$ can be defined in two ways:

$$\begin{aligned} \mathcal{H}[f](x) &= \lim_{\varepsilon\to 0} \frac{1}{\pi} \int_{|t|\geq\varepsilon} \frac{f(x-t)}{t}\,dt, \\ &= \lim_{\varepsilon\to 0} \frac{1}{\pi} \int_{\mathbb{R}} f(x-t)\frac{t}{t^2+\varepsilon^2}\,dt, \end{aligned} \qquad (\varepsilon > 0). \tag{3.17}$$

In fact (see [46], p. 218, or [51], p. 255), the two expressions coincide on the Lebesgue set of f; thus they are equal almost everywhere, and in particular for all points at which f is continuous. The Hilbert transform is linear, translation-invariant ($g(x) = f(x-h)$ implies $\mathcal{H}[g](x) = \mathcal{H}[f](x-h)$), scale-invariant (for $a > 0$, $g(x) = f(x/a)$ implies $\mathcal{H}[g](x) = \mathcal{H}[f](x/a)$), and antisymmetric ($\mathcal{H}[f_\rho] = -\mathcal{H}[f]_\rho$). It has two important properties. First, the following result due to M. Riesz (see [51], p. 287, or [46], p. 188):

— **$\mathbf{L^p}$ stability:** *For $1 < p < \infty$, there is some constant A_p such that for every f in L^p, $\mathcal{H}[f]$ is also in L^p, with $\|\mathcal{H}[f]\|_p \leq A_p \|f\|_p$.*

Pichorides [53] found the best value of A_p for *real-valued* functions in L^p: $\tan(\pi/2p)$ for $1 < p \leq 2$, and $\cot(\pi/2p)$ for $2 \leq p < \infty$. Next, the following is well-known (see [51], p. 257):

— **Quadrature formula in $\mathbf{L^2}$:** *For any f in L^2, $\mathcal{H}[f]^\wedge(\nu) = -i\,\mathrm{sgn}(\nu)\widehat{f}(\nu)$ for almost all $\nu \in \mathbb{R}$.*

Thus Hilbert transform pairs of functions in L^2 are in quadrature; we have then $\|\mathcal{H}[f]\|_2 = \|f\|_2$ by the Plancherel theorem, and this agrees with the Pichorides bound for real-valued functions: $A_2 = \tan(\pi/4) = \cot(\pi/4) = 1$. We get also $\mathcal{H}\big[\mathcal{H}[f]\big] \equiv -f$, since both have the same Fourier transform; this is called the *skewed symmetry* of the Hilbert transform. In fact for $1 < p < \infty$ the skewed symmetry is also valid for functions in L^p, while the quadrature formula holds for functions in $L^1 \cap L^p$. This will be shown afterwards in the more general framework of the directional Hilbert transform for functions on $\mathbb{R}^d$.

For f in L^1, one can only prove results such as the following: for every $t > 0$, the Lebesgue measure of the set of all $x \in \mathbb{R}$ such that $|\mathcal{H}[f](x)| > t$ is at most $e \cdot \|f\|_1/t$ (see [46], p. 188). In fact the Pichorides bound gives $A_p \to \infty$ for $p \to 1$. We will show later on how to obtain Hilbert transform pairs in L^1.

Let us now consider functions defined on the space $\mathcal{E} = \mathbb{R}^d$, where $d \geq 1$. We will show how to extend the notions of quadrature and Hilbert transform defined for $d = 1$ to the multidimensional case. Let $\mathbf{n}$ be any unit vector. Two functions $F, G : \mathcal{E} \to \mathbb{C}$ are said to be *in* $\mathbf{n}$-*quadrature*, or to form a $\mathbf{n}$-*quadrature pair* if we have $\widehat{G}(\mathbf{u}) = -i\,\mathrm{sgn}_{\mathbf{n}}(\mathbf{u})\widehat{F}(\mathbf{u})$ for almost all $\mathbf{u} \in \mathcal{E}$; when F and G are real-valued, this means that they have the same Fourier amplitude, but that the phase of G is shifted by $-\pi/2$ for frequencies in $\mathcal{E}_{\mathbf{n}}^{+}$ (or by $\pi/2$ for frequencies in $\mathcal{E}_{\mathbf{n}}^{-}$). Note that the following four equalities are equivalent:

$$\begin{aligned}
\widehat{G}(\mathbf{u}) &= -i\,\mathrm{sgn}_{\mathbf{n}}(\mathbf{u})\widehat{F}(\mathbf{u});\\
\widehat{F}(\mathbf{u}) &= i\,\mathrm{sgn}_{\mathbf{n}}(\mathbf{u})\widehat{G}(\mathbf{u});\\
\widehat{G}(\mathbf{u}) &= i\,\mathrm{sgn}_{-\mathbf{n}}(\mathbf{u})\widehat{F}(\mathbf{u});\\
\widehat{F}(\mathbf{u}) &= -i\,\mathrm{sgn}_{-\mathbf{n}}(\mathbf{u})\widehat{G}(\mathbf{u}).
\end{aligned}$$

Hence the order of F and G in such a relation does not really matter; anyway F and G will be squared in the energy function.

Let $\mathcal{L}_{\mathbf{n}}$ be the vector space generated by $\mathbf{n}$. We can consider $\mathcal{E}$ as the product of the two orthogonal subspaces $\mathcal{E}_{\mathbf{n}}^{0}$ and $\mathcal{L}_{\mathbf{n}}$, the latter being identified with $\mathbb{R}$ (every $t \in \mathbb{R}$ corresponding to $t\mathbf{n} \in \mathcal{L}_{\mathbf{n}}$). Thus every $\mathbf{x} \in \mathcal{E}$ can be written in a unique way as a pair $(\mathbf{y}, t) \in \mathcal{E}_{\mathbf{n}}^{0} \times \mathbb{R}$; in fact $t = \mathbf{x}\cdot\mathbf{n}$ and $\mathbf{y} = \mathbf{x} - t\mathbf{n}$. We illustrate this in Figure 6 for $d = 2$. Recall that for a function $F : \mathcal{E} \to \mathbb{C}$, every $\mathbf{y} \in \mathcal{E}_{\mathbf{n}}^{0}$ gives the $\mathcal{E}_{\mathbf{n}}^{0}$-section $F_{\mathbf{y}} : \mathbb{R} \to \mathbb{C} : t \mapsto F(\mathbf{y}, t)$, while every $t \in \mathbb{R}$ gives the $\mathcal{L}_{\mathbf{n}}$-section $F^{t} : \mathcal{E}_{\mathbf{n}}^{0} \to \mathbb{C} : \mathbf{y} \mapsto F(\mathbf{y}, t)$. Then the $\mathbf{n}$-quadrature formula becomes: $\widehat{G}(\mathbf{y}, t) = -i\,\mathrm{sgn}(t)\widehat{F}(\mathbf{y}, t)$ for almost all $(\mathbf{y}, t) \in \mathcal{E}_{\mathbf{n}}^{0} \times \mathbb{R}$. The following result links $\mathbf{n}$-quadrature to quadrature for functions on $\mathbb{R}$ (here all functions are assumed to be integrable or square-integrable):

LEMMA 3.6. *For $F, G : \mathcal{E} \to \mathbb{C}$, F and G are in $\mathbf{n}$-quadrature if and only if for almost all $\mathbf{y} \in \mathcal{E}_{\mathbf{n}}^{0}$, $F_{\mathbf{y}}$ and $G_{\mathbf{y}}$ are in quadrature. In particular for every function $H : \mathcal{E}_{\mathbf{n}}^{0} \to \mathbb{C}$:*

- (i) *For $f, g : \mathbb{R} \to \mathbb{C}$ in quadrature, the functions $Hf : (\mathbf{y}, t) \mapsto H(\mathbf{y})f(t)$ and $Hg : (\mathbf{y}, t) \mapsto H(\mathbf{y})g(t)$ are in $\mathbf{n}$-quadrature.*
- (ii) *For $F, G : \mathcal{E} \to \mathbb{C}$ in $\mathbf{n}$-quadrature, the functions $HF : (\mathbf{y}, t) \mapsto H(\mathbf{y})F(\mathbf{y}, t)$ and $HG : (\mathbf{y}, t) \mapsto H(\mathbf{y})G(\mathbf{y}, t)$ are in $\mathbf{n}$-quadrature.*

PROOF. Define $\widetilde{F}$ and $\widetilde{G}$ as the respective Fourier transforms inside $\mathcal{L}_{\mathbf{n}}$ of F and G, in other words $\widetilde{F}_{\mathbf{y}} = \big[F_{\mathbf{y}}\big]^{\wedge}$ and $\widetilde{G}_{\mathbf{y}} = \big[G_{\mathbf{y}}\big]^{\wedge}$for all $\mathbf{y} \in \mathcal{E}_{\mathbf{n}}^{\mathbf{0}}$. By the decomposability property of the Fourier transform, $\widehat{F}$ and $\widehat{G}$ are the Fourier transforms inside $\mathcal{E}_{\mathbf{n}}^{0}$ of $\widetilde{F}$ and $\widetilde{G}$ respectively, that is $\widehat{F}^{t} = \big[\widetilde{F}^{t}\big]^{\wedge}$ and $\widehat{G}^{t} = \big[\widetilde{G}^{t}\big]^{\wedge}$ for every $t \in \mathbb{R}$. Now since $\big[\widetilde{F}^{t}\big]^{\wedge}$ is the Fourier transform of $\widetilde{F}^{t}$ for variables in $\mathcal{E}_{\mathbf{n}}^{0}$, w.r.t. which $\mathrm{sgn}(t)$ is a constant, we have $-i\,\mathrm{sgn}(t)\big[\widetilde{F}^{t}\big]^{\wedge} = \big[-i\,\mathrm{sgn}(t)\widetilde{F}^{t}\big]^{\wedge}$; this gives thus $-i\,\mathrm{sgn}(t)\widehat{F}^{t} = \big[-i\,\mathrm{sgn}(t)\widetilde{F}^{t}\big]^{\wedge}$ for all $t \in \mathbb{R}$. Hence the following statements are equivalent:

- For almost all $(\mathbf{y}, t) \in \mathcal{E}_{\mathbf{n}}^{0} \times \mathbb{R}$, $\widehat{G}(\mathbf{y}, t) = -i\,\mathrm{sgn}(t)\widehat{F}(\mathbf{y}, t)$.
- For almost all $t \in \mathbb{R}$, $\big[\widetilde{G}^{t}\big]^{\wedge} = \widehat{G}^{t} \equiv -i\,\mathrm{sgn}(t)\widehat{F}^{t} = \big[-i\,\mathrm{sgn}(t)\widetilde{F}^{t}\big]^{\wedge}$.
- For almost all $t \in \mathbb{R}$, $\widetilde{G}^{t} \equiv -i\,\mathrm{sgn}(t)\widetilde{F}^{t}$.
- For almost all $(\mathbf{y}, t) \in \mathcal{E}_{\mathbf{n}}^{0} \times \mathbb{R}$, we have $\big[G_{\mathbf{y}}\big]^{\wedge}(t) = \widetilde{G}_{\mathbf{y}}(t) = \widetilde{G}^{t}(\mathbf{y}) = -i\,\mathrm{sgn}(t)\widetilde{F}^{t}(\mathbf{y}) = -i\,\mathrm{sgn}(t)\widetilde{F}_{\mathbf{y}}(t) = -i\,\mathrm{sgn}(t)\big[F_{\mathbf{y}}\big]^{\wedge}(t)$.

Hence F and G are in $\mathbf{n}$-quadrature if and only if for almost all $\mathbf{y} \in \mathcal{E}_{\mathbf{n}}^{0}$, $F_{\mathbf{y}}$ and $G_{\mathbf{y}}$ are in quadrature.

Take f, g defined on $\mathbb{R}$ and H defined on $\mathcal{E}_{\mathbf{n}}^{0}$. By the decomposability of the Fourier transform, the Fourier transforms of the functions $Hf : (\mathbf{y}, t) \mapsto H(\mathbf{y})f(t)$ and $Hg : (\mathbf{y}, t) \mapsto H(\mathbf{y})g(t)$ satisfy $[Hf]^{\wedge}(\mathbf{y}, t) = \widehat{H}(\mathbf{y})\widehat{f}(t)$ and $[Hg]^{\wedge}(\mathbf{y}, t) = \widehat{H}(\mathbf{y})\widehat{g}(t)$; hence $\widehat{g}(t) = -i\,\mathrm{sgn}(t)\widehat{f}(t)$ implies that $[Hg]^{\wedge}(\mathbf{y}, t) = -i\,\mathrm{sgn}(t)[Hf]^{\wedge}(\mathbf{y}, t)$, and ($i$) holds.

For F, G in $\mathbf{n}$-quadrature and H defined on $\mathcal{E}^0_{\mathbf{n}}$, we have $[HF]_{\mathbf{y}} = H(\mathbf{y})F_{\mathbf{y}}$ and $[HG]_{\mathbf{y}} = H(\mathbf{y})G_{\mathbf{y}}$, and as $H(\mathbf{y})$ is a constant w.r.t. $t \in \mathbb{R}$, $H(\mathbf{y})F_{\mathbf{y}}$ and $H(\mathbf{y})G_{\mathbf{y}}$ are in quadrature whenever $F_{\mathbf{y}}$ and $G_{\mathbf{y}}$ are in quadrature. Thus (ii) holds. ∎

The following is a multidimensional generalization of Proposition 3 of [15]:

PROPOSITION 3.7. *Let F, G be in L^1 and forming an $\mathbf{n}$-quadrature pair. Then for almost all $\mathbf{y} \in \mathcal{E}^0_{\mathbf{n}}$ we have $\int_{\mathbb{R}} F_{\mathbf{y}} = \int_{\mathbb{R}} G_{\mathbf{y}} = 0$; in particular $\int_{\mathcal{E}} F = \int_{\mathcal{E}} G$.*

PROOF. Since F and G are integrable and in $\mathbf{n}$-quadrature, for almost all $\mathbf{y} \in \mathcal{E}^0_{\mathbf{n}}$, $F_{\mathbf{y}}$ and $G_{\mathbf{y}}$ are integrable (by Fubini's theorem) and in quadrature (by Lemma 3.6). For such a $\mathbf{y}$ we have we have $\left[G_{\mathbf{y}}\right]^\wedge(\nu) = -i\left[F_{\mathbf{y}}\right]^\wedge(\nu)$ for almost all $\nu > 0$ and $\left[G_{\mathbf{y}}\right]^\wedge(\nu) = i\left[F_{\mathbf{y}}\right]^\wedge(\nu)$ for allmost all $\nu < 0$. By the Riemann-Lebesgue theorem, $\left[F_{\mathbf{y}}\right]^\wedge$ and $\left[G_{\mathbf{y}}\right]^\wedge$ are continuous; the continuity at 0 gives then both $\left[G_{\mathbf{y}}\right]^\wedge(0) = i\left[F_{\mathbf{y}}\right]^\wedge(0)$ and $\left[G_{\mathbf{y}}\right]^\wedge(0) = -i\left[F_{\mathbf{y}}\right]^\wedge(0)$, in other words $\int_{\mathbb{R}} G_{\mathbf{y}} = \left[G_{\mathbf{y}}\right]^\wedge(0) = 0 = \left[F_{\mathbf{y}}\right]^\wedge(0) = \int_{\mathbb{R}} F_{\mathbf{y}}$. By Fubini's theorem we have $\int_{\mathcal{E}} F = \int_{\mathcal{E}^0_{\mathbf{n}}} d\mathbf{y} \int_{\mathbb{R}} F_{\mathbf{y}} = \int_{\mathcal{E}^0_{\mathbf{n}}} d\mathbf{y}\, 0 = 0$ and similarly $\int_{\mathcal{E}} G = 0$. ∎

In electrical engineering parlance, "F and G have zero dc level". Write ξ_n for the function $\mathcal{E} \to \mathbb{C} : \mathbf{x} \mapsto x_n = \mathbf{x} \cdot \mathbf{n}$. The following generalizes Proposition 4 of [15] to multidimensional functions:

PROPOSITION 3.8. *Let F, G be in L^1, forming an $\mathbf{n}$-quadrature pair, and such that $\xi_n F$ and $\xi_n G$ are in L^1. Then $\xi_n F$ and $\xi_n G$ are in $\mathbf{n}$-quadrature.*

PROOF. By the L^1 Fourier derivative formula, we have $[\xi_n F]^\wedge = (-2\pi\, i)^{-1}\partial/\partial x_n \widehat{F}$ and $[\xi_n G]^\wedge = (-2\pi\, i)^{-1}\partial/\partial x_n \widehat{G}$. Since F and G are in $\mathbf{n}$-quadrature and $\widehat{F}$ and $\widehat{G}$ are continuous (by the Riemann-Lebesgue theorem), we have $\widehat{G}(\mathbf{u}) = -i\,\widehat{F}(\mathbf{u})$ for all $\mathbf{u} \in \mathcal{E}^+_{\mathbf{n}}$ and $\widehat{G}(\mathbf{u}) = i\,\widehat{F}(\mathbf{u})$ for all $\mathbf{u} \in \mathcal{E}^-_{\mathbf{n}}$; it follows then that $\partial/\partial x_n \widehat{G}(\mathbf{u}) = -i\,\partial/\partial x_n \widehat{F}(\mathbf{u})$ for all $\mathbf{u} \in \mathcal{E}^+_{\mathbf{n}}$ and $\partial/\partial x_n \widehat{G}(\mathbf{u}) = i\,\partial/\partial x_n \widehat{F}(\mathbf{u})$ for all $\mathbf{u} \in \mathcal{E}^-_{\mathbf{n}}$. Hence for all $\mathbf{u} \notin \mathcal{E}^0_{\mathbf{n}}$ we have

$$[\xi_n G]^\wedge(\mathbf{u}) = (-2\pi\, i)^{-1}\partial/\partial x_n \widehat{G}(\mathbf{u}) = -i\,\mathrm{sgn}_{\mathbf{n}}(\mathbf{u})(-2\pi\, i)^{-1}\partial/\partial x_n \widehat{F}(\mathbf{u}) = -i\,\mathrm{sgn}_{\mathbf{n}}(\mathbf{u})[\xi_n F]^\wedge(\mathbf{u}),$$

and so $\xi_n F$ and $\xi_n G$ are in $\mathbf{n}$-quadrature. ∎

Let us now define a multidimensional generalization of the Hilbert transform which will produce pairs of functions in $\mathbf{n}$-quadrature. For $F : \mathcal{E} \to \mathbb{C}$, its *Hilbert transform in the direction* $\mathbf{n}$, or $\mathbf{n}$-*directional Hilbert transform*, is the function $\mathcal{H}_{\mathbf{n}}[F]$ defined by

$$\big(\mathcal{H}_{\mathbf{n}}[F]\big)_{\mathbf{y}} = \mathcal{H}\big[F_{\mathbf{y}}\big] \qquad \text{for all} \quad \mathbf{y} \in \mathcal{E}^0_{\mathbf{n}}. \tag{3.18}$$

When F is in L^p (where $1 \leq p < \infty$), then (by Fubini's theorem) $F_{\mathbf{y}}$ is in L^p for almost all $\mathbf{y} \in \mathcal{E}^0_{\mathbf{n}}$, and for any such $\mathbf{y}$, $\mathcal{H}\big[F_{\mathbf{y}}\big]$ is defined; hence $\mathcal{H}_{\mathbf{n}}[F]$ is defined almost everywhere. By (3.17) and the definition of the $\mathcal{E}^0_{\mathbf{n}}$-sections (with $\mathbf{z} = (\mathbf{y}, x) = \mathbf{y} + x\mathbf{n}$), we have:

$$\begin{aligned}\mathcal{H}_{\mathbf{n}}[F](\mathbf{z}) &= \lim_{\varepsilon \to 0} \frac{1}{\pi} \int_{|t| \geq \varepsilon} \frac{F(\mathbf{z} - t\mathbf{n})}{t}\, dt, \\ &= \lim_{\varepsilon \to 0} \frac{1}{\pi} \int_{\mathbb{R}} F(\mathbf{z} - t\mathbf{n}) \frac{t}{t^2 + \varepsilon^2}\, dt,\end{aligned} \qquad (\varepsilon > 0). \tag{3.19}$$

The two formulas given here coincide almost everywhere. This definition (3.19) can be found in [52], p. 49. As the one-dimensional Hilbert transform, this multidimentional directional Hilbert transform is linear, translation-invariant ($G(\mathbf{x}) = F(\mathbf{x} - \mathbf{h})$ implies $\mathcal{H}_{\mathbf{n}}[G](\mathbf{x}) = \mathcal{H}_{\mathbf{n}}[F](\mathbf{x} - \mathbf{h})$), scale-invariant (for $a > 0$, $G(\mathbf{x}) = F(a^{-1}\mathbf{x})$ implies $\mathcal{H}_{\mathbf{n}}[G](\mathbf{x}) = \mathcal{H}_{\mathbf{n}}[F](a^{-1}\mathbf{x})$), and antisymmetric ($\mathcal{H}_{\mathbf{n}}[F_\rho] = -\mathcal{H}_{\mathbf{n}}[F]_\rho$). The Hilbert transform L^p stability (with the norm A_p), its quadrature formula and skewed symmetry in L^2 extend naturally to $\mathcal{H}_{\mathbf{n}}$:

LEMMA 3.9. *For every function $F : \mathcal{E} \to \mathbb{C}$:*

(*i*) *If F is in L^p for $1 < p < \infty$, then $\mathcal{H}_{\mathbf{n}}[F]$ is also in L^p, with $\|\mathcal{H}_{\mathbf{n}}[F]\|_p \leq A_p \|F\|_p$.*

(*ii*) *If F is in L^2, then $\mathcal{H}_{\mathbf{n}}[F]^\wedge(\mathbf{u}) = -i\,\mathrm{sgn}_{\mathbf{n}}(\mathbf{u})\widehat{F}(\mathbf{u})$ for almost all $\mathbf{u} \in \mathbb{R}$. Moreover, $\|\mathcal{H}_{\mathbf{n}}[F]\|_2 = \|F\|_2$ and $\mathcal{H}_{\mathbf{n}}\big[\mathcal{H}_{\mathbf{n}}[F]\big] \equiv -F$.*

For every function $G : \mathcal{E}^0_{\mathbf{n}} \to \mathbb{C}$:

(*iii*) *For $f : \mathbb{R} \to \mathbb{C}$ in L^p ($1 \leq p < \infty$), the functions $Gf : (\mathbf{y}, t) \mapsto G(\mathbf{y})f(t)$ and $G\mathcal{H}[f] : (\mathbf{y}, t) \mapsto G(\mathbf{y})\mathcal{H}[f](t)$ verify $G\mathcal{H}[f] = \mathcal{H}_{\mathbf{n}}[Gf]$.*

(*iv*) *For $F : \mathcal{E} \to \mathbb{C}$ in L^p ($1 \leq p < \infty$), the functions $GF : (\mathbf{y}, t) \mapsto G(\mathbf{y})F(\mathbf{y}, t)$ and $G\mathcal{H}_{\mathbf{n}}[F] : (\mathbf{y}, t) \mapsto G(\mathbf{y})\mathcal{H}_{\mathbf{n}}[F](\mathbf{y}, t)$ verify $G\mathcal{H}_{\mathbf{n}}[F] = \mathcal{H}_{\mathbf{n}}[GF]$.*

PROOF. (*i*) By the Hilbert transform L^p stability we have $\int_{\mathbb{R}}\big|\mathcal{H}\big[F_{\mathbf{y}}\big]\big|^p = \big\|\mathcal{H}\big[F_{\mathbf{y}}\big]\big\|_p^p \leq A_p^p\big\|F_{\mathbf{y}}\big\|_p^p = A_p^p \int_{\mathbb{R}}\big|F_{\mathbf{y}}\big|^p$. By Fubini's theorem, we get:

$$\int_{\mathcal{E}}\big|\mathcal{H}_{\mathbf{n}}[F]\big|^p = \int_{\mathcal{E}^0_{\mathbf{n}}} d\mathbf{y} \int_{\mathbb{R}}\big|\big(\mathcal{H}_{\mathbf{n}}[F]\big)_{\mathbf{y}}\big|^p = \int_{\mathcal{E}^0_{\mathbf{n}}} d\mathbf{y} \int_{\mathbb{R}}\big|\mathcal{H}\big[F_{\mathbf{y}}\big]\big|^p \leq \int_{\mathcal{E}^0_{\mathbf{n}}} d\mathbf{y}\, A_p^p \int_{\mathbb{R}}\big|F_{\mathbf{y}}\big|^p = A_p^p \int_{\mathcal{E}} |F|^p.$$

Hence $\|\mathcal{H}_{\mathbf{n}}[F]\|_p^p \leq A_p^p \|F\|_p^p$.

(*ii*) By Fubini's theorem, $F_{\mathbf{y}}$ is in L^2 for almost all $\mathbf{y} \in \mathcal{E}^0_{\mathbf{n}}$. For such a $\mathbf{y}$ we have by the Hilbert transform quadrature formula in L^2: $\big[\big(\mathcal{H}_{\mathbf{n}}[F]\big)_{\mathbf{y}}\big]^\wedge(t) = \mathcal{H}\big[F_{\mathbf{y}}\big]^\wedge(t) = -i\,\mathrm{sgn}(t) \cdot \big[F_{\mathbf{y}}\big]^\wedge(t)$ for almost all $t \in \mathbb{R}$. By Lemma 3.6, F and $\mathcal{H}_{\mathbf{n}}[F]$ are in $\mathbf{n}$-quadrature. It follows that $|\mathcal{H}_{\mathbf{n}}[F]^\wedge(\mathbf{x})| = |\widehat{F}(\mathbf{x})|$ almost everywhere, and so the Plancherel theorem gives $\|\mathcal{H}_{\mathbf{n}}[F]\|_2 = \|\mathcal{H}_{\mathbf{n}}[F]^\wedge\|_2 = \|\widehat{F}\|_2 = \|F\|_2$. Since $\mathcal{H}_{\mathbf{n}}[F]$ is in L^2, $\mathcal{H}_{\mathbf{n}}[F]$ and $\mathcal{H}_{\mathbf{n}}\big[\mathcal{H}_{\mathbf{n}}[F]\big]$ are in $\mathbf{n}$-quadrature. Thus $\mathcal{H}_{\mathbf{n}}\big[\mathcal{H}_{\mathbf{n}}[F]\big]^\wedge(\mathbf{u}) = -i\,\mathrm{sgn}_{\mathbf{n}}(\mathbf{u})\mathcal{H}_{\mathbf{n}}[F]^\wedge(\mathbf{u}) = (-i\,\mathrm{sgn}_{\mathbf{n}}(\mathbf{u}))^2\widehat{F}(\mathbf{u})$ a.e., and so $\mathcal{H}_{\mathbf{n}}\big[\mathcal{H}_{\mathbf{n}}[F]\big] \equiv -F$.

(*iii*) We have $[Gf]_{\mathbf{y}} = G(\mathbf{y})f$ and $\big[G\mathcal{H}[f]\big]_{\mathbf{y}} = G(\mathbf{y})\mathcal{H}[f]$, and as $G(\mathbf{y})$ is a constant w.r.t. $t \in \mathbb{R}$, we get $\mathcal{H}\big[G(\mathbf{y})f\big] = G(\mathbf{y})\mathcal{H}[f]$. Hence $\mathcal{H}\big[[Gf]_{\mathbf{y}}\big] = \big[G\mathcal{H}[f]\big]_{\mathbf{y}}$. The result follows then from the definition (3.18) of $\mathcal{H}_{\mathbf{n}}$.

(*iv*) We have $[GF]_{\mathbf{y}} = G(\mathbf{y})F_{\mathbf{y}}$ and $\big[G\mathcal{H}_{\mathbf{n}}[F]\big]_{\mathbf{y}} = G(\mathbf{y})\mathcal{H}_{\mathbf{n}}[F]_{\mathbf{y}} = G(\mathbf{y})\mathcal{H}\big[F_{\mathbf{y}}\big]$, and as $G(\mathbf{y})$ is a constant w.r.t. $t \in \mathbb{R}$, we get $\mathcal{H}\big[G(\mathbf{y})F_{\mathbf{y}}\big] = G(\mathbf{y})\mathcal{H}\big[F_{\mathbf{y}}\big]$. Hence $\mathcal{H}\big[[GF]_{\mathbf{y}}\big] = \big[G\mathcal{H}_{\mathbf{n}}[F]\big]_{\mathbf{y}}$. The result follows then from the definition (3.18) of $\mathcal{H}_{\mathbf{n}}$. ∎

The following result generalizes item (*ii*) of Lemma 3.9 to L^p for $p \neq 2$. In particular for $d = 1$ it applies also to the ordinary Hilbert transform, for which we found no trace of this property in the literature. We will use it in order to give a criterion for obtaining $\mathbf{n}$-directional Hilbert transform pairs in L^1 which are in $\mathbf{n}$-quadrature.

LEMMA 3.10. *Let $1 < p < \infty$. For F in L^p we have $\mathcal{H}_{\mathbf{n}}\big[\mathcal{H}_{\mathbf{n}}[F]\big] \equiv -F$. For F in $L^1 \cap L^p$, the Fourier transform of $\mathcal{H}_{\mathbf{n}}[F]$ (in the tempered distribution sense) is a function, and we have $\mathcal{H}_{\mathbf{n}}[F]^\wedge(\mathbf{u}) = -i\,\mathrm{sgn}_{\mathbf{n}}(\mathbf{u})\widehat{F}(\mathbf{u})$ a.e.*

PROOF. Take F in L^p. Since $L^2 \cap L^p$ is dense in L^p, for every $\varepsilon > 0$ there is some G in $L^2 \cap L^p$ such that $\|F - G\|_p < \varepsilon$. Item (*ii*) of Lemma 3.9 gives $\mathcal{H}_{\mathbf{n}}\big[\mathcal{H}_{\mathbf{n}}[G]\big] \equiv -G$, so that by item (*i*) of Lemma 3.9 we get:

$$\big\|\mathcal{H}_{\mathbf{n}}\big[\mathcal{H}_{\mathbf{n}}[F]\big] + F\big\|_p \leq \big\|\mathcal{H}_{\mathbf{n}}\big[\mathcal{H}_{\mathbf{n}}[F]\big] - \mathcal{H}_{\mathbf{n}}\big[\mathcal{H}_{\mathbf{n}}[G]\big]\big\|_p + \|F - G\|_p < A_p^2\varepsilon + \varepsilon,$$

leading to $\mathcal{H}_{\mathbf{n}}\big[\mathcal{H}_{\mathbf{n}}[F]\big] \equiv -F$.

Suppose now that F is in $L^1 \cap L^p$; then $\widehat{F}$ is a function, while $\mathcal{H}_{\mathbf{n}}[F]^\wedge$ is a priori a tempered distribution. Take this same G; we have thus $\|\mathcal{H}_{\mathbf{n}}[F - G]\|_p \leq A_p\|F - G\|_p < \varepsilon A_p$. Since $\mathcal{H}_{\mathbf{n}}[G]^\wedge \equiv$

$-i\,\mathrm{sgn}_{\mathbf{n}}\widehat{G}$ (by item (*ii*) of Lemma 3.9), and both the Hilbert and Fourier transforms are linear, the following equality holds in the tempered distribution sense:

$$\mathcal{H}_{\mathbf{n}}[F]^{\wedge} + i\,\mathrm{sgn}_{\mathbf{n}}\widehat{F} = \mathcal{H}_{\mathbf{n}}[F-G]^{\wedge} - i\,\mathrm{sgn}_{\mathbf{n}}\widehat{G} + i\,\mathrm{sgn}_{\mathbf{n}}\widehat{F};$$

thus for any Schwartz function S we have:

$$\begin{aligned}
&\langle \mathcal{H}_{\mathbf{n}}[F]^{\wedge} + i\,\mathrm{sgn}_{\mathbf{n}}\,\widehat{F}, S\rangle = \langle \big(\mathcal{H}_{\mathbf{n}}[F-G]\big)^{\wedge}, S\rangle - i\,\langle \mathrm{sgn}_{\mathbf{n}}\,\widehat{G}, S\rangle + i\,\langle \mathrm{sgn}_{\mathbf{n}}\,\widehat{F}, S\rangle \\
&= \langle \big(\mathcal{H}_{\mathbf{n}}[F-G]\big)^{\wedge}, S\rangle - i\int_{\mathcal{E}} \mathrm{sgn}_{\mathbf{n}}\,\widehat{G}S + i\int_{\mathcal{E}} \mathrm{sgn}_{\mathbf{n}}\,\widehat{F}S \qquad \text{(since } \widehat{F} \text{ and } \widehat{G} \text{ are functions)} \\
&= \langle \mathcal{H}_{\mathbf{n}}[F-G], \widehat{S}\rangle - i\int_{\mathcal{E}} G(\mathrm{sgn}_{\mathbf{n}}\,S)^{\wedge} + i\int_{\mathcal{E}} F(\mathrm{sgn}_{\mathbf{n}}\,S)^{\wedge} \qquad \text{(by the multiplication formula)} \\
&= \int_{\mathcal{E}} \mathcal{H}_{\mathbf{n}}[F-G]\widehat{S} + i\int_{\mathcal{E}} (F-G)(\mathrm{sgn}_{\mathbf{n}}\,S)^{\wedge} \qquad \text{(since } \mathcal{H}_{\mathbf{n}}[F-G] \text{ is a function)}
\end{aligned}$$

Now by Hölder's inequality we have (for $p' = p/(p-1)$):

$$\Big|\int_{\mathcal{E}} \mathcal{H}_{\mathbf{n}}[F-G]\widehat{S}\Big| \le \|\mathcal{H}_{\mathbf{n}}[F-G]\|_p \cdot \|\widehat{S}\|_{p'} < A_p\varepsilon\|\widehat{S}\|_{p'}$$

and

$$\Big|\int_{\mathcal{E}} (F-G)(\mathrm{sgn}_{\mathbf{n}}S)^{\wedge}\Big| \le \|F-G\|_p \cdot \|(\mathrm{sgn}_{\mathbf{n}}S)^{\wedge}\|_{p'} < \varepsilon\|(\mathrm{sgn}_{\mathbf{n}}S)^{\wedge}\|_{p'}.$$

Hence

$$\big|\langle \mathcal{H}_{\mathbf{n}}[F]^{\wedge} + i\,\mathrm{sgn}_{\mathbf{n}}\,\widehat{F}, S\rangle\big| < \varepsilon\big(A_p\|\widehat{S}\|_{p'} + \|(\mathrm{sgn}_{\mathbf{n}}S)^{\wedge}\|_{p'}\big)$$

for any $\varepsilon > 0$, from which we conclude that $\mathcal{H}_{\mathbf{n}}[F]^{\wedge} = -i\,\mathrm{sgn}_{\mathbf{n}}\,\widehat{F}$ in the tempered distribution sense; as $\widehat{F}$ is a function, this means that $\mathcal{H}_{\mathbf{n}}[F]^{\wedge} \equiv -i\,\mathrm{sgn}_{\mathbf{n}}\,\widehat{F}$. ∎

Note that the proof of the **n**-quadrature formula does not extend naturally to the case where F is not integrable, even if we assume that $\widehat{F}$ (in the tempered distribution sense) is a function; this is due to the fact that we cannot write $\langle \mathrm{sgn}_{\mathbf{n}}\,\widehat{F}, S\rangle = \langle \widehat{F}, \mathrm{sgn}_{\mathbf{n}}\,S\rangle$ in general, because $\mathrm{sgn}_{\mathbf{n}}\,S$ is not necessarily a Schwartz function (it can be discontinuous).

The following result gives us a criterion for obtaining Hilbert transform pairs in L^1 and in quadrature. Write ξ for the identity mapping $x \mapsto x$ on $\mathbb{R}$:

Lemma 3.11. *Let $1 < p, q < \infty$, and let f be a function in L^p such that the function ξf is in L^q. Then f is integrable and*

$$\mathcal{H}[\xi f] = \xi\mathcal{H}[f] - \frac{1}{\pi}\int_{\mathbb{R}} f. \tag{3.20}$$

Furthermore, the following four statements are equivalent:

(*i*) $\int_{\mathbb{R}} f = 0$.

(*ii*) $\mathcal{H}[\xi f] = \xi\mathcal{H}[f]$.

(*iii*) $\xi\mathcal{H}[f]$ *is in* L^q.

(*iv*) $\mathcal{H}[f]$ *is in* L^1.

Proof. The integrability of f follows from Lemma 3.1. The first expression of (3.17) implies that:

$$\begin{aligned}
\mathcal{H}[\xi f](x) &= \lim_{\varepsilon\to 0}\frac{1}{\pi}\int_{|t|\ge\varepsilon}\frac{(x-t)f(x-t)}{t}\,dt = \lim_{\varepsilon\to 0}\frac{1}{\pi}\Big(x\cdot\int_{|t|\ge\varepsilon}\frac{f(x-t)}{t}\,dt - \int_{|t|\ge\varepsilon}\frac{tf(x-t)}{t}\,dt\Big) \\
&= x\cdot\lim_{\varepsilon\to 0}\frac{1}{\pi}\int_{|t|\ge\varepsilon}\frac{f(x-t)}{t}\,dt - \lim_{\varepsilon\to 0}\frac{1}{\pi}\int_{|t|\ge\varepsilon} f(x-t)\,dt = x\mathcal{H}[f](x) - \frac{1}{\pi}\int_{\mathbb{R}} f.
\end{aligned}$$

Thus (3.20) holds.

Let us now show the equivalence between (i), (ii), (iii), and (iv). As f is in L^p and ξf is in L^q, by the Hilbert transform stability property, $\mathcal{H}[f]$ is in L^p and $\mathcal{H}[\xi f]$ is in L^q. As f is in $L^1 \cap L^p$, f and $\mathcal{H}[f]$ are in quadrature by Lemma 3.10.

If (i): $\int_{\mathbb{R}} f = 0$, then (3.20) gives (ii): $\mathcal{H}[\xi f] = \xi\mathcal{H}[f]$. As $\mathcal{H}[\xi f]$ is in L^q, (ii): $\mathcal{H}[\xi f] = \xi\mathcal{H}[f]$ implies that (iii): $\xi\mathcal{H}[f]$ is in L^q. If (iii): $\xi\mathcal{H}[f]$ is in L^q, as $\mathcal{H}[f]$ is in L^p, applying Lemma 3.1 to $\mathcal{H}[f]$, we obtain (iv): $\mathcal{H}[f]$ is in L^1. As f is in L^1, and f and $\mathcal{H}[f]$ are in quadrature, if (iv): $\mathcal{H}[f]$ is in L^1, then by Proposition 3.7 (in fact, by Proposition 3 of [15]) we deduce that (i): $\int_{\mathbb{R}} f = 0$. ∎

Thus f and $\mathcal{H}[f]$ will form a quadrature pair in L^1. Note that by Proposition 3.7 we have then also $\int_{\mathbb{R}} \mathcal{H}[f] = 0$.

Let us give two examples where this result can be used. First, if f is bounded in a neighbourhood V of the origin, and if for some $\varepsilon > 0$ we have $|f(x)| \leq 1/x^{1+\varepsilon}$ outside V, then f belongs to L^p for $1 < p < \infty$, and ξf is in L^q for $\max(1, 1/\varepsilon) < q < \infty$. Second, if f has a bounded support V (that is, $f(x) = 0$ for $x \notin V$), and if for $0 < \delta < 1$ we have $|f(x)| \leq 1/x^\delta$ for $x \in V$, then f is in L^p for $1 < p < 1/\delta$, while ξf belongs to L^q for $1 < q < \infty$. In both cases f will be integrable, forming a quadrature pair with $\mathcal{H}[f]$, and $\mathcal{H}[f]$ will be integrable if and only if $\int_{\mathbb{R}} f = 0$.

Let us briefly explain what happens when condition (i) is not satisfied. We suppose that f is in $L^1 \cap L^2$, with ξf also in L^2, but $\int_{\mathbb{R}} f = m \neq 0$. Then f and $\mathcal{H}[f]$ are in quadrature, but $\mathcal{H}[f]$ is not integrable; in fact the Fourier transform of $\mathcal{H}[f]$ is discontinuous at the origin, where it jumps from $i\,m$ to $-i\,m$. Furthermore, (3.20) gives $\xi\mathcal{H}[f] = \mathcal{H}[\xi f] + m/\pi$, and as $\mathcal{H}[\xi f]$ is in L^2, $\xi\mathcal{H}[f]$ is not square-integrable; in fact for every $\varepsilon > 0$, the set of all $x \in \mathbb{R}$ such that $|x\mathcal{H}[f](x) - m/\pi| > \varepsilon$ has finite measure. Furthermore, ξf and $\xi\mathcal{H}[f]$ are not in quadrature. This shows in particular that Propositions 3.7 and 3.8 are not true for functions in L^2 instead of L^1.

In [15] we considered the particular case where f is given by the Gaussian

$$G_\sigma(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[\frac{-x^2}{2\sigma^2}\right], \tag{3.21}$$

whose Hilbert transform is the function

$$K_\sigma(x) = \frac{2}{\sqrt{\pi}}\, G_\sigma(x) \int_0^{x/\sigma\sqrt{2}} \exp(s^2)\, ds. \tag{3.22}$$

Clearly G_σ and ξG_σ belong to L^p for all p, and $\int_{\mathbb{R}} G_\sigma = 1$. We have $\xi K_\sigma = \mathcal{H}[\xi G_\sigma] + 1/\pi$, and in fact $\lim_{x\to\pm\infty} xK_\sigma(x) = 1/\pi$, so that $K_\sigma(x)$ is asymptotically in $1/\pi x$. If one take for f the n-th derivative $G_\sigma^{(n)}$ of G_σ, then $\mathcal{H}[f]$ will be the n-th derivative $K_\sigma^{(n)}$ of K_σ, which is asymptotically proportional to $1/\xi^{n+1}$ (because n is the least $m \geq 0$ such that $\int_{\mathbb{R}} \xi^m G_\sigma^{(n)} \neq 0$). On the other hand, if f is a difference of Gaussians $G_{\sigma_1} - G_{\sigma_2}$ (where $\sigma_1 \neq \sigma_2$), then $\mathcal{H}[f] = K_{\sigma_1} - K_{\sigma_2}$ is asymptotically proportional to $1/\xi^3$ (because $\int_{\mathbb{R}} \xi^2 \big(G_{\sigma_1} - G_{\sigma_2}\big) = {\sigma_1}^2 - {\sigma_2}^2 \neq 0$). Thus functions involving the Hilbert transform K_σ of the Gaussian G_σ have a relatively slow decay (in addition to their computational complexity). This example shows that the celebrated Gaussian, which is usually considered as an optimal smoothing function, is not a very good candidate for building Hilbert transform pairs of integrable filters.

In [9] the following function $g : \mathbb{R} \to \mathbb{R}$ is considered:

$$g(\nu) = \exp\Big(-\frac{[\ln(|\nu|/P)]^2}{2[q\ln 2]^2}\Big). \tag{3.23}$$

Since g is square-integrable, there is a square-integrable function f such that $\widehat{f} = g$. Here P is the peak frequency of f, and q its half bandwidth in octaves at height $\exp(-1/2)$. This choice of $g = \widehat{f}$ results from data obtained in psychophysical measurements of human visual response. Since g and all its n-th derivatives $g^{(n)}$ are integrable (for all $n > 0$), repeated application of integration by parts gives $g^{(n)\vee} = (-2\pi\, i\, \xi)^n g^\vee \equiv (-2\pi\, i\, \xi)^n f$, and as $g^{(n)}$ is square-integrable, it follows that $(-2\pi\, i\, \xi)^n f$ is square-integrable and $\big[(-2\pi\, i\, \xi)^n f\big]^\wedge \equiv g^{(n)}$ for all $n > 0$. As f and all $\xi^n f$ are square-integrable, by repeated application of Lemma 3.11, f and all $\xi^n f$ will also be integrable; furthermore, as $g(0) = g^{(n)}(0)$, we have $\int f = \int \xi^n f = 0$. Now $\mathcal{H}[f]^\wedge \equiv -i \operatorname{sgn} \cdot g$, and the same argument shows that the above properties hold with $\mathcal{H}[f]$ and $\operatorname{sgn} \cdot g$ instead of f and g. The functions f and $\mathcal{H}[f]$ were proposed in [9] as filters for building the energy function.

In order to generate $\mathbf{n}$-directional Hilbert transform pairs of integrable functions on $\mathcal{E}$ which are in $\mathbf{n}$-quadrature, we can take functions of the form given in item (iii) of Lemma 3.9, namely for $f : \mathbb{R} \to \mathbb{C}$ satisfying the conditions of Lemma 3.11, and for an integrable function $G : \mathcal{E}^0_\mathbf{n} \to \mathbb{C}$, the functions $Gf : (\mathbf{y}, t) \mapsto G(\mathbf{y})f(t)$ and $G\mathcal{H}[f] : (\mathbf{y}, t) \mapsto G(\mathbf{y})\mathcal{H}[f](t)$ are integrable and verify $G\mathcal{H}[f] = \mathcal{H}_\mathbf{n}[Gf]$. We can also generalize Lemma 3.11 to the multidimensional case. Recall the function $\xi_n : \mathcal{E} \to \mathbb{C} : \mathbf{x} \mapsto x_n = \mathbf{x} \cdot \mathbf{n}$.

PROPOSITION 3.12. (a) *Let $1 < p, q < \infty$, and let F be a function in L^p such that the function $\xi_\mathbf{n} F$ is in L^q. Then for almost all $\mathbf{y} \in \mathcal{E}^0_\mathbf{n}$, $F_\mathbf{y}$ is integrable and for every $x_n \in \mathbb{R}$,*

$$\mathcal{H}_\mathbf{n}[\xi_\mathbf{n} F](\mathbf{y}, x_n) = x_n \mathcal{H}_\mathbf{n}[F](\mathbf{y}, x_n) - \frac{1}{\pi} \int_{\mathbb{R}} F(\mathbf{y}, t)\, dt. \tag{3.24}$$

Moreover, the following four statements are equivalent:

(i) *For almost all $\mathbf{y} \in \mathcal{E}^0_\mathbf{n}$, $\int_{\mathbb{R}} F(\mathbf{y}, t)\, dt = 0$.*

(ii) *$\mathcal{H}_\mathbf{n}[\xi_\mathbf{n} F] \equiv \xi_\mathbf{n} \mathcal{H}_\mathbf{n}[F]$.*

(iii) *$\xi_\mathbf{n} \mathcal{H}_\mathbf{n}[F]$ is in L^q.*

(iv) *For almost all $\mathbf{y} \in \mathcal{E}^0_\mathbf{n}$, $\big(\mathcal{H}_\mathbf{n}[F]\big)_\mathbf{y} = \mathcal{H}\big[F_\mathbf{y}\big]$ is integrable.*

(b) *Let $\{\mathbf{e}_1, \ldots, \mathbf{e}_{d-1}\}$ be an orthonormal basis of $\mathcal{E}^0_\mathbf{n}$ and let $\mathbf{e}_d = \mathbf{n}$; for $i = 1, \ldots, d$ let ξ_i be the map $\mathcal{E} \to \mathbb{R} : \mathbf{x} \mapsto \mathbf{x} \cdot \mathbf{e}_i$. Let the function F be such that for every subset I of $\{1, \ldots, d\}$ there is some $p(I)$ with $1 < p(I) < \infty$ for which the function $\big(\prod_{i \in I} \xi_i\big) \cdot F$ is in $L^{p(I)}$ (in particular F is in $L^{p(\emptyset)}$ and $\xi_\mathbf{n} F = \xi_d F$ is in $L^{p(\{d\})}$). Then F is integrable, the statement of (a) holds, and F and $\mathcal{H}_\mathbf{n}[F]$ are in $\mathbf{n}$-quadrature. Moreover, if F satisfies anyone of the conditions (i) to (iv) of (a), then $\mathcal{H}_\mathbf{n}[F]$ is integrable.*

PROOF. (a) Since F is in L^p, by Fubini's theorem we have $\int_{\mathcal{E}^0_\mathbf{n}} d\mathbf{y} \int_{\mathbb{R}} \big|F_\mathbf{y}\big|^p = \int_{\mathcal{E}} |F|^p < \infty$, and so $\int_{\mathbb{R}} \big|F_\mathbf{y}\big|^p < \infty$ for almost every $\mathbf{y} \in \mathcal{E}^0_\mathbf{n}$. Similarly, since $\xi_\mathbf{n} F$ is in L^q, we get $\int_{\mathbb{R}} \big|\big(\xi_\mathbf{n} F\big)_\mathbf{y}\big|^q = \int_{\mathbb{R}} \big|\xi F_\mathbf{y}\big|^q < \infty$ for almost every $\mathbf{y} \in \mathcal{E}^0_\mathbf{n}$. Hence for almost all $\mathbf{y} \in \mathcal{E}^0_\mathbf{n}$ we have $F_\mathbf{y}$ in L^p and $\xi F_\mathbf{y}$ in L^q, and so $F_\mathbf{y}$ is integrable by Lemma 3.11. Then:

$$\begin{aligned}
\big(\mathcal{H}_\mathbf{n}[\xi_\mathbf{n} F]\big)_\mathbf{y} &= \mathcal{H}\big[\big(\xi_\mathbf{n} F\big)_\mathbf{y}\big] = \mathcal{H}\big[\xi F_\mathbf{y}\big] && \text{by (3.18)} \\
&= \xi \mathcal{H}\big[F_\mathbf{y}\big] - \frac{1}{\pi} \int_{\mathbb{R}} F_\mathbf{y} && \text{by (3.20)} \\
&= \xi \big(\mathcal{H}_\mathbf{n}[F]\big)_\mathbf{y} - \frac{1}{\pi} \int_{\mathbb{R}} F_\mathbf{y} = \big(\xi_\mathbf{n} \mathcal{H}_\mathbf{n}[F]\big)_\mathbf{y} - \frac{1}{\pi} \int_{\mathbb{R}} F_\mathbf{y} && \text{by (3.18)}
\end{aligned}$$

This gives thus (3.24).

Now the equivalence between (*i*), (*ii*), (*iii*), and (*iv*) is obtained in the same way as in Lemma 3.11. By item (*i*) of Lemma 3.9, $\mathcal{H}_{\mathbf{n}}[F]$ is in L^p, while $\mathcal{H}_{\mathbf{n}}[\xi_{\mathbf{n}}F]$ is in L^q. For almost all $\mathbf{y} \in \mathcal{E}^0_{\mathbf{n}}$, $F_{\mathbf{y}}$ is in $L^1 \cap L^p$, so that $F_{\mathbf{y}}$ and $\mathcal{H}\big[F_{\mathbf{y}}\big]$ are in quadrature by Lemma 3.10. Thanks to (3.24), (*i*) implies (*ii*). As $\mathcal{H}_{\mathbf{n}}[\xi_{\mathbf{n}}F]$ is in L^q, (*ii*) implies (*iii*). If (*iii*) holds, as $\mathcal{H}_{\mathbf{n}}[F]$ is in L^p, the above argument with $\mathcal{H}_{\mathbf{n}}[F]$ instead of F gives (*iv*). If (*iv*) holds, $F_{\mathbf{y}}$ and $\mathcal{H}\big[F_{\mathbf{y}}\big]$ being in L^1 and in quadrature imply that $\int_{\mathbb{R}} F(\mathbf{y}, t)\, dt = 0$ (by Proposition 3.7, or Proposition 3 of [15]).

(*b*) By Lemma 3.1, F is integrable. Let $p = p(\emptyset)$ and $q = p(\{d\})$. Since F is in $L^1 \cap L^p$, F and $\mathcal{H}_{\mathbf{n}}[F]$ are in $\mathbf{n}$-quadrature by Lemma 3.10. Now F satisfies the conditions of statement (*a*), and in particular conditions (*i*) to (*iv*) are equivalent. Assume that one of them holds; thus we have (*ii*): $\mathcal{H}_{\mathbf{n}}[\xi_{\mathbf{n}}F] \equiv \xi_{\mathbf{n}}\mathcal{H}_{\mathbf{n}}[F]$. For a subset J of $\{1, \ldots, d-1\}$, $\big(\prod_{j\in J} \xi_j\big) \cdot F$ is in $L^{p(J)}$ and $\big(\prod_{j\in J} \xi_j\big)\xi_{\mathbf{n}} \cdot F = \big(\prod_{j\in J} \xi_j\big) \cdot \xi_d F$ is in $L^{p(J\cup\{p\})}$, where $1 < p(J), p(J \cup \{p\}) < \infty$. By item (*iv*) of Lemma 3.9, $\mathcal{H}_{\mathbf{n}}\big[\big(\prod_{j\in J} \xi_j\big)\cdot F\big] = \big(\prod_{j\in J} \xi_j\big)\cdot\mathcal{H}_{\mathbf{n}}[F]$ and $\mathcal{H}_{\mathbf{n}}\big[\big(\prod_{j\in J} \xi_j\big)\cdot\xi_{\mathbf{n}}F\big] = \big(\prod_{j\in J} \xi_j\big)\cdot\mathcal{H}_{\mathbf{n}}[\xi_{\mathbf{n}}F] \equiv \big(\prod_{j\in J} \xi_j\big) \cdot \xi_{\mathbf{n}}\mathcal{H}_{\mathbf{n}}[F]$. By item (*i*) of Lemma 3.9, we deduce that $\big(\prod_{j\in J} \xi_j\big) \cdot \mathcal{H}_{\mathbf{n}}[F]$ is in $L^{p(J)}$ and $\big(\prod_{j\in J} \xi_j\big) \cdot \xi_{\mathbf{n}}\mathcal{H}_{\mathbf{n}}[F]$ is in $L^{p(J\cup\{p\})}$. Thus $\mathcal{H}_{\mathbf{n}}[F]$ satisfies the same above-mentioned condition as F, and so $\mathcal{H}_{\mathbf{n}}[F]$ is integrable by Lemma 3.1. ∎

A simplified form of statement of (*b*) applies when we take the same p for all $I \subseteq \{1, \ldots, d\}$. Suppose here that we have any basis $\{\mathbf{u}_1, \ldots, \mathbf{u}_d\}$ of $\mathcal{E}$ (not necessarily orthonormal) for which there is some p with $1 < p < \infty$ such that for every subset I of $\{1, \ldots, d\}$ the function $\mathbf{x} \mapsto \big(\prod_{i\in I} \mathbf{x} \cdot \mathbf{u}_i\big) \cdot F(\mathbf{x})$ is in L^p; then for any other basis $\{\mathbf{e}_1, \ldots, \mathbf{e}_d\}$ of $\mathcal{E}$, the functions $\mathbf{x} \mapsto \big(\prod_{i\in I} \mathbf{x} \cdot \mathbf{e}_i\big) \cdot F(\mathbf{x})$, being linear combinations of those of the form $\mathbf{x} \mapsto \big(\prod_{j\in J} \mathbf{x} \cdot \mathbf{u}_j\big) \cdot F(\mathbf{x})$, will also be in L^p, and so statement (*b*) holds then.

If we have a function F satisfying the statement of (*a*) such that (*i*) is not verified, one can construct a new one satisfying (*i*) as follows: Let $a > 0$, and define H_a by

$$H_a(\mathbf{y}, t) = F(\mathbf{y}, t) - a^{-1}F(\mathbf{y}, a^{-1}t);$$

then clearly $\int_{\mathbb{R}} H_a(\mathbf{y}, t)\, dt = 0$ for all $\mathbf{y} \in \mathcal{E}^0_{\mathbf{n}}$. In the Fourier domain we have

$$\widehat{H_a}(\mathbf{u}, v) = \widehat{F}(\mathbf{u}, v) - \widehat{F}(\mathbf{u}, av);$$

thus if F has on $\mathcal{E}^+_{\mathbf{n}}$ a constant Fourier phase ϕ and a Fourier amplitude $F^{\mathcal{A}}(\mathbf{u}, v)$ decreasing in v, then for $a > 1$ we will get $\widehat{H_a}(\mathbf{u}, v) = e^{i\phi}\big(F^{\mathcal{A}}(\mathbf{u}, v) - F^{\mathcal{A}}(\mathbf{u}, av)\big)$, and so H_a will have the same Fourier phase as F. For example, taking $d = 1$, if F is the Gaussian G_σ of (3.21), then H_a is the difference of Gaussians $G_\sigma - G_{a\sigma}$, which has in common with the Gaussian a constant zero Fourier phase.

Thanks to Proposition 3.12, we have a criterion for determining whether a function F and its $\mathbf{n}$-directional Hilbert transform $\mathcal{H}_{\mathbf{n}}(F)$ are in $\mathbf{n}$-quadrature and both integrable. Such pairs of functions will be the filters used in the phase congruence model for constructing energy functions leading to edge detection.

There is another generalization of the Hilbert transform to d dimensions, namely the *M. Riesz transform* (see [46], p. 224), given by

$$\mathcal{R}[F](\mathbf{x}) = \lim \frac{\Gamma[(n+1)/2]}{\pi^{(n+1)/2}} \int_{\delta\geq|\mathbf{t}|\geq\varepsilon} F(\mathbf{x} - \mathbf{t}) \frac{t_n}{|\mathbf{t}|^{n+1}}\, d\mathbf{t}, \qquad \varepsilon \to 0, \delta \to \infty,$$

where $t_n = \mathbf{t} \cdot \mathbf{n}$. For F square-integrable, this leads to the following equality in the Fourier domain:

$$\mathcal{R}[F]^\wedge(\mathbf{u}) = -i \, \frac{u_n}{|\mathbf{u}|} \widehat{F}(\mathbf{u}).$$

We did not use this transform, because it does not lead to a Fourier phase quadrature, since it modifies also the Fourier amplitude of a function F.

3.5. *The complex-valued function associated to a quadrature pair, and its energy*

Given a square-integrable function f on $\mathbb{R}$, the complex-valued function $f + i\,\mathcal{H}[f]$ has some interesting mathematical properties, in particular the fact that it can be extended to a complex function which is analytic in the upper half of the complex plane. The use of this function in signal processing has been proposed by Gabor [54], who called it the *analytic signal* associated to f, and who introduced the term *energy* in order to designate the square of its absolue value, namely $f^2 + \mathcal{H}[f]^2 = \big|f + i\,\mathcal{H}[f]\big|^2$. Owens [55] used extensively this function $f + i\,\mathcal{H}[f]$ and its analytic extension in the upper half of the complex plane in order to describe the behaviour of the energy function $f^2 + \mathcal{H}[f]^2$ by means of complex analysis. We will consider here the complex-valued function $F + i\,G$ for a $\mathbf{n}$-quadrature pair of functions on $\mathcal{E}$.

Given two functions $F, G : \mathcal{E} \to \mathbb{C}$, the fact that F and G are in $\mathbf{n}$-quadrature is equivalent to each of the following identities:

$$\begin{aligned} [F + i\,G]^\wedge &\equiv 2 \,\mathrm{pos}_{\mathbf{n}}\, \widehat{F}; \\ [F - i\,G]^\wedge &\equiv 2 \,\mathrm{neg}_{\mathbf{n}}\, \widehat{F}. \end{aligned} \tag{3.25}$$

More precisely, consider the following four conditions:

(a) For almost all $\mathbf{u} \in \mathcal{E}^+_{\mathbf{n}}$, $[F + i\,G]^\wedge(\mathbf{u}) = 2\,\widehat{F}(\mathbf{u})$.
(b) For almost all $\mathbf{u} \in \mathcal{E}^-_{\mathbf{n}}$, $[F + i\,G]^\wedge(\mathbf{u}) = 0$.
(c) For almost all $\mathbf{u} \in \mathcal{E}^+_{\mathbf{n}}$, $[F - i\,G]^\wedge(\mathbf{u}) = 0$.
(b) For almost all $\mathbf{u} \in \mathcal{E}^-_{\mathbf{n}}$, $[F - i\,G]^\wedge(\mathbf{u}) = 2\,\widehat{F}(\mathbf{u})$.

Then (a) $\Leftrightarrow$ (c), (b) $\Leftrightarrow$ (d), and any combination of (a) or (c) with (b) or (d) expresses the fact that F and G are in $\mathbf{n}$-quadrature; in particular (a) and (b) together mean the first identity in (3.25), while (c) and (d) together mean the second one. Furthermore, if F and G are real-valued, (3.7) and (3.8) applied to $F \pm i\,G$ imply that (a, b, c, d) are pairwise equivalent, and so each one is equivalent to the fact that F and G are in $\mathbf{n}$-quadrature.

Thanks to (3.7), for F and G complex-valued, F and G are in $\mathbf{n}$-quadrature if and only if $\overline{F}$ and $\overline{G}$ are in $\mathbf{n}$-quadrature, in other words if and only if both $(\Re F, \Re G)$ and $(\Im F, \Im G)$ are $\mathbf{n}$-quadrature pairs. Then $F + i\,G = (\Re F - \Im G) + i\,(\Re G + \Im F)$, where $\Re F - \Im G$ and $\Re G + \Im F$ are in $\mathbf{n}$-quadrature. Thus when considering the complex-valued function $F + i\,G$ associated to a $\mathbf{n}$-quadrature pair (F, G), we can assume that F and G are real-valued.

Consider now the family $\mathcal{F}^+_{\mathbf{n}}$ of all functions $H : \mathcal{E} \to \mathbb{C}$ such that $\widehat{H}(\mathbf{u}) = 0$ a.e. on $\mathcal{E}^-_{\mathbf{n}}$, in other words of all functions of the form $F + i\,\widetilde{F}$ for F and $\widetilde{F}$ in $\mathbf{n}$-quadrature (and we can assume here that F and $\widetilde{F}$ are real-valued). Clearly $\mathcal{F}^+_{\mathbf{n}}$ is a vector space on $\mathbb{C}$; in particular for any angle θ and $F, \widetilde{F}$ in $\mathbf{n}$-quadrature,

$$e^{i\,\theta}(F + i\,\widetilde{F}) = \big(\cos\theta\, F - \sin\theta\, \widetilde{F}\big) + i\,\big(\sin\theta\, F + \cos\theta\, \widetilde{F}\big) \tag{3.26}$$

belongs to $\mathcal{F}_{\mathbf{n}}^{+}$, so that $\big(\cos\theta\, F - \sin\theta\, \widetilde{F}\big)$ and $i\,\big(\sin\theta\, F + \cos\theta\, \widetilde{F}\big)$ are in **n**-quadrature; moreover we have a.e.:

$$\big(\cos\theta\, F - \sin\theta\, \widetilde{F}\big)^{\wedge}(\mathbf{u}) = \exp[i\,\theta\, \mathrm{sgn}_{\mathbf{n}}(\mathbf{u})]\, \widehat{F}(\mathbf{u}). \qquad (3.27)$$

By the convolution formula, for $F + i\,\widetilde{F}$ and $G + i\,\widetilde{G}$ in $\mathcal{F}_{\mathbf{n}}^{+}$ with at least one of them integrable, their convolution

$$(F + i\,\widetilde{F}) * (G + i\,\widetilde{G}) = (F * G - \widetilde{F} * \widetilde{G}) + i\,(F * \widetilde{G} + \widetilde{F} * G)$$

belongs to $\mathcal{F}_{\mathbf{n}}^{+}$, so that $F * G - \widetilde{F} * \widetilde{G}$ and $F * \widetilde{G} + \widetilde{F} * G$ are in **n**-quadrature.

Similarly, if both $F + i\,\widetilde{F}$ and $G + i\,\widetilde{G}$ in $\mathcal{F}_{\mathbf{n}}^{+}$ are square-integrable, by the dual convolution formula and the fact that the convolution of two functions vanishing on $\mathcal{E}_{\mathbf{n}}^{-}$ vanishes on $\mathcal{E}_{\mathbf{n}}^{-}$, their product

$$(F + i\,\widetilde{F})(G + i\,\widetilde{G}) = (FG - \widetilde{F}\widetilde{G}) + i\,(F\widetilde{G} + \widetilde{F}G) \qquad (3.28)$$

belongs to $\mathcal{F}_{\mathbf{n}}^{+}$, so that $FG - \widetilde{F}\widetilde{G}$ and $F\widetilde{G} + \widetilde{F}G$ are in **n**-quadrature. Here $\widetilde{F} = \mathcal{H}_{\mathbf{n}}[F]$ and $\widetilde{G} = \mathcal{H}_{\mathbf{n}}[G]$, and in the one-dimensional case where $\mathcal{E} = \mathrm{I\!R}$ and the **n**-directional Hilbert transform reduces to the ordinary one, such a formula $FG - \widetilde{F}\widetilde{G}$ was used in [55] as a new type of product of F and G, and it was written $F \odot G$. The advantage of such a product $F \odot G$ is that the energy associated to it is the ordinary product of the energies associated to F and G; for example if G has constant energy, the energy of $F \odot G$ is proportional to that of F.

Let us now explain how in the one-dimensional case such a complex-valued fonction extends to a complex function which is analytic in the upper half of the complex plane. Given a square-integrable function $f : \mathrm{I\!R} \to \mathrm{I\!R}$, let $g = f + i\,\mathcal{H}[f]$ and $h = \widehat{g}$, that is $h(\nu) = 2\,\widehat{f}(\nu)$ for $\nu > 0$ and $h(\nu) = 0$ for $\nu < 0$. We extend g to a complex function G by extending the inverse Fourier transform of h to a complex variable:

$$G(z) = \int_0^{\infty} h(\nu)\, \exp[2\pi i\, \nu z]\, d\nu \qquad (z \in \mathbb{C}), \qquad (3.29)$$

in other words for $z = x + i\,y$ we have:

$$G(x + i\,y) = \int_0^{\infty} h(\nu)\, \exp[-2\pi\nu y]\, \exp[2\pi i\, \nu x]\, d\nu. \qquad (3.30)$$

Then for for a fixed $y > 0$, the function E_y defined by

$$E_y(\nu) = \begin{cases} \exp[-2\pi\nu y] & \text{for } \nu \geq 0, \\ 0 & \text{for } \nu < 0, \end{cases}$$

is square-integrable, and $G(x+i\,y)$ as a function of x is the inverse Fourier transform of the integrable function $h \cdot E_y$. Thus $G(z)$ is well-defined for $\Im z > 0$. It is not hard to see that for $\Im z > 0$, G can be derived by deriving w.r.t. z under the integral in (3.29):

$$G'(z) = \int_0^{\infty} h(\nu)\, \exp[2\pi i\, \nu z]\, 2\pi i\, \nu\, d\nu \qquad (z \in \mathbb{C}, \Im z > 0). \qquad (3.31)$$

This result can be proved in the same way as the L^1 Fourier derivative formula (see also Theorem 2.27 of [48]). Since G is derivable in the open set $\Im z > 0$, it is analytic on it. Now g can be obtained from G as its "non-tangential limit". This means that for every $a > 0$, for x real and $\Im z > 0$, $z \to x$

subject to $|\Re(z-x)| < a|\Im(z-x)|$ gives $G(z) \to g(x)$. See [56,57] for more details; a short exposition of this theory is given in [55].

Let us give a concrete example. We define h by

$$h(\nu) = \begin{cases} (n!)^{-1}(2\pi)^{n+1}t^n \exp[-2\pi\nu] & \text{for } \nu \geq 0, \\ 0 & \text{for } \nu < 0, \end{cases} \tag{3.32}$$

where $n \geq 0$. Then $h \cdot E_y$ is integrable for $y > -1$, and so for $\Im z > -1$, $G(z)$ can be defined by (3.29) and has its derivative as in (3.31); in fact G has a pole at $z = -i$. One can show by induction on n that $G(z) = (1 - i\,z)^{-n-1}$, which is indeed analytic for $\Im z > -1$. In particular for x real, $g(x) = (1 - i\,x)^{-n-1}$, and we get the quadrature pair $(f, \widetilde{f})$ given by:

$$\begin{aligned} f(x) &= \Re\big\{(1-ix)^{-n-1}\big\} = \frac{\Re\big\{(1+ix)^{n+1}\big\}}{(1+x^2)^{n+1}}, \\ \widetilde{f}(x) &= \Im\big\{(1-ix)^{-n-1}\big\} = \frac{\Im\big\{(1+ix)^{n+1}\big\}}{(1+x^2)^{n+1}}. \end{aligned} \tag{3.33}$$

These functions were proposed in [58], where they were called *Cauchy functions*. They are square-integrable, and integrable for $n \geq 1$. The constant n determines also the number of ripples in the profiles of f and $\widetilde{f}$. In [58] the values $n = 3$ and $n = 5$ were taken in order to define filters modeling some properties of the human visual response. The main interest of Cauchy functions resides in their computational simplicity.

Let us now give properties of the energy function $F^2 + \widetilde{F}^2$ for F and $\widetilde{F}$ in $\mathbf{n}$-quadrature. Recall the definition given in (3.14) of $F^\Phi(\mathbf{u}, \mathbf{p})$, the local Fourier phase of F at $\mathbf{p}$ for the frequency $\mathbf{u}$.

LEMMA 3.13. *Let $G : \mathcal{E} \to \mathbb{C}$ be square-integrable and such that there is some $\mathcal{V} \subseteq \mathcal{E}$ with $\widehat{G}(\mathbf{u}) = 0$ for $\mathbf{u} \notin \mathcal{V}$. Then:*

(i) $(G\overline{G})^\wedge = \widehat{G} * \overline{G^\vee}$. *In particular, if G^Φ is constant (on $\mathcal{V}$), then $(G\overline{G})^\wedge = G^\mathcal{A} * (G^\mathcal{A})_\rho$, a real-valued non-negative function.*

(ii) *For $\mathbf{u} \in \mathcal{E}$ and $\mathcal{V}_\mathbf{u}$ the translate of $\mathcal{V}$ by $\mathbf{u}$, we have:*

$$\begin{aligned} (G\overline{G})^\wedge(\mathbf{u}) &= \int_{\mathcal{V}\cap\mathcal{V}_\mathbf{u}} \widehat{G}(\mathbf{v})\overline{\widehat{G}(\mathbf{v}-\mathbf{u})}\,d\mathbf{v} \\ &= \int_{\mathcal{V}\cap\mathcal{V}_\mathbf{u}} G^\mathcal{A}(\mathbf{v})G^\mathcal{A}(\mathbf{v}-\mathbf{u}) \exp\big[i\big(G^\Phi(\mathbf{v}) - G^\Phi(\mathbf{v}-\mathbf{u})\big)\big]\,d\mathbf{v}. \end{aligned} \tag{3.34}$$

In particular $(G\overline{G})^\wedge(\mathbf{u}) = 0$ when $\mathcal{V} \cap \mathcal{V}_\mathbf{u}$ is negligible.

(iii) *If $\widehat{G}$ is integrable, then for almost all $\mathbf{x} \in \mathcal{E}$ we have*

$$\begin{aligned} (G\overline{G})(\mathbf{x}) &= \int_\mathcal{V}\int_\mathcal{V} \widehat{G}(\mathbf{u})\overline{\widehat{G}(\mathbf{v})} \exp[2\pi i\,(\mathbf{u}-\mathbf{v})\cdot\mathbf{x}]\,d\mathbf{u}d\mathbf{v} \\ &= \int_\mathcal{V}\int_\mathcal{V} G^\mathcal{A}(\mathbf{u})G^\mathcal{A}(\mathbf{v}) \cos\big[G^\Phi(\mathbf{u},\mathbf{x}) - G^\Phi(\mathbf{v},\mathbf{x})\big]\,d\mathbf{u}d\mathbf{v}. \end{aligned} \tag{3.35}$$

PROOF. (*i*) By (3.8) and the dual convolution formula, we have $(G\overline{G})^\wedge = \widehat{G} * \overline{G^\vee}$. If G^Φ is a constant ϕ , we get $\widehat{G} = e^{i\phi}G^\mathcal{A}$ and $\overline{G^\vee} = e^{-i\phi}(G^\mathcal{A})_\rho$, so that:

$$(G\overline{G})^\wedge = \widehat{G} * \overline{G^\vee} = e^{i\phi}G^\mathcal{A} * e^{-i\phi}(G^\mathcal{A})_\rho = G^\mathcal{A} * (G^\mathcal{A})_\rho,$$

the convolution of two non-negative real-valued functions, which will be non-negative.

(ii) Expanding (i) gives:

$$(G\overline{G})^{\wedge}(\mathbf{u}) = \int_{\mathcal{E}} \widehat{G}(\mathbf{v})\overline{G^{\vee}(\mathbf{u}-\mathbf{v})}\, d\mathbf{v} = \int_{\mathcal{E}} \widehat{G}(\mathbf{v})\overline{\widehat{G}(\mathbf{v}-\mathbf{u})}\, d\mathbf{v}.$$

Since $\widehat{G}$ vanishes outside $\mathcal{V}$, in the above integral we can restrict the domain to where both $\mathbf{v}$ and $\mathbf{v}-\mathbf{u}$ are in $\mathcal{V}$, that is $\mathbf{v} \in \mathcal{V} \cap \mathcal{V}_{\mathbf{u}}$; this gives the first equality in (3.34). In particular the integral is null when $\mathcal{V} \cap \mathcal{V}_{\mathbf{u}}$ has measure zero. Now by definition of Fourier amplitude and phase we have $\widehat{G}(\mathbf{v}) = G^{\mathcal{A}}(\mathbf{v}) \exp[i\, G^{\Phi}(\mathbf{v})]$ and $\overline{\widehat{G}(\mathbf{v}-\mathbf{u})} = G^{\mathcal{A}}(\mathbf{v}-\mathbf{u}) \exp[-i\, G^{\Phi}(\mathbf{v}-\mathbf{u})]$, giving the second equality in (3.34).

(iii) If $\widehat{G}$ is integrable, by Lemma 3.2 the formula

$$G(x) = \int_{\mathcal{V}} \widehat{G}(\mathbf{u}) \exp[2\pi i\, \mathbf{u}\cdot\mathbf{x}]\, d\mathbf{u}$$

holds almost everywhere; this gives thus

$$(G\overline{G})(\mathbf{x}) = \int_{\mathcal{V}} \widehat{G}(\mathbf{u}) \exp[2\pi i\, \mathbf{u}\cdot\mathbf{x}]\, d\mathbf{u} \int_{\mathcal{V}} \overline{\widehat{G}(\mathbf{v}) \exp[2\pi i\, \mathbf{v}\cdot\mathbf{x}]}\, d\mathbf{v}.$$

Fubini's theorem gives then the first equality in (3.35). By definition of Fourier amplitude and phase, this becomes

$$(G\overline{G})(\mathbf{x}) = \int_{\mathcal{V}}\int_{\mathcal{V}} G^{\mathcal{A}}(\mathbf{u})G^{\mathcal{A}}(\mathbf{v}) \exp\bigl[2\pi i\, \mathbf{u}\cdot\mathbf{x} + i\, G^{\Phi}(\mathbf{u}) - 2\pi i\, \mathbf{v}\cdot\mathbf{x} - i\, G^{\Phi}(\mathbf{v})\bigr]\, d\mathbf{u}d\mathbf{v}.$$

Using definition (3.14) of the local phase at $\mathbf{x}$ for frequency $\mathbf{u}$, this gives

$$(G\overline{G})(\mathbf{x}) = \int_{\mathcal{V}}\int_{\mathcal{V}} G^{\mathcal{A}}(\mathbf{u})G^{\mathcal{A}}(\mathbf{v}) \exp\bigl[i\bigl(G^{\Phi}(\mathbf{u},\mathbf{x}) - G^{\Phi}(\mathbf{v},\mathbf{x})\bigr)\bigr]\, d\mathbf{u}d\mathbf{v}.$$

As $G\overline{G}$ is real-valued, the imaginary part of this double integral vanishes; the real part of it gives then the last member of (3.35). ∎

In the case where $G = F + \widetilde{F}$, where F and $\widetilde{F}$ are in $\mathbf{n}$-quadrature, we have $\mathcal{V} = \mathcal{E}^{+}_{\mathbf{n}}$, and combining ($i$) with (3.25) we get:

$$(F^2 + \widetilde{F}^2)^{\wedge} = (F + i\,\widetilde{F})^{\wedge} * (F - i\,\widetilde{F})^{\wedge} = 4\,(\mathrm{pos}_{\mathbf{n}}\,\widehat{F}) * (\mathrm{neg}_{\mathbf{n}}\,\widehat{F}). \tag{3.36}$$

If F^{Φ} is constant on $\mathcal{E}^{+}_{\mathbf{n}}$, it follows that $(F^2 + \widetilde{F}^2)^{\wedge}$ is non-negative real-valued. From the results of Subsection 3.3, provided that $F + \widetilde{F}$ is continuous, $(F + i\,\widetilde{F})^{\wedge}$ and $(F - i\,\widetilde{F})^{\wedge}$ will be integrable, and $(F^2 + \widetilde{F}^2)(\mathbf{0}) > (F^2 + \widetilde{F}^2)(\mathbf{u})$ for all $\mathbf{u} \neq \mathbf{0}$ (this inequality can also be verified from (3.35)). Formula (3.35) gives here:

$$\begin{aligned}(F^2 + \widetilde{F}^2)(\mathbf{x}) &= 4 \int_{\mathcal{E}^{+}_{\mathbf{n}}}\int_{\mathcal{E}^{+}_{\mathbf{n}}} \widehat{F}(\mathbf{u})\overline{\widehat{F}(\mathbf{v})} \exp[2\pi i\,(\mathbf{u}-\mathbf{v})\cdot\mathbf{x}]\, d\mathbf{u}d\mathbf{v} \\ &= 4 \int_{\mathcal{E}^{+}_{\mathbf{n}}}\int_{\mathcal{E}^{+}_{\mathbf{n}}} F^{\mathcal{A}}(\mathbf{u})F^{\mathcal{A}}(\mathbf{v}) \cos\bigl[F^{\Phi}(\mathbf{u},\mathbf{x}) - F^{\Phi}(\mathbf{v},\mathbf{x})\bigr]\, d\mathbf{u}d\mathbf{v}.\end{aligned} \tag{3.37}$$

This formula holds almost everywhere, in particular whenever F and $\widetilde{F}$ are continuous. It is at the basis of the phase congruence approach: the energy function at point $\mathbf{x}$ measures the degree to which the Fourier phases at $\mathbf{x}$, namely $F^{\Phi}(\mathbf{u},\mathbf{x})$, are similar when $\mathbf{u}$ ranges over $\mathcal{E}^{+}_{\mathbf{n}}$. Indeed, the closer $F^{\Phi}(\mathbf{u},\mathbf{x})$ is to $F^{\Phi}(\mathbf{v},\mathbf{x})$, the greater is $\cos\bigl[F^{\Phi}(\mathbf{u},\mathbf{x}) - F^{\Phi}(\mathbf{v},\mathbf{x})\bigr]$, and so the greater is the integrant in (3.37).

The following result will be useful in our discussion in Section 4 concerning quadratic operators in edge detection:

Proposition 3.14. *Let $G : \mathcal{E} \to \mathbb{C}$ and $H, K : \mathcal{E} \to \mathbb{R}$ be square-integrable functions such that:*

(i) *G^Φ is constant.*

(ii) *H and K are both a.e. even-symmetric or both a.e. odd-symmetric.*

(iii) *$G^\mathcal{A} \geq H^\mathcal{A}$ and $G^\mathcal{A} \geq K^\mathcal{A}$.*

Then for any real λ such that $-1 \leq \lambda \leq 1$, $(G\overline{G} + \lambda HK)^\wedge$ is real-positive.

Proof. By Lemma 3.13 (i), $(G\overline{G})^\wedge = G^\mathcal{A} * (G^\mathcal{A})_\rho$, the convolution of two real-valued non-negative functions. If H and K are both a.e. even-symmetric, then $\widehat{H}$ and $\widehat{K}$ are both real-valued, and so is $\widehat{H} * \widehat{K}$; if H and K are both a.e. odd-symmetric, then $\widehat{H}$ and $\widehat{K}$ are both imaginary-valued, and so $\widehat{H} * \widehat{K}$ is real-valued. Thus in any case $\widehat{H} * \widehat{K}$ is real-valued, and so

$$(G\overline{G} + \lambda HK)^\wedge = G^\mathcal{A} * (G^\mathcal{A})_\rho + \lambda \widehat{H} * \widehat{K}$$

is real-valued. Now $G^\mathcal{A} \geq H^\mathcal{A}$ and $G^\mathcal{A} \geq K^\mathcal{A}$; since K is real-valued, $K^\mathcal{A}$ is symmetric, and so $(G^\mathcal{A})_\rho \geq K^\mathcal{A}$. We deduce that:

$$(G\overline{G})^\wedge = G^\mathcal{A} * (G^\mathcal{A})_\rho \geq H^\mathcal{A} * K^\mathcal{A} = |\widehat{H}| * |\widehat{K}| \geq |\widehat{H} * \widehat{K}| \geq |\lambda| \cdot |\widehat{H} * \widehat{K}| = |\lambda \widehat{H} * \widehat{K}|.$$

Therefore $(G\overline{G} + \lambda HK)^\wedge \geq 0$. ∎

4. Edge detection in the phase congruence model

We will now describe the phase congruence model for edge detection and its mathematical properties, using the results of Section 3. We deal successively with the filters and their properties, the phase congruence, the behaviour of this type of edge detector on standard edge profiles (cfr. Section 2), other quadratic combinations of the filters, in particular single-filter approaches to edge detection (such as Canny's operator [7]), and finally the problem of orientation selection.

From now on, we assume that the dimension d of our space $\mathcal{E}$ is one or two. When $d = 2$, we select a unit vector $\mathbf{n}$, which will be considered as normal to the edge orientation, and we will take a perpendicular unit vector $\mathbf{t}$, which will be considered as tangential to the edge orientation.

4.1. *The filters, their constraints and properties*

We take an even-symmetric filter C (for "cosine"), and an odd-symmetric filter S (for "sine"); write I for the image to be convolved with them; all three must be considered as functions $\mathcal{E} \to \mathbb{R}$, and furthermore C and S are implicitly assumed to be integrable or square-integrable (since we consider their Fourier transforms as functions). In view of Subsection 3.5, we define the complex-valued filter $F = C + i\,S$, and we write $J = I * F$, that is $J = (I * C) + i\,(I * S)$. Then $E = |J|^2 = (I * C)^2 + (I * S)^2$ will be called the *energy function*. Note that several authors, in particular [9,55], define $\sqrt{E} = |J|$ as the energy function, but this does not really matter.

As edge detection involves very particular filters being applied to rather arbitrary images, we choose to impose sharp constraints on C and S, but loose ones on I. The first requirement is a two-dimensional generalization of the Morrone-Burr condition [9]:

Requirement 1. *$C \not\equiv 0$, C is even-symmetric, S is odd-symmetric, $\widehat{C}$ has real non-negative values, and (C, S) is an $\mathbf{n}$-quadrature pair.*

Let us write A for the Fourier amplitude of C; then the following is a reformulation of this requirement:

$$\text{For all} \quad \mathbf{u} \in \mathcal{E}, \qquad \begin{aligned} \widehat{C}(\mathbf{u}) &= A(\mathbf{u}), \\ \widehat{S}(\mathbf{u}) &= -i\,\mathrm{sgn}_{\mathbf{n}}(\mathbf{u})A(\mathbf{u}), \end{aligned} \qquad \text{where} \quad A(\mathbf{u}) \geq 0. \tag{4.1}$$

Note that several previous studies (in particular [9,15]) have considered one-dimensional functions satisfying $\widehat{S}(\nu) = i\,\mathrm{sgn}(\nu)\widehat{C}(\mathbf{u})$, in another words such that (S, C) (rather than (C, S)) is a quadrature pair, but again this does not matter, we have then only to take $-S$ instead of S.

The next requirement will allow us to use the theory of the previous section:

Requirement 2. *C is bounded in a neighbourhood of the origin.*

This gives us the following consequence:

Proposition 4.1. *$\widehat{C}$ and $\widehat{S}$ are integrable, and there exist two bounded and uniformly continuous functions C_u and S_u vanishing at infinity, such that $C_u \equiv C$ and $S_u \equiv S$. Moreover, if C or S belongs to L^p (where $p \geq 1$), then it belongs to L^q for every $q > p$.*

Proof. $\widehat{C}$ is integrable by Proposition 3.5, and as $|\widehat{S}| = |\widehat{C}|$, it follows that $\widehat{S}$ is also integrable. Defining C_u and S_u according to (3.11), Lemma 3.2 implies that they are bounded and uniformly continuous, that they vanish at infinity, and that $C_u \equiv C$ and $S_u \equiv S$. If C is in L^p, then for $q > p$ we have $\int_{\mathcal{E}} |C|^q \leq \|C\|_\infty^{q-p} \cdot \int_{\mathcal{E}} |C|^p$, and then C is also in L^q; the same holds for S. ∎

Since $C_u \equiv C$ and $S_u \equiv S$, it follows that $I * C = I * C_u$ and $I * S = I * S_u$, in other words C and S considered as filters have the same behaviour as C_u and S_u. We can thus take the latter in place of the former:

Requirement 2'. *C and S are continuous, in other words $C = C_u$ and $S = S_u$.*

Note that Requirement 2' is stronger than Requirement 2; we can thus forget the latter.

By Proposition 4.1, if C or S is integrable, then it is also square-integrable. Thus requiring square-integrability on C and S is less restrictive than requiring integrability. As we choose to take the strongest conditions on the filters and the weakest ones on the image, we postulate the following:

Requirement 3. *C and S are integrable.*

Another reason for postulating integrability instead of square-integrability, is that a function which is square-integrable but not integrable has a slow decay: asymptotically it cannot decrease faster than $1/|\mathbf{x}|^d$.

Requirements 1, 2', and 3 are the basis of our theory. They lead to the following fundamental result:

Proposition 4.2. $\int_{\mathcal{E}} C = \int_{\mathcal{E}} S = 0$. *Furthermore:*

(*i*) *$H * C = H * S = 0$ for every constant $H : \mathcal{E} \to \mathbb{R}$.*

(*ii*) *For $I = I_1 + I_2$, where I_1 is in L^1 and I_2 is in L^2, both $I * C$ and $I * S$ are square-integrable, bounded, uniformly continuous, and they vanish at infinity. Furthermore $(I * C)^\wedge = \widehat{I}\widehat{C}$ and $(I * S)^\wedge = \widehat{I}\widehat{S}$ are integrable and square-integrable.*

Proof. By Proposition 3.7, $\int_{\mathcal{E}} C = \int_{\mathcal{E}} S = 0$. If $H(\mathbf{x}) = c$ for all $\mathbf{x} \in \mathcal{E}$, then $(H * C)(\mathbf{x}) = c \cdot \int_{\mathcal{E}} C = 0$ and similarly $(H * S(\mathbf{x}) = 0$ for all $\mathbf{x} \in \mathcal{E}$; thus (*i*) holds.

We know that C is integrable and bounded; hence it is also square-integrable. Take I_1 in L^1 and I_2 in L^2. Since C is both in L^1 and in L^2, both $I_1 * C$ and $I_2 * C$ will also be in L^2 by Young's inequality. Since C is bounded and vanishes at infinity, the p-p' convolution property for $p = 1$ implies that $I_1 * C$ will be bounded, uniformly continuous, and will vanish at infinity; since C is in L^2, the p-p' convolution property for $p = 2$ implies that $I_2 * C$ will also be bounded, uniformly continuous, and will vanish at infinity. Thus $I * C = I_1 * C + I_2 * C$ will share the common properties of $I_1 * C$ and $I_2 * C$, namely being square-integrable, bounded, uniformly continuous, and vanishing at infinity.

The convolution formula gives $(I_1 * C)^\wedge = \widehat{I_1}\widehat{C}$ and $(I_2 * C)^\wedge = \widehat{I_2}\widehat{C}$. Since C is integrable, $\widehat{C}$ is bounded by the Riemann-Lebesgue theorem, and we know that $\widehat{C}$ is integrable by Lemma 4.1; hence $\widehat{C}$ is also square-integrable. As I_1 is integrable, $\widehat{I_1}$ is bounded by the Riemann-Lebesgue theorem, and so $\widehat{I_1}\widehat{C}$ is integrable and bounded; it is in particular square-integrable. As I_2 is square-integrable, $\widehat{I_2}$ is square-integrable by the Plancherel Theorem, and as $\widehat{C}$ is square-integrable and bounded, it follows from Hölder's inequality that $\widehat{I_2}\widehat{C}$ is integrable and square-integrable. Thus

$$(I * C)^\wedge = (I_1 * C)^\wedge + (I_2 * C)^\wedge = \widehat{I_1}\widehat{C} + \widehat{I_2}\widehat{C} = \widehat{I}\widehat{C},$$

and it will share the common properties of $\widehat{I_1}\widehat{C}$ and $\widehat{I_2}\widehat{C}$, namely being both integrable and square-integrable.

The same argument works wih S instead of C, and so (ii) holds. ∎

Note that since $I * C$ and $I * S$ are continuous and have an integrable Fourier transform, their values are given pointwise by the inverse Fourier integral applied to their Fourier transform (cfr. Lemma 3.2). The corresponding formulas using (4.1) will be given later.

The above three requirements, and their consequences (Propositions 4.1 and 4.2), are the basis for the interpretation of the energy function in terms of Fourier phase congruence; this will be the subject of the next subsection. Note that these results are not limited by the assumption that $d \leq 2$; they remain valid in spaces with a higher number of dimensions.

We will now consider further requirements concerning C and S, which are specific to the two-dimensional case. Then we will examine properties of C and S in the spatial domain, and additional constraints to be imposed on them, in particular concerning their response on some types of idealized line, step, or roof signals.

Assume temporarily that $d = 2$. Given $\mathbf{x} \in \mathcal{E}$, we write x_n and x_t for the coordinates of $\mathbf{x}$ in the directions of the two perpendicular unit vectors $\mathbf{n}$ and $\mathbf{t}$, in other words $\mathbf{x} = x_t \cdot \mathbf{t} + x_n \cdot \mathbf{n}$; we can thus write $\mathbf{x} = (x_t, x_n)$. It is natural to assume that when we detect significant events in the grey-level profile along the normal direction $\mathbf{n}$, both orientations $\mathbf{t}$ and $-\mathbf{t}$ in the tangential direction should be treated symmetrically; we have thus the following:

Requirement 4. *When $d = 2$, C and S are symmetric in the direction of $\mathbf{t}$, that is $C(x_t, x_n) = C(-x_t, x_n)$ and $S(x_t, x_n) = S(-x_t, x_n)$ for all $x_t, x_n \in \mathbb{R}$.*

Often one assumes the separability of C and S, each one being the product of a "smoothing" function on x_t and an "edge detection function" on x_n, the "smoothing" being the same for C and S:

$$C(x_t, x_n) = b(x_t) \cdot c(x_n) \qquad \text{and} \qquad S(x_t, x_n) = b(x_t) \cdot s(x_n). \tag{4.2}$$

Usually, b is a bell-shaped function, such as the Gaussian (cfr. (3.21)). Here Requirements 1 and 4 combined give: b and c are even-symmetric, s is odd-symmetric, $\widehat{b}$ and $\widehat{c}$ have real non-negative values, and (c, s) is a quadrature pair. Requirements 2' and 3 mean that b, c, and s are continuous and integrable. By Proposition 4.1, $\widehat{b}$, $\widehat{c}$, and $\widehat{s}$ are integrable.

Several authors (for example [29,30,31]) have considered that either C and S, or $\widehat{C}$ and $\widehat{S}$, are polar-separable, being the product of a radial function and an angular one (the angle being measured w.r.t. $\mathbf{n}$), for example:

$$\begin{aligned} \widehat{C}(x_t, x_n) = R(\rho) \cdot \gamma(\theta) \qquad &\text{and} \qquad S(x_t, x_n) = R(\rho) \cdot \sigma(\theta), \\ \text{with} \qquad x_n = \rho \cos\theta \qquad &\text{and} \qquad x_t = \rho \sin\theta. \end{aligned} \tag{4.3}$$

Here Requirements 1 and 4 combined give: γ is symmetric, has period dividing π, and non-negative values, $\sigma(\theta) = \gamma(\theta)$ for $-\pi/2 < \theta < \pi/2$, and $\sigma(\theta) = -\gamma(\theta)$ for $\pi/2 < \theta < 3\pi/2$. Requirements 2' and 3 mean that R and γ are continuous and integrable. Polar-separable filters have some advantages for edge detection when the orientation of $\mathbf{n}$ is allowed to vary.

We do not require either separability condition (4.2) or (4.3), although it could simplify both theory and implementation; we leave open the possibility of finding interesting non-separable filters.

Well-known models of the receptive field profiles of simple cells in the monkey visual cortex (the "edge and bar detectors" of Hubel and Wiesel [4]), indicate that these profiles are similar to those of Gabor cosine and sine functions [59,60]; in other words the even-symmetric and odd-symmetric functions C and S look like

$$\exp\left[-\frac{x_t^2}{w_t^2} - \frac{x_n^2}{w_n^2}\right] \cdot \cos[2\pi\nu x_n] \qquad \text{and} \qquad \exp\left[-\frac{x_t^2}{w_t^2} - \frac{x_n^2}{w_n^2}\right] \cdot \sin[2\pi\nu x_n] \tag{4.4}$$

respectively, where ν is a fixed frequency, while w_n and w_t are constants measuring the width of these functions in the normal and tangential directions; normally w_t is significantly larger than w_n [31]. The separability condition (4.2) is verified. Note that the two functions given here are not in $\mathbf{n}$-quadrature, in particular the even-symmetric function does not satisfy item (i) of Proposition 4.2; we have already [15] pointed some disadvantages of this fact for edge detection. There have also been criticisms of this model on experimental grounds [58].

From this quantitative model (4.4) we retain only a few qualitative guidelines concerning the shape of C and S:

— The zeroes of C and S form a discrete set of lines parallel to $\mathbf{t}$. In particular, the signs of $C(x_t, x_n)$ and $S(x_t, x_n)$ do not depend on x_t, but only on x_n. We illustrate this in Figure 7.

— For a fixed x_n, the x_n-sections $x_t \to C(x_t, x_n)$ and $x_t \to S(x_t, x_n)$ are "bell-shaped"; in other words $|C(x_t, x_n)|$ and $|S(x_t, x_n)|$ are decreasing in $|x_t|$.

A rationale for these guidelines is that C and S should signal edges in the normal direction $\mathbf{n}$, and not in the tangential direction $\mathbf{t}$, for which they should act only as smoothing filters. We will see later that the behaviour of C and S in the tangential direction is essential for their orientation selectivity.

Let us introduce some further notation. For an integrable function $H : \mathcal{E} \to \mathbb{C}$ and a unit vector $\mathbf{v}$, we define the function $H_{/\mathbf{v}} : \mathbb{R} \to \mathbb{R}$ by

$$H_{/\mathbf{v}}(x) = \int_R H(x \cdot \mathbf{v} + y \cdot \mathbf{w})\, dy, \tag{4.5}$$

where $\mathbf{w}$ is the unit vector orthonormal to $\mathbf{v}$. Note in particular that $H_{/\mathbf{t}}(x_t) = \int_{\mathbb{R}} H(x_t, x_n)\, dx_n$ and $H_{/\mathbf{n}}(x_n) = \int_{\mathbb{R}} H(x_t, x_n)\, dx_t$. By Fubini's theorem, $H_{/\mathbf{v}}$ is integrable, and for every $u \in \mathbb{R}$,

$$[H_{/\mathbf{v}}]^\wedge(u) = \int_{\mathcal{E}} H(x \cdot \mathbf{v} + y \cdot \mathbf{w}) \exp[-2\pi i\, ux]\, dxdy = \widehat{H}(u \cdot \mathbf{v}). \tag{4.6}$$

We can now state our next requirement, which will be used in its full expression later on, when we will consider orientation selectivity. Here we will use it in the restricted case where $\mathbf{v}$ is $\mathbf{n}$:

REQUIREMENT 5. *When $d = 2$, for every unit vector $\mathbf{v}$, $C_{/\mathbf{v}}$ and $S_{/\mathbf{v}}$ are continuous.*

By Proposition 3.7, $C_{/\mathbf{t}}$ and $S_{/\mathbf{t}}$ are a.e. equal to zero; by Requirement 5, they are identically zero, in other words

$$C_{/\mathbf{t}}(x_t) = \int_{\mathbb{R}} C(x_t, x_n)\, dx_n = S_{/\mathbf{t}}(x_t) = \int_{\mathbb{R}} S(x_t, x_n)\, dx_n = 0 \tag{4.7}$$

for all $x_t \in \mathbb{R}$. More generally, by (4.6) and Requirement 5, for a unit vector $\mathbf{v}$, the following three statements are equivalent:

(*i*) $C_{/\mathbf{v}}(t) = 0$ for all $t \in \mathbb{R}$;

(*ii*) $C_{/\mathbf{v}}(t) = 0$ for almost all $t \in \mathbb{R}$;

(*iii*) $\widehat{C}(u \cdot \mathbf{v}) = 0$ for all $u \in \mathbb{R}$.

We have then the following requirement, complementing the previous one:

REQUIREMENT 6. *When $d = 2$, for every unit vector $\mathbf{v} \neq \pm\mathbf{t}$, $C_{/\mathbf{v}}$ is not identically zero, in other words $\widehat{C}(u \cdot \mathbf{v}) \neq 0$ for some $u \in \mathbb{R}$.*

Thus, if one draws in the Fourier plane a line L through the origin, either L is in the direction of $\mathbf{t}$ and $\widehat{C}$ vanishes on L, or L is in another direction, and $\widehat{C}$ does not vanish on L. This agrees with the above guideline (see also Figure 7) concerning the location of the zeroes of C on lines parallel to $\mathbf{t}$.

We have then the following result, which is an immediate consequence of (4.6):

PROPOSITION 4.3. *For $d = 2$, Requirements 5 and 6 imply that for every unit vector $\mathbf{v} \in \mathcal{E}^+_{\mathbf{n}}$, $C_{/\mathbf{v}}$ and $S_{/\mathbf{v}}$ satisfy Requirements 1, 2', and 3 with $d = 1$.*

We can thus apply Propositions 4.1 and 4.2 with $C_{/\mathbf{v}}$ and $S_{/\mathbf{v}}$ instead of C and S.

When C and S satisfy the separability condition (4.2), Requirement 6 for $\mathbf{v} = \mathbf{n}$ implies that $\int_{\mathbb{R}} b \neq 0$; we can then without loss of generality assume that $\int_{\mathbb{R}} b = 1$, and then we get $C_{/\mathbf{n}} = c$ and $S_{/\mathbf{n}} = s$. Thus $C_{/\mathbf{n}}$ and $S_{/\mathbf{n}}$ can be considered as a generalization of c and s in the non-separable case.

Let us now give describe the response of $F = C + i\,S$ on idealized one-dimensional line, step, and roof edges, and the requirements to be demanded for each of them. In the two-dimentional case ($d = 2$), each such edge has normal orientation $\mathbf{n}$ and tangential orientation $\mathbf{t}$, and forms a purely one-dimensional feature. This means that it is an image I which is constant along the direction of $\mathbf{t}$, and forms an edge along the direction of $\mathbf{n}$; it satisfies thus an equation of the form

$$\forall x_t, x_n \in \mathbb{R}, \qquad I(x_t, x_n) = P(x_n) \tag{4.8}$$

for a one-dimensional signal P giving the edge profile, and by Fubini's theorem we get then

$$(I * F)(x_t, x_n) = (P * F_{/\mathbf{n}})(x_n) \tag{4.9}$$

for all $x_t, x_n \in \mathbb{R}$ where $(I * F)(x_t, x_n)$ is defined. On the other hand, in the one-dimensional case ($d = 1$), I reduces to P. Thus by (4.9) and Proposition 4.3 the response on a one-dimensional profile in the two-dimensional case reduces to the one-dimensional case by replacing $F = C + i\,S$

with $F_{/\mathbf{n}} = C_{/\mathbf{n}} + i\, S_{/\mathbf{n}}$, which will satisfy Requirements 1, 2', and 3. Thus the requirements on the response of F on a one-dimensional ideal edge will apply to F for $d = 1$, and to $F_{/\mathbf{n}}$ for $d = 2$.

We will choose several types of one-dimensional edge profiles P having clearly a unique edge located at the line $x_n = 0$, and we will demand that the corresponding energy function $|I * F|^2$ has a local maximum located precisely at the line $x_n = 0$, but no local maximum for $x_n \neq 0$; in other words $|P * F_{/\mathbf{n}}|$ has a unique local maximum, located at the origin. This will lead to a series of necessary conditions (one for each ideal edge profile), that we will call "Spatial Constraints" rather than "Requirements", because they concern only the existence of local maxima of particular functions built from $F_{/\mathbf{n}}$, and do not intervene in the phase congruence model; to our knowledge, these constraints cannot be expressed in terms of Fourier analysis.

We first consider the ideal line. In [13] the one-dimensional profile in the normal direction across an ideal line was modeled as the *Dirac delta distribution* δ defined by $\langle \delta, W\rangle = W(0)$ for any Schwartz function W, so that $\delta * W = W$; considering F as a Schwartz function, this gives $\delta * F = F$ as the response of the one-dimensional filter F on this ideal line profile. In the two-dimensional case, an ideal line in the direction of $\mathbf{t}$ will be the product of a Dirac delta in x_n and the constant 1 in x_t, in other words the tempered distribution λ given by $\langle \lambda, W\rangle = \int_{\mathbb{R}} W(x_t, 0)\, dx_t$ for any Schwartz function W, giving thus $(\lambda * W)(x_t, x_n) = W_{/\mathbf{n}}(x_n)$; considering F as a Schwartz function, this gives $F_{/\mathbf{n}}$ as the one-dimensional response of the two-dimensional filter F on this ideal line in the plane.

As our edge model locates edges at local maxima of the energy function, we demand that this one-dimensional energy function $|F_{/\mathbf{n}}|^2$ has a local maximum at $x_n = 0$ only. As $F_{/\mathbf{n}} = C_{/\mathbf{n}} + i\, S_{/\mathbf{n}}$ is continuous, and $(F_{/\mathbf{n}})^\wedge$ is real-valued non-negative (by (4.1)) and integrable (by Proposition 4.1), Corollary 3.4 implies that $|F_{/\mathbf{n}}(\mathbf{x})| < F_{/\mathbf{n}}(\mathbf{0})$ for $\mathbf{x} \neq \mathbf{0}$; in other words $|F_{/\mathbf{n}}|$ has a global maximum at the origin. Thus the constraint reduces to the following:

SPATIAL CONSTRAINT 1. *$|F_{/\mathbf{n}}|$ for $d = 2$, or $|F|$ for $d = 1$, has no local maximum outside the origin.*

As explained in [13] (but only in the one-dimensional case), this constraint expresses the fact that the edge detector is *idempotent*: it sees the same edges in the edge map as in the original image. See also [15] for further comments on this question.

Let us introduce some further notation; given a continuous and integrable function $f : \mathbb{R} \to \mathcal{C}$, let us write $\Pi[f]$ for the primitive of f vanishing at $-\infty$; in other words $\Pi[f](x) = \int_{-\infty}^{x} f(t)\, dt$. If $\xi f : x \mapsto x f(x)$ is also integrable, we define $\Pi^2[f]$ by

$$\Pi^2[f](x) = \int_{-\infty}^{x} (x - t) f(t)\, dt = x \int_{-\infty}^{x} f(t)\, dt - \int_{-\infty}^{x} t f(t)\, dt,$$

in other words $\Pi^2[f] = \xi \Pi[f] - \Pi[\xi f]$. It is easily seen that $\Pi^2[f] = \Pi[\Pi[f]]$, either because it vanishes at $-\infty$ and its derivative is $\Pi[f]$, or by using Fubini's theorem:

$$\Pi^2[f](x) = \int_{-\infty}^{x} dt\, f(t)(x - t) = \int_{-\infty}^{x} dt\, f(t) \int_{t}^{x} du = \iint_{t \le u \le x} dt du\, f(t)$$
$$= \int_{-\infty}^{x} du \int_{-\infty}^{u} dt\, f(t) = \int_{-\infty}^{x} du\, \Pi[f](u) = \Pi[\Pi[f]](x).$$

Now let us define the ideal step. It is an ordinary function I (rather than a tempered distribution), defined by the one-dimensional profile function P given by $P(x_n) = a$ for $x_n > 0$ and $P(x_n) = b$ for $x_n < 0$, where $a \neq b$ are two constants; by Proposition 4.2 (i) and the linearity of convolution, we see easily that $I * F = P * F_{/\mathbf{n}} = (b-a) \cdot H * F_{/\mathbf{n}}$, where H is the *Heavyside step function* given by $H(x_n) = 1$ for $x_n \geq 0$ and $H(x_n) = 0$ for $x_n < 0$. Thus the maxima of $|P * F_{/\mathbf{n}}|$ are those of $|H * F_{/\mathbf{n}}|$. Now we have

$$(H * F_{/\mathbf{n}})(x_n) = \int_{-\infty}^{x_n} F_{/\mathbf{n}}(t)\, dt = \Pi[F_{/\mathbf{n}}](x_n).$$

We demand that $|\Pi[F_{/\mathbf{n}}]|$ has a unique local maximum located at the origin. Since $|\Pi[F_{/\mathbf{n}}]|^2 = (\Pi[C_{/\mathbf{n}}])^2 + (\Pi[S_{/\mathbf{n}}])^2$, while $C_{/\mathbf{n}}$ and $S_{/\mathbf{n}}$ are the derivatives of $\Pi[C_{/\mathbf{n}}]$ and $\Pi[S_{/\mathbf{n}}]$, we make the slightly stronger requirement that the derivative

$$2C_{/\mathbf{n}}\Pi[C_{/\mathbf{n}}] + 2S_{/\mathbf{n}}\Pi[S_{/\mathbf{n}}]$$

of $|\Pi[F_{/\mathbf{n}}]|^2$ is equal to 0 only at the origin:

SPATIAL CONSTRAINT 2. *For $d = 2$ we have*

$$\big(C_{/\mathbf{n}}\Pi[C_{/\mathbf{n}}] + S_{/\mathbf{n}}\Pi[S_{/\mathbf{n}}]\big)(x_n) \begin{cases} < 0 & \text{for } x_n > 0; \\ = 0 & \text{for } x_n = 0; \\ > 0 & \text{for } x_n < 0. \end{cases}$$

For $d = 1$ we have the same with C and S instead of $C_{/\mathbf{n}}$ and $S_{/\mathbf{n}}$.

Let us finally consider ideal roofs and Mach bands. Here we have a profile function P given by $P(x_n) = ax_n$ for $x_n > 0$ and $P(x_n) = bx_n$ for $x_n < 0$, where $a \neq b$ are two constants; in order for the convolution $(I * F)(x_t, x_n)$ to be defined for any such a, b, it is necessary and sufficient that the function $\xi_\mathbf{n} F : (x_t, x_n) \mapsto x_n \cdot F(x_t, x_n)$ be integrable (or equivalently, both $\xi_\mathbf{n} C$ and $\xi_\mathbf{n} S$ are integrable). The following result is obtained by applying successively Propositions 3.8 and 3.7 (with Fubini's theorem):

LEMMA 4.4. *If $\xi_\mathbf{n} C$ and $\xi_\mathbf{n} S$ are integrable, then they are in quadrature, and we have*

$$\int_{\mathcal{E}} x_n \cdot F(x_t, x_n)\, dx_t dx_n = \int_{\mathbb{R}} x_n \cdot F_{/\mathbf{n}}(x_n)\, dx_n = 0.$$

In particular $I * F = P * F_{/\mathbf{n}} = 0$ for $I(x_t, x_n) = P(x_n) = cx_n + d$, with c and d constants.

Let R be the ramp defined $R(x_n) = x_n$ for $x_n > 0$ and $R(x_n) = 0$ for $x_n < 0$; then by Lemma 4.4, for the above profile P given by $P(x_n) = ax_n$ for $x_n > 0$ and $P(x_n) = bx_n$ for $x_n < 0$, where a and b are arbitrary constants, we have then $P * F_{/\mathbf{n}} = (b-a) \cdot R * F_{/\mathbf{n}}$, and the maxima of $|P * F_{/\mathbf{n}}|$ are those of $|R * F_{/\mathbf{n}}|$. We have

$$(R * F_{/\mathbf{n}})(x_n) = \int_{-\infty}^{x_n} (x_n - t) F_{/\mathbf{n}}(t)\, dt = \Pi^2[F_{/\mathbf{n}}](x_n).$$

We make then the same requirement on the derivative of

$$|R * F_{/\mathbf{n}}|^2 = |\Pi^2[F_{/\mathbf{n}}]|^2 = (\Pi^2[C_{/\mathbf{n}}])^2 + (\Pi^2[S_{/\mathbf{n}}])^2$$

as we did for the derivative of $|\Pi[F_{/\mathbf{n}}]|^2$:

Spatial Constraint 3. *Both $\xi_{\mathbf{n}} C$ and $\xi_{\mathbf{n}} S$ are integrable. For $d = 2$ we have*

$$\big(\Pi[C_{/\mathbf{n}}]\Pi^2[C_{/\mathbf{n}}] + \Pi[S_{/\mathbf{n}}]\Pi^2[S_{/\mathbf{n}}]\big)(x_n) \begin{cases} < 0 & \text{for } x_n > 0; \\ = 0 & \text{for } x_n = 0; \\ > 0 & \text{for } x_n < 0. \end{cases}$$

For $d = 1$ we have the same with C and S instead of $C_{/\mathbf{n}}$ and $S_{/\mathbf{n}}$.

As seen above with Lemma 4.4, the fact that $\xi_{\mathbf{n}} C$ and $\xi_{\mathbf{n}} S$ are integrable implies that all ideal ramp edges and Mach bands give, up to a constant factor, the same energy function $|I * F|^2$; in particular the same edges should be detected, at the same locations. We said in Section 2 that there is not a complete agreement as to the exact position of the line perceived by human observers in a Mach band or roof edge [23,21]. This indicates that the edge detectors of the human visual system may have a non-zero response to the underlying linear ramp, contradicting the conclusion of Lemma 4.4. Therefore the above third constraint is less important than the two previous ones.

Other spatial constraints could be envisaged, for example that $F_{/\mathbf{n}}$ is derivable and $|(F_{/\mathbf{n}})'|$ has a unique local maximum at the origin.

One choice of filters C and S for which these three spatial constraints are satisfied is by taking $C_{/\mathbf{n}}$ and $S_{/\mathbf{n}}$ equal to the Cauchy functions given in (3.33), in other words $F_{/\mathbf{n}}(x) = 1/(1-ix)^{m+1}$ for $m \geq 2$. Here the primitives of $F_{/\mathbf{n}}$ take the form $\Pi[F_{/\mathbf{n}}](x) = (1/im) \cdot 1/(1-ix)^m$ and $\Pi^2[F_{/\mathbf{n}}](x) = (1/m(1-m)) \cdot 1/(1-ix)^{m-1}$, so that $|(F_{/\mathbf{n}})(x)|^2$, $|\Pi[F_{/\mathbf{n}}](x)|^2$, and $|\Pi^2[F_{/\mathbf{n}}](x)|^2$ are of the form $\alpha/(1+x^2)^t$ (α constant, $t \geq 1$), a function which has a local maximum for $x = 0$ only. The same holds also for the derivatives of $F_{/\mathbf{n}}$.

4.2. *Phase congruence*

We take an image I in $L^1 + L^2$. As said at the beginning of this section, we define $F = C + i\,S$, $J = I * F = (I * C) + i\,(I * S)$, and the energy function $E = |J|^2 = (I * C)^2 + (I * S)^2$. From (4.1), we write $\widehat{C} = A$, and then we have $\widehat{S} = \operatorname{sgn}_{\mathbf{n}} A$. We will see now that at every point $\mathbf{p}$ the energy function $E(\mathbf{p})$ measures the degree to which all phases $I^\Phi(\mathbf{u}, \mathbf{p})$ for frequencies $\mathbf{u} \in \mathcal{E}_{\mathbf{n}}^+$ are clustered around a unique value.

Proposition 4.5. *Assume that I is in $L^1 + L^2$, in other words $I = I_1 + I_2$, where I_1 is in L^1 and I_2 is in L^2. Then $\widehat{E} = 4\,(\operatorname{pos}_{\mathbf{n}} \widehat{I} A) * (\operatorname{neg}_{\mathbf{n}} \widehat{I} A)$, an integrable function. Furthermore:*

- (i) *If I^Φ is constant on $\mathcal{E}_{\mathbf{n}}^+$, then $\widehat{E} = 4\,(\operatorname{pos}_{\mathbf{n}} I^{\mathcal{A}} A) * (\operatorname{neg}_{\mathbf{n}} I^{\mathcal{A}} A)$, a real-valued non-negative function, and E^Φ is constant zero.*
- (ii) *If for a given $\mathbf{p} \in \mathcal{E}$ we have $I^\Phi(\mathbf{u}, \mathbf{p})$ constant for all $\mathbf{u} \in \mathcal{E}_{\mathbf{n}}^+$, then $E^\Phi(\mathbf{u}, \mathbf{p}) = 0$ for all $\mathbf{u} \in \mathcal{E}$, and $E(\mathbf{p}) > E(\mathbf{x})$ for all $\mathbf{x} \in \mathcal{E}$ such that $\mathbf{x} \neq \mathbf{p}$.*
- (iii) *For every $\mathbf{p} \in \mathcal{E}$ we have*

$$E(\mathbf{p}) = 4 \iint_{\mathcal{E}_{\mathbf{n}}^+ \times \mathcal{E}_{\mathbf{n}}^+} I^{\mathcal{A}}(\mathbf{u}) A(\mathbf{u}) I^{\mathcal{A}}(\mathbf{v}) A(\mathbf{v}) \cos\big[I^\Phi(\mathbf{u}, \mathbf{p}) - I^\Phi(\mathbf{v}, \mathbf{p})\big]\, d\mathbf{u} d\mathbf{v}. \tag{4.10}$$

Proof. By Proposition 4.2 we have $\widehat{J} = \widehat{I}\widehat{F} = 2\operatorname{pos}_{\mathbf{n}} A\widehat{I}$ and $\widehat{\overline{J}} = 2\operatorname{neg}_{\mathbf{n}} A\widehat{I}$, so that $\widehat{E} = 4\,(\operatorname{pos}_{\mathbf{n}} \widehat{I} A) * (\operatorname{neg}_{\mathbf{n}} \widehat{I} A)$ by Lemma 3.13 (i) (see also (3.36)). By Proposition 4.1, $\widehat{J}$ and $\widehat{\overline{J}}$ are integrable, so that $\widehat{E}$, being the convolution of two integrable functions, is integrable.

Now $\widehat{J}(\mathbf{u}) = 0$ for $\mathbf{u} \notin \mathcal{E}_{\mathbf{n}}^+$, while for $\mathbf{u} \in \mathcal{E}_{\mathbf{n}}^+$ we have $J^{\mathcal{A}}(\mathbf{u}) = 2\,A(\mathbf{u}) I^{\mathcal{A}}(\mathbf{u})$ and $J^\Phi(\mathbf{u}) = I^\Phi(\mathbf{u})$. By Lemma 3.13 ($i$), if I^Φ is constant on $\mathcal{E}_{\mathbf{n}}^+$, we get $\widehat{E} = J^{\mathcal{A}} * (J^{\mathcal{A}})_\rho = 4\,(\operatorname{pos}_{\mathbf{n}} I^{\mathcal{A}} A) * (\operatorname{neg}_{\mathbf{n}} I^{\mathcal{A}} A)$, and since $\widehat{E}$ is non-negative real, E^Φ is constant zero. Therefore (i) follows.

By definition, the phase of I at point $\mathbf{p}$ is the Fourier phase of $\tau_{-\mathbf{p}}(I)$. Now $\tau_{-\mathbf{p}}(I)$ leads to the energy function $\tau_{-\mathbf{p}}(E)$. Suppose that $\tau_{-\mathbf{p}}(I)$ has constant Fourier phase. Then by (*i*), $\tau_{-\mathbf{p}}(E)$ has zero Fourier phase. Since $\tau_{-\mathbf{p}}(E)$ is continuous and the Fourier transform of $\tau_{-\mathbf{p}}(E)$ is integrable, Corollary 3.4 implies that $\tau_{-\mathbf{p}}(E)(\mathbf{0}) > \tau_{-\mathbf{p}}(E)(\mathbf{z})$ for all $\mathbf{z} \neq \mathbf{0}$. Therefore (*ii*) holds.

Finally (*iii*) follows from (3.35) and (3.37). ∎

This result, especially (*iii*), is at the basis of the model. Indeed, at a given point $\mathbf{p}$, the more all Fourier phases $I^\Phi(\mathbf{u},\mathbf{p}) = I^\Phi(\mathbf{u}) + 2\pi\,\mathbf{p}\cdot\mathbf{u}$ for $\mathbf{u} \in \mathcal{E}^+_\mathbf{n}$ are clustered together, the more all $\cos\big[I^\Phi(\mathbf{u},\mathbf{p}) - I^\Phi(\mathbf{v},\mathbf{p})\big]$ appearing in (4.10) are close to 1, and the higher is $E(\mathbf{p})$. Therefore maxima of E correspond to *points of maximal phase congruence* for frequencies in the half-plane $\mathcal{E}^+_\mathbf{n}$. In particular by (*ii*), a point where all phases become strictly equal gives an absolute maximum of E. Thus edges whose normal direction is parallel to $\mathbf{n}$ will be localized at points where E has a maximum in the normal direction. Whether such maxima are purely local or should be over a certain range will be discussed later.

The next result shows how the value around which all phases at $\mathbf{p}$ are clustered is obtained from the argument of the complex number $J(\mathbf{p})$. This generalizes a similar finding in [14] in the case of one-dimensional periodic signals. For any $\mathbf{x} \in \mathcal{E}$, define the angle $\varphi(\mathbf{x})$ (uniquely modulo 2π) by

$$J(\mathbf{x}) = |J(\mathbf{x})| \cdot e^{i\,\varphi(\mathbf{x})}, \tag{4.11}$$

in other words

$$(I*C)(\mathbf{x}) = |J(\mathbf{x})| \cdot \cos\varphi(\mathbf{x}) \qquad \text{and} \qquad (I*S)(\mathbf{x}) = |J(\mathbf{x})| \cdot \sin\varphi(\mathbf{x}). \tag{4.12}$$

This will be the angle around which all phases cluster:

PROPOSITION 4.6. *For every* $\mathbf{p} \in \mathcal{E}$,

$$|J(\mathbf{p})| \cdot e^{i\,\varphi(\mathbf{p})} = 2\int_{\mathcal{E}^+_\mathbf{n}} I^\mathcal{A}(\mathbf{u})A(\mathbf{u}) \exp\big[i\, I^\Phi(\mathbf{u},\mathbf{p})\big]\, d\mathbf{u}, \tag{4.13}$$

$$\int_{\mathcal{E}^+_\mathbf{n}} I^\mathcal{A}(\mathbf{u})A(\mathbf{u}) \sin\big[I^\Phi(\mathbf{u},\mathbf{p}) - \varphi(\mathbf{p})\big]\, d\mathbf{u} = 0, \tag{4.14}$$

and

$$2\int_{\mathcal{E}^+_\mathbf{n}} I^\mathcal{A}(\mathbf{u})A(\mathbf{u}) \cos\big[I^\Phi(\mathbf{u},\mathbf{p}) - \theta\big]\, d\mathbf{u} \begin{cases} = |J(\mathbf{p})| & \theta = \varphi(\mathbf{p}); \\ < |J(\mathbf{p})| & \text{for } \theta \neq \varphi(\mathbf{p}). \end{cases} \tag{4.15}$$

Furthermore if for all $\mathbf{u} \in \mathcal{E}^+_\mathbf{n}$, $I^\Phi(\mathbf{u},\mathbf{p}) = \theta$, *where* θ *is constant, then* $\theta = \varphi(\mathbf{p})$.

PROOF. (4.13) follows from the inverse Fourier transform formula (3.11), combined with (3.14) and (4.1). Now (4.13) gives

$$\begin{aligned} |J(\mathbf{p})| &= 2\exp\big[-i\,\varphi(\mathbf{p})\big] \int_{\mathcal{E}^+_\mathbf{n}} I^\mathcal{A}(\mathbf{u})A(\mathbf{u}) \exp\big[i\, I^\Phi(\mathbf{u},\mathbf{p})\big]\, d\mathbf{u} \\ &= 2\int_{\mathcal{E}^+_\mathbf{n}} I^\mathcal{A}(\mathbf{u})A(\mathbf{u}) \exp\big[i\,\big(I^\Phi(\mathbf{u},\mathbf{p}) - \varphi(\mathbf{p})\big)\big]\, d\mathbf{u}. \end{aligned}$$

The imaginary part of this equality gives (4.14), and the real part of it gives the equality in (4.15) for $\theta = \varphi(\mathbf{p})$. Now for $\theta \neq \varphi(\mathbf{p})$, we have

$$|J(\mathbf{p})| > \Re\big(e^{i\,(\varphi(\mathbf{p})-\theta)}|J(\mathbf{p})|\big) = \Re\big(e^{-i\,\theta} J(\mathbf{p})\big) = \Re\Big(2e^{-i\,\theta} \int_{\mathcal{E}^+_\mathbf{n}} I^\mathcal{A}(\mathbf{u})A(\mathbf{u}) \exp\big[i\, I^\Phi(\mathbf{u},\mathbf{p})\big]\, d\mathbf{u}\Big)$$

$$=2\Re\Big(\int_{\mathcal{E}^+_\mathbf{n}} I^\mathcal{A}(\mathbf{u})A(\mathbf{u}) \exp\big[i\,\big(I^\Phi(\mathbf{u},\mathbf{p}) - \theta\big)\big]\, d\mathbf{u}\Big) = 2\int_{\mathcal{E}^+_\mathbf{n}} I^\mathcal{A}(\mathbf{u})A(\mathbf{u}) \cos\big[I^\Phi(\mathbf{u},\mathbf{p}) - \theta\big]\, d\mathbf{u},$$

giving the inequality in (4.15). Finally if $I^\Phi(\mathbf{u},\mathbf{p}) = \theta$ for all $\mathbf{u} \in \mathcal{E}^+_\mathbf{n}$, then (4.13) gives

$$|J(\mathbf{p})| \cdot e^{i\,\varphi(\mathbf{p})} = 2e^{i\,\theta} \int_{\mathcal{E}^+_\mathbf{n}} I^\mathcal{A}(\mathbf{u})A(\mathbf{u})\,d\mathbf{u},$$

from which we deduce that $\theta = \varphi(\mathbf{p})$. ∎

Let us comment this result, and the equations within it. Equation (4.13) shows that $e^{i\,\varphi(\mathbf{p})}$ is a weighted linear combination (with non-negative weights) of all $\exp\bigl[i\,I^\Phi(\mathbf{u},\mathbf{p})\bigr]$ for $\mathbf{u} \in \mathcal{E}^+_\mathbf{n}$; in particular when all $I^\Phi(\mathbf{u},\mathbf{p})$ are equal, they must coincide with $\varphi(\mathbf{p})$. Equation (4.14) is another way of expressing that $\varphi(\mathbf{p})$ is some form of average between the $I^\Phi(\mathbf{u},\mathbf{p})$, $\mathbf{u} \in \mathcal{E}^+_\mathbf{n}$. Finally (4.15) implies that

$$|J(\mathbf{p})| = 2 \max_{\theta\in[0,2\pi[} \int_{\mathcal{E}^+_\mathbf{n}} I^\mathcal{A}(\mathbf{u})A(\mathbf{u}) \cos\bigl[I^\Phi(\mathbf{u},\mathbf{p}) - \theta\bigr]\,d\mathbf{u}. \tag{4.16}$$

Here, the closer are all $I^\Phi(\mathbf{u},\mathbf{p})$ to θ, the higher are all $\cos\bigl[I^\Phi(\mathbf{u},\mathbf{p}) - \theta\bigr]$, and the higher is the resulting integral. This means that $\varphi(\mathbf{p})$ is the angle which is on the average the closest to each $I^\Phi(\mathbf{u},\mathbf{p})$, in other words the angle of maximum congruence of the phases $I^\Phi(\mathbf{u},\mathbf{p})$ at $\mathbf{p}$ for $\mathbf{u} \in \mathcal{E}^+_\mathbf{n}$. We call thus $\varphi(\mathbf{p})$ the *average phase at* $\mathbf{p}$.

We define the *phase congruence function* of image I as the function $\bigl(\int_\mathcal{E} I^\mathcal{A} A\bigr)^{-1} \cdot |J|$; indeed (4.16) and the symmetry of $I^\mathcal{A} A$ give:

$$\frac{|J(\mathbf{p})|}{\int_\mathcal{E} I^\mathcal{A} A} = \max_{\theta\in[0,2\pi[} \frac{\int_{\mathcal{E}^+_\mathbf{n}} I^\mathcal{A}(\mathbf{u})A(\mathbf{u}) \cos\bigl[I^\Phi(\mathbf{u},\mathbf{p}) - \theta\bigr]\,d\mathbf{u}}{\int_{\mathcal{E}^+_\mathbf{n}} I^\mathcal{A}(\mathbf{u})A(\mathbf{u})\,d\mathbf{u}}. \tag{4.17}$$

This function measures the degree to which all Fourier phases at $\mathbf{p}$ are concentrated around $\varphi(\mathbf{p})$. Clearly it does not change when I is multiplied by a constant factor, and it takes values in the interval $[-1, 1]$. It has value 1 at points $\mathbf{p}$ where all $I^\Phi(\mathbf{u},\mathbf{p})$ for $\mathbf{u} \in \mathcal{E}^+_\mathbf{n}$ are equal, in other words I has constant Fourier phase at $\mathbf{p}$. Such a phase congruence function was considered in [14] in the case of one-dimensional periodic signals.

Note that in all the above equations, to each frequency $\mathbf{u}$ corresponds the non-negative weight $I^\mathcal{A}(\mathbf{u})A(\mathbf{u})$; thus the value of the average phase $\varphi(\mathbf{p})$, and the existence of a maximum of E at point $\mathbf{p}$, depend not only on the phases $I^\Phi(\mathbf{u},\mathbf{p})$, but also on the amplitude spectra of both I and C. In particular, it is sensitive to the choice of C and S; another pair of filters satisfying the same requirements could lead to other values for $\varphi(\mathbf{p})$, and to other positions for the maxima of the energy function E. However this change should in general be moderate, since the profile of our filters and of their Fourier transforms is highly constrained by the requirements and spatial constraints we imposed on them.

Let us now see how behave Fourier phases of edge profiles described in Section 2, in particular whether such edges will be properly detected and localized by the phase congruence model. As when dealing with the spatial constraints in the preceding subsection, we assume that we have an image I forming a one-dimensional profile P, namely $I(x_t, x_n) = P(x_n)$ as in (4.8). Given (4.9) and Proposition 4.3, we can apply then the phase congruence model with P and $F/\mathbf{n}$ instead of I and F, and so we have only to look at the Fourier transform of P. For the sake of simplicity, we assume that P is integrable.

We consider first an ideal line edge profile. We have the following:

PROPOSITION 4.7. *Let $P : \mathbb{R} \to \mathbb{R}$ be integrable, even-symmetric, having non-negative values, and convex on $\mathbb{R}^+$, in other words such that for $0 \leq a < b$ and $0 \leq \lambda \leq 1$ we have $P\big(\lambda a + (1-\lambda)b\big) \leq \lambda P(a) + (1-\lambda)P(b)$. Then P has constant zero Fourier phase.*

PROOF. Since P is even-symmetric, we have

$$\widehat{P}(\nu) = 2\int_0^\infty P(x)\cos[2\pi\nu x]\,dx. \tag{4.18}$$

Obviously $\widehat{P}(0) \geq 0$ since $P \geq 0$. For $\nu > 0$, let us define the period $L = 1/\nu$. Take any integer $n \geq 0$. The change of variable $y = x - nL$ ($x = nL + y$) gives

$$\int_{nL}^{nL+L/4} P(x)\cos[2\pi\nu x]\,dx = \int_0^{L/4} P(nL+y)\cos[2\pi\nu y]\,dy;$$

the change of variable $y = nL + L/2 - x$ ($x = nL + L/2 - y$) gives

$$\int_{nL+L/4}^{nL+L/2} P(x)\cos[2\pi\nu x]\,dx = -\int_0^{L/4} P(nL+L/2-y)\cos[2\pi\nu y]\,dy;$$

the change of variable $y = x - nL - L/2$ ($x = nL + L/2 + y$) gives

$$\int_{nL+L/2}^{nL+3L/4} P(x)\cos[2\pi\nu x]\,dx = -\int_0^{L/4} P(nL+L/2+y)\cos[2\pi\nu y]\,dy;$$

the change of variable $y = nL + L - x$ ($x = nL + L - y$) gives

$$\int_{nL+3L/4}^{nL+L} P(x)\cos[2\pi\nu x]\,dx = \int_0^{L/4} P(nL+L-y)\cos[2\pi\nu y]\,dy.$$

Adding the four equations, we get thus

$$\begin{aligned}&\int_{nL}^{nL+L} P(x)\cos[2\pi\nu x]\,dx = \\ &\int_0^{L/4}\big[P(nL+y)+P(nL+L-y)-P(nL+L/2-y)-P(nL+L/2+y)\big]\cos[2\pi\nu y]\,dy.\end{aligned} \tag{4.19}$$

Now for every $y \in [0, L/4]$, by taking $\lambda = (L-4y)/(2L-4y)$, we have $0 \leq \lambda \leq 1$, $1-\lambda = L/(2L-4y)$, and also

$$\lambda(nL+y)+(1-\lambda)(nL+L-y) = nL+L/2+y \qquad \text{and} \qquad (1-\lambda)(nL+y)+\lambda(nL+L-y) = nL+L/2-y;$$

since $nL+y, nL+L/2-y, nL+L/2+y, nL+L-y \geq 0$, the convexity condition implies that

$$\begin{aligned}&\lambda P(nL+y)+(1-\lambda)P(nL+L-y) \geq P(nL+L/2+y)\\ \text{and}\qquad &(1-\lambda)P(nL+y)+\lambda P(nL+L-y) \geq P(nL+L/2-y);\end{aligned}$$

adding the two inequations, and subtracting the resulting right member from the left one, we get:

$$P(nL+y)+P(nL+L-y)-P(nL+L/2-y)-P(nL+L/2+y) \geq 0.$$

Combining this with (4.19), we get

$$\int_{nL}^{nL+L} P(x)\cos[2\pi\nu x]\,dx \geq 0.$$

Since this holds for every integer $n \geq 0$, (4.18) gives $\widehat{P}(\nu) \geq 0$ for all $\nu > 0$, and so P has constant zero phase. ∎

If we look at the ideal line edge profile shown in Figure 1, we see that it satisfies the hypothesis of Proposition 4.7; this result implies then that such a profile has constant zero phase, and so by Proposition 4.5, the energy function will have an absolute maximum at the edge position. On the other hand, the ideal bar profile shown next to it in Figure 1 will have a Fourier transform of the form $\widehat{P}(\nu) = \sin(a\nu)/(b\nu)$, where a and b are positive constants, so that its Fourier phase alternates between 0 and π. We explained in Section 2 that following Horn's model [2], we considered the line as physically more realistic than the bar. We see here that this physical choice has also a mathematical advantage.

Note that there are other line profiles which do not satisfy the hypothesis of Proposition 4.7, but which have nevertheless constant zero phase (for example a Gaussian).

Let us now consider ideal step edge profiles. We have the following:

PROPOSITION 4.8. *Let $P : \mathbb{R} \to \mathbb{R}$ be integrable, odd-symmetric, having decreasing non-negative values on $\mathbb{R}^+$, in other words such that for $0 \leq a < b$ we have $P(a) \geq P(b) \geq 0$. Then P has constant $-\pi/2$ Fourier phase.*

PROOF. Since P is odd-symmetric, we have

$$\widehat{P}(\nu) = -2i \int_0^\infty P(x) \sin[2\pi\nu x]\, dx. \tag{4.20}$$

In particular $\widehat{P}(0) = 0$. For $\nu > 0$, let us define the period $L = 1/\nu$. Take any integer $n \geq 0$. The change of variable $y = x - nL$ ($x = nL + y$) gives

$$\int_{nL}^{nL+L/2} P(x) \sin[2\pi\nu x]\, dx = \int_0^{L/2} P(nL + y) \sin[2\pi\nu y]\, dy;$$

the change of variable $y = x - nL - L/2$ ($x = nL + L/2 + y$) gives

$$\int_{nL+L/2}^{nL+L} P(x) \sin[2\pi\nu x]\, dx = -\int_0^{L/2} P(nL + L/2 + y) \sin[2\pi\nu y]\, dy.$$

Adding both equations, we get

$$\int_{nL}^{nL+L} P(x) \sin[2\pi\nu x]\, dx = \int_0^{L/2} \big[P(nL + y) - P(nL + L/2 + y)\big] \sin[2\pi\nu y]\, dy,$$

and as $P(nL + y) \geq P(nL + L/2 + y)$ for all $y \in [0, L/2]$ (since P is decreasing on R^+), we get

$$\int_{nL}^{nL+L} P(x) \sin[2\pi\nu x]\, dx \geq 0.$$

Since this holds for every integer $n \geq 0$, (4.20) gives $\mathrm{sgn}(\widehat{P}(\nu)) = -i$ for all $\nu > 0$, and so P has constant $-\pi/2$ phase. ∎

It follows that an odd-symmetric *sharp* step, where the grey-level jumps discontinuously from negative to positive, and then decreases, will have constant $-\pi/2$ phase. This is however not the necessarily the case with the gradual step shown in Figure 1.

Note that there are other step profiles which do not satisfy the hypothesis of Proposition 4.8, but which have nevertheless constant $-\pi/2$ phase (for example minus the derivative of a Gaussian).

We consider now compound edges consisting of the linear superposition of a line and a step. We can assume that the line and step have constant phases 0 and $-\pi/2$ (resp. $\pi/2$), so that all phases of the compound edge will be in the same quadrant $[-\pi/2, 0]$ (resp. $[0, \pi/2]$). Thus the congruence of phases will be relatively high at that edge position, but we might have a higher congruence at a neighbouring location. Consider for example a middle row in Figure 3, whose grey-level profile, the addition of a square wave and a triangular wave, is shown in Figure 4. As a periodic function of x, it can be decomposed as a Fourier series of the form

$$a + \sum_{n=1}^{\infty} \big(\frac{\alpha}{n^2} \cos[2\pi n f x] - \frac{\beta}{n} \sin[2\pi n f x] \big),$$

where f is the fundamental frequency, a is a constant corresponding to the frequency 0, and the constants $\alpha, \beta > 0$ determine the degree of mixture between the two waves; thus the Fourier coefficient for frequency nf will be $\frac{\alpha}{n^2} + i\frac{\alpha}{n}$, and in particular the corresponding Fourier phase will be $\arctan[n\beta/\alpha]$. Hence the phase increases with frequency, so that for small $\varepsilon > 0$, the phases

$$P^{\Phi}(nf, -\varepsilon) = \arctan\big[\frac{n\beta}{\alpha}\big] - 2\pi n f \varepsilon$$

at $-\varepsilon$ will generally be more congruent than at the origin. This accords with our visual perception of Figure 3, where in a middle row the feature appears as a combination of an edge and of a Mach band extending slightly to the left of the feature's true position.

Therefore the edge localization given by the phase congruence model can be slightly to the left (or to the right) of the true edge position, but this is not a serious problem, since this offset is generally small. As we will see in the next subsection, a much worse problem would arise with usual methods where one applies the two filters separately, the even-symmetric one in order to detect lines, and the odd-symmetric one in order to detect steps: the two edges detected by both filters would not coincide.

4.3. Other quadratic operators, and the relation to classical edge detectors

We will study here all quadratic combinations of $I * C$ and $I * S$, and give their interpretation in terms of phases and phase congruence; in particular we will consider traditional edge detectors using a single filter, and show that they lead to an edge model which is a restricted form of the phase congruence approach.

Let us define

$$\Gamma = I * C \qquad \text{and} \qquad \Sigma = I * S, \tag{4.21}$$

so that $J = \Gamma + i\,\Sigma$. We already have the quadratic operator associating to the image I its energy $E = |J|^2 = \Gamma^2 + \Sigma^2$. We introduce two new quadratic combinations of Γ and Σ:

$$X = \Gamma^2 - \Sigma^2 = \Re(J^2) \qquad \text{and} \qquad Y = 2\Gamma\Sigma = \Im(J^2), \tag{4.22}$$

in other words

$$X + i\,Y = J^2 = (\Gamma + i\,\Sigma)^2.$$

We can also write

$$X = \frac{J^2 + \overline{J}^2}{2} \qquad \text{and} \qquad Y = \frac{J^2 - \overline{J}^2}{2i}.$$

We have then

$$X^2 + Y^2 = |J^2|^2 = E^2. \tag{4.23}$$

Note that since Γ and Σ are in $\mathbf{n}$-quadrature, $X = \Gamma \odot \Gamma$ and $Y = \Gamma \odot \Sigma$ according to the definition of $\odot$ by [55] (cfr. Subsection 3.5). We have the following interpretation of X and Y in terms of phases:

PROPOSITION 4.9. *Assume that I is in $L^1 + L^2$, in other words $I = I_1 + I_2$, where I_1 is in L^1 and I_2 is in L^2. Then X and Y are in $\mathbf{n}$-quadrature, and $\widehat{X}$ and $\widehat{Y}$ are integrable. Furthermore:*

- *(i) If for a given $\mathbf{p} \in \mathcal{E}$ we have $I^\Phi(\mathbf{u}, \mathbf{p}) = \phi$ for all $\mathbf{u} \in \mathcal{E}_\mathbf{n}^+$, then $X^\Phi(\mathbf{u}, \mathbf{p}) = 2\phi$ and $Y^\Phi(\mathbf{u}, \mathbf{p}) = 2\phi - \pi/2$ for all $\mathbf{u} \in \mathcal{E}_\mathbf{n}^+$.*
- *(ii) For every $\mathbf{p} \in \mathcal{E}$ we have*

$$\begin{aligned} X(\mathbf{p}) &= 4 \iint_{\mathcal{E}_\mathbf{n}^+ \times \mathcal{E}_\mathbf{n}^+} I^\mathcal{A}(\mathbf{u}) A(\mathbf{u}) I^\mathcal{A}(\mathbf{v}) A(\mathbf{v}) \, \cos\bigl[I^\Phi(\mathbf{u}, \mathbf{p}) + I^\Phi(\mathbf{v}, \mathbf{p})\bigr] \, d\mathbf{u} d\mathbf{v}; \\ Y(\mathbf{p}) &= 4 \iint_{\mathcal{E}_\mathbf{n}^+ \times \mathcal{E}_\mathbf{n}^+} I^\mathcal{A}(\mathbf{u}) A(\mathbf{u}) I^\mathcal{A}(\mathbf{v}) A(\mathbf{v}) \, \sin\bigl[I^\Phi(\mathbf{u}, \mathbf{p}) + I^\Phi(\mathbf{v}, \mathbf{p})\bigr] \, d\mathbf{u} d\mathbf{v}. \end{aligned} \tag{4.24}$$

PROOF. We refer to the proof of Proposition 4.5; thus $\widehat{J} = 2\mathrm{pos}_\mathbf{n} A\widehat{I}$, $(J^2)^\wedge = \widehat{J} * \widehat{J}$, and the latter is integrable. As $J^2 = X + i\,Y$, $(J^2)^\wedge = \widehat{X} + i\,\widehat{Y}$, and so both $\widehat{X}$ and $\widehat{Y}$ are integrable. Since $\widehat{J}$ vanished outside $\mathcal{E}_\mathbf{n}^+$, so does $(J^2)^\wedge = \widehat{J} * \widehat{J}$, and from Subsection 3.5 it follows that X and Y are in $\mathbf{n}$-quadrature.

If $I^\Phi(\mathbf{u}) = \phi$ for all $\mathbf{u} \in \mathcal{E}_\mathbf{n}^+$, then $\widehat{J} = 2e^{i\phi}\mathrm{pos}_\mathbf{n} AI^\mathcal{A}$ and $\widehat{X} + i\,\widehat{Y} = (J^2)^\wedge = 4e^{2i\phi}(\mathrm{pos}_\mathbf{n} AI^\mathcal{A}) * (\mathrm{pos}_\mathbf{n} AI^\mathcal{A})$, from which we get (with (3.25)) that

$$\widehat{X}(\mathbf{u}) = 2e^{2i\phi}\Bigl[(\mathrm{pos}_\mathbf{n} AI^\mathcal{A}) * (\mathrm{pos}_\mathbf{n} AI^\mathcal{A})\Bigr](\mathbf{u}),$$

for all $\mathbf{u} \in \mathcal{E}_\mathbf{n}^+$, so that $X^\Phi(\mathbf{u}) = 2\phi$. As X and Y are in $\mathbf{n}$-quadrature, we get $Y^\Phi(\mathbf{u}) = 2\phi - \pi/2$. If $I^\Phi(\mathbf{u}, \mathbf{p}) = \phi$ for all $\mathbf{u} \in \mathcal{E}_\mathbf{n}^+$, then this means that $\bigl(\tau_{-\mathbf{p}}(I)\bigr)^\Phi = \phi$ for all $\mathbf{u} \in \mathcal{E}_\mathbf{n}^+$, thus $\bigl(\tau_{-\mathbf{p}}(X)\bigr)^\Phi = 2\phi$ and $\bigl(\tau_{-\mathbf{p}}(Y)\bigr)^\Phi = 2\phi - \pi/2$, that is $X^\Phi(\mathbf{u}, \mathbf{p}) = 2\phi$ and $Y^\Phi(\mathbf{u}, \mathbf{p}) = 2\phi - \pi/2$, and (i) holds.

By (3.11) and (3.14) we have

$$J(\mathbf{p}) = 2 \int_{\mathcal{E}_\mathbf{n}^+} I^\mathcal{A}(\mathbf{u}) A(\mathbf{u}) \exp\bigl[i\,I^\Phi(\mathbf{u}, \mathbf{p})\bigr] \, d\mathbf{u},$$

from which we derive that

$$\begin{aligned} X + i\,Y = J^2 = 4 &\iint_{\mathcal{E}_\mathbf{n}^+ \times \mathcal{E}_\mathbf{n}^+} I^\mathcal{A}(\mathbf{u}) A(\mathbf{u}) I^\mathcal{A}(\mathbf{v}) A(\mathbf{v}) \, \exp\bigl[i\,I^\Phi(\mathbf{u}, \mathbf{p})\bigr] \exp\bigl[i\,I^\Phi(\mathbf{v}, \mathbf{p})\bigr] \, d\mathbf{u} d\mathbf{v} \\ =4 &\iint_{\mathcal{E}_\mathbf{n}^+ \times \mathcal{E}_\mathbf{n}^+} I^\mathcal{A}(\mathbf{u}) A(\mathbf{u}) I^\mathcal{A}(\mathbf{v}) A(\mathbf{v}) \Bigl(\cos\bigl[I^\Phi(\mathbf{u}, \mathbf{p}) + I^\Phi(\mathbf{v}, \mathbf{p})\bigr] + i\,\sin\bigl[I^\Phi(\mathbf{u}, \mathbf{p}) + I^\Phi(\mathbf{v}, \mathbf{p})\bigr]\Bigr) \, d\mathbf{u} d\mathbf{v}. \end{aligned}$$

Taking the real and imaginary part of both sides, (4.24) results. ∎

Note that changing the sign of I does not modify X and Y; thus the latter two functions are invariant to a shift of all phases by π.

It follows from (i) that if for a given $\mathbf{p} \in \mathcal{E}$ we have $I^\Phi(\mathbf{u}, \mathbf{p}) = \phi$ for all $\mathbf{u} \in \mathcal{E}_\mathbf{n}^+$, then:

— If $\phi = 0$ or $\phi = \pi$, then $\widehat{\tau}_{-\mathbf{p}}(X) \geq 0$.

— If $\phi = \pi/4$ or $\phi = -3\pi/4$, then $\widehat{\tau}_{-\mathbf{p}}(Y) \geq 0$.

— If $\phi = \pm\pi/2$, then $\widehat{\tau}_{-\mathbf{p}}(X) \leq 0$.

— If $\phi = 3\pi/4$ or $\phi = -\pi/4$, then $\widehat{\tau}_{-\mathbf{p}}(Y) \leq 0$.

Alternately, using (ii), we see that $X(\mathbf{p})$ and $Y(\mathbf{p})$ measure the extent to which all phases $I^{\Phi}(\mathbf{u}, \mathbf{p})$ are clustered around certain multiples of $\pi/4$.

(i) As all $I^{\Phi}(\mathbf{u}, \mathbf{p})$ get close to 0 (resp. π), all $I^{\Phi}(\mathbf{u}, \mathbf{p}) + I^{\Phi}(\mathbf{v}, \mathbf{p})$ will approach 0, and so the their cosines and sines will tend to 1 and 0 respectively; thus $X(\mathbf{p})$ will increase towards $E(\mathbf{p})$ and $Y(\mathbf{p})$ will approach 0.

(ii) As all $I^{\Phi}(\mathbf{u}, \mathbf{p})$ get close to $\pi/4$ (resp. $-3\pi/4$), all $I^{\Phi}(\mathbf{u}, \mathbf{p}) + I^{\Phi}(\mathbf{v}, \mathbf{p})$ will approach $\pi/2$, and so the their cosines and sines will tend to 0 and 1 respectively; thus $X(\mathbf{p})$ will approach 0 and $Y(\mathbf{p})$ will increase towards $E(\mathbf{p})$.

(iii) As all $I^{\Phi}(\mathbf{u}, \mathbf{p})$ get close to $\pi/2$ (resp. $-\pi/2$), all $I^{\Phi}(\mathbf{u}, \mathbf{p}) + I^{\Phi}(\mathbf{v}, \mathbf{p})$ will approach π, and so the their cosines and sines will tend to -1 and 0 respectively; thus $X(\mathbf{p})$ will decrease towards $-E(\mathbf{p})$ and $Y(\mathbf{p})$ will approach 0.

(iv) As all $I^{\Phi}(\mathbf{u}, \mathbf{p})$ get close to $3\pi/4$ (resp. $-\pi/4$), all $I^{\Phi}(\mathbf{u}, \mathbf{p}) + I^{\Phi}(\mathbf{v}, \mathbf{p})$ will approach $-\pi/2$, and so the their cosines and sines will tend to 0 and -1 respectively; thus $X(\mathbf{p})$ will approach 0 and $Y(\mathbf{p})$ will decrease towards $-E(\mathbf{p})$.

More generally, we can link $X(\mathbf{p})$ and $Y(\mathbf{p})$ with $\varphi(\mathbf{p})$, the average phase at $\mathbf{p}$. Combining (4.11) and (4.22), we see that

$$X(\mathbf{p}) + i\,Y(\mathbf{p}) = J^2(\mathbf{p}) = E(\mathbf{p})e^{2i\varphi(\mathbf{p})},$$

in other words

$$X(\mathbf{p}) = E(\mathbf{p})\cos[2\varphi(\mathbf{p})] \qquad \text{and} \qquad Y(\mathbf{p}) = E(\mathbf{p})\sin[2\varphi(\mathbf{p})]. \tag{4.25}$$

While $E(\mathbf{p})$ measures the extent to which there is a feature at a point $\mathbf{p}$, the additional information provided by $X(\mathbf{p})$ and $Y(\mathbf{p})$ allows us to give the average phase modulo π at that point, and so to describe the *type* of feature encountered, where each type includes both the positive and negative feature. For example a line edge at $\mathbf{p}$ (either dark or light) has $X(\mathbf{p})$ close to $E(\mathbf{p})$ and $Y(\mathbf{p})$ close to 0, while a step edge at $\mathbf{p}$ (either dark to light or light to dark) has $X(\mathbf{p})$ close to $-E(\mathbf{p})$ and $Y(\mathbf{p})$ close to 0. On the other hand a compound line plus step edge as in Figure 2 (a), having phases close to $-\pi/4$, will give $X(\mathbf{p})$ close to 0 and $Y(\mathbf{p})$ close to $-E(\mathbf{p})$; a left-right symmetry of this profile would have phases close to $\pi/4$, and so $Y(\mathbf{p})$ close to $E(\mathbf{p})$. As we consider lines and steps as basic edges, we will give more importance to the function X than to its counterpart Y.

Another classification of the type of a feature is given in [14]; it relies on an examination of maxima, minima, and zero-crossings of Γ and Σ rather than X and Y; in other words (cfr. (4.12) and (4.25)), it is based on $\varphi(\mathbf{p})$ rather than on $2\varphi(\mathbf{p})$.

An interesting fact is that any quadratic combination of $\Gamma = I * C$ and $\Sigma = I * S$, being of the form $a\Gamma^2 + b\Sigma^2 + c\Gamma\Sigma$, will be a linear combination of $E = \Gamma^2 + \Sigma^2$, $X = \Gamma^2 - \Sigma^2$, and $Y = 2\Gamma\Sigma$. Thus every quadratic combination of Γ and Σ can be interpreted in terms of Fourier phases.

As an illustration, we consider Γ^2 and Σ^2, which are associated to traditional approaches to edge detection. Usually one convolves the image I with a single filter G, which can be a derivative of a Gaussian, a Gabor cosine or sine function, etc.; generally one of the following holds:

(a) With the aim of detecting line edges, G is even-symmetric, and has constant zero Fourier phase; thus one can consider that $G = C$.

(b) With the aim of detecting step edges, G is odd-symmetric, and has constant Fourier phase $\pi/2$ (or $-\pi/2$); thus one can consider that $G = \pm S$.

Then edges are localized at local maxima of $|I * G|$, or equivalently of $(I * G)^2$. Thus we have only to consider maxima of Γ^2 (in (a)) or Σ^2 (in (b)). Since $\Gamma^2 = (E + X)/2$ and $\Sigma^2 = (E - X)/2$, (4.10) and (4.24) give:

$$
\begin{aligned}
&\Gamma^2(\mathbf{p}) = \\
&2 \iint_{\mathcal{E}_\mathbf{n}^+ \times \mathcal{E}_\mathbf{n}^+} I^\mathcal{A}(\mathbf{u}) A(\mathbf{u}) I^\mathcal{A}(\mathbf{v}) A(\mathbf{v}) \left(\cos\bigl[I^\Phi(\mathbf{u},\mathbf{p}) - I^\Phi(\mathbf{v},\mathbf{p})\bigr] + \cos\bigl[I^\Phi(\mathbf{u},\mathbf{p}) + I^\Phi(\mathbf{v},\mathbf{p})\bigr] \right) d\mathbf{u} d\mathbf{v}; \\
&\Sigma^2(\mathbf{p}) = \\
&2 \iint_{\mathcal{E}_\mathbf{n}^+ \times \mathcal{E}_\mathbf{n}^+} I^\mathcal{A}(\mathbf{u}) A(\mathbf{u}) I^\mathcal{A}(\mathbf{v}) A(\mathbf{v}) \left(\cos\bigl[I^\Phi(\mathbf{u},\mathbf{p}) - I^\Phi(\mathbf{v},\mathbf{p})\bigr] - \cos\bigl[I^\Phi(\mathbf{u},\mathbf{p}) + I^\Phi(\mathbf{v},\mathbf{p})\bigr] \right) d\mathbf{u} d\mathbf{v}.
\end{aligned}
\tag{4.26}
$$

Thus Γ^2 and Σ^2 measure mixed aspects of phase congruence. Both phase congruence in general and congruence of phases around 0 (or π) contribute to maxima of Γ^2; thus congruence of phases around an angle $\theta \in [0, \pi/4]$ will give a relatively high value fo Γ^2. Similarly, both phase congruence in general and congruence of phases around $\pi/2$ (or $-\pi/2$) contribute to maxima of Σ^2; thus congruence of phases around an angle $\theta \in [\pi/4, \pi/2]$ will give a relatively high value fo Σ^2. In other words the "phase tuning" of Γ^2 and Σ^2 is broader than that of X. We have also the following analogue of item (ii) of Proposition 4.5:

- (a) If for a given $\mathbf{p} \in \mathcal{E}$ we have $I^\Phi(\mathbf{u}, \mathbf{p}) = 0$ for all $\mathbf{u} \in \mathcal{E}_\mathbf{n}^+$, then $\Gamma^\Phi(\mathbf{u}, \mathbf{p}) = (\Gamma^2)^\Phi(\mathbf{u}, \mathbf{p}) = 0$ for all $\mathbf{u} \in \mathcal{E}_\mathbf{n}^+$, and $\Gamma^2(\mathbf{p}) > \Gamma^2(\mathbf{x})$ for all $\mathbf{x} \in \mathcal{E}$ such that $\mathbf{x} \neq \mathbf{p}$.
- (b) If for a given $\mathbf{p} \in \mathcal{E}$ we have $I^\Phi(\mathbf{u}, \mathbf{p}) = \pi/2$ for all $\mathbf{u} \in \mathcal{E}_\mathbf{n}^+$, then $\Sigma^\Phi(\mathbf{u}, \mathbf{p}) = (\Sigma^2)^\Phi(\mathbf{u}, \mathbf{p}) = 0$ for all $\mathbf{u} \in \mathcal{E}_\mathbf{n}^+$, and $\Sigma^2(\mathbf{p}) > \Sigma^2(\mathbf{x})$ for all $\mathbf{x} \in \mathcal{E}$ such that $\mathbf{x} \neq \mathbf{p}$.

Thus perfect edges for the single filter line detector using Γ^2 are given by constant zero (or π) phase signals, perfect edges for the single filter step detector using Σ^2 are given by constant $\pi/2$ (or $-\pi/2$) phase signals, while perfect edges in the phase congruence model are given by all constant phase signals, whatever the value of that constant phase. In other words the phase congruence model generalizes previous approaches.

We can also consider a variant of the model where the phase of C in $\mathcal{E}_\mathbf{n}^+$ would be θ instead of 0, with (C, S) still an $\mathbf{n}$-quadrature pair. We define thus C_θ and S_θ by ${C_\theta}^\mathcal{A} = {S_\theta}^\mathcal{A} = A$ (cfr. (4.1)), while we have ${C_\theta}^\Phi(\mathbf{u}) = \theta$ and ${S_\theta}^\Phi(\mathbf{u}) = \theta - \pi/2$ for $\mathbf{u} \in \mathcal{E}_\mathbf{n}^+$. In fact we have

$$
\begin{aligned}
C_\theta &= \cos\theta C - \sin\theta S, \\
S_\theta &= \sin\theta C + \cos\theta S,
\end{aligned}
$$

in other words

$$C_\theta + i\, S_\theta = e^{i\theta} F.$$

Defining $\Gamma_\theta = I * C_\theta$ and $\Sigma_\theta = I * S_\theta$, then we get $e^{i\theta} J$ instead of J, the same energy E (since $\Gamma_\theta^2 + \Sigma_\theta^2 = |J|^2 = \Gamma^2 + \Sigma^2$), and

$$
\begin{aligned}
X_\theta &= \cos 2\theta X - \sin 2\theta Y, \\
Y_\theta &= \sin 2\theta X + \cos 2\theta Y,
\end{aligned}
$$

in other words

$$X_\theta + i\, Y_\theta = e^{2i\theta} J^2.$$

Note that such C_θ and S_θ give the same result on an image I as would C and S on an image I_θ having the same amplitude spectrum as I, but satisfying ${I_\theta}^\Phi(\mathbf{u}) = I^\Phi(\mathbf{u}) - \theta$ for $\mathbf{u} \in \mathcal{E}_\mathbf{n}^+$. All the above formulas apply if we replace in them $I^\Phi(\mathbf{u})$ by $I^\Phi(\mathbf{u}) - \theta$ and $I^\Phi(\mathbf{u},\mathbf{p})$ by $I^\Phi(\mathbf{u},\mathbf{p}) - \theta$. This generalization of C and S allows us to see C and S, and similarly X and Y, as two particular case of a single class of functions:

$$C = C_0 = S_{\pi/2}, \quad S = S_0 = C_{-\pi/2}, \quad S_\theta = C_{\theta-\pi/2};$$
$$\Gamma = \Gamma_0 = \Sigma_{\pi/2}, \quad \Sigma = \Sigma_0 = \Gamma_{-\pi/2}, \quad \Sigma_\theta = \Gamma_{\theta-\pi/2};$$
$$X = X_0 = Y_{\pi/4}, \quad Y = Y_0 = X_{-\pi/4}, \quad Y_\theta = X_{\theta-\pi/4}.$$

Some authors [13,16,17,18,19,29,30,31,33] have considered quadratic edge detectors using two filters having constant Fourier phases 0 and $-\pi/2$ respectively, but having different amplitude spectra, for example: the Gabor cosine and sine functions, the first and second derivatives of a Gaussian, etc. What kind of edges do they detect? We have the following result:

PROPOSITION 4.10. *Let G and Z be two real-valued continuous integrable functions such that $G^\Phi(\mathbf{u}) = 0$ and $Z^\Phi(\mathbf{u}) = -\pi/2$ for all $\mathbf{u} \in \mathcal{E}_\mathbf{n}^+$. Let I be in $L^1 + L^2$, and set $Q = (I * G)^2 + (I * Z)^2$. Suppose that there is some point $\mathbf{p} \in \mathcal{E}$ and some angle θ such that $I^\Phi(\mathbf{u},\mathbf{p}) = \theta$ for all $\mathbf{u} \in \mathcal{E}_\mathbf{n}^+$. Then $Q^\Phi(\mathbf{u},\mathbf{p}) = 0$ for all $\mathbf{u} \in \mathcal{E}$, and $Q(\mathbf{p}) > Q(\mathbf{x})$ for all $\mathbf{x} \in \mathcal{E}$ such that $\mathbf{x} \neq \mathbf{p}$, provided that one of the following holds:*

(*i*) *$G^\mathcal{A} \geq Z^\mathcal{A}$ and $\theta = 0$ or π.*

(*ii*) *$G^\mathcal{A} \leq Z^\mathcal{A}$ and $\theta = \pm\pi/2$.*

PROOF. As in Proposition 4.2, $I * G$ and $I * Z$ are uniformly continuous and square-integrable, with $(I * G)^\wedge = \widehat{I}\widehat{G}$ and $(I * Z)^\wedge = \widehat{I}\widehat{Z}$. We have $(\tau_{-\mathbf{p}}(I))^\Phi(\mathbf{u}) = \theta$ for all $\mathbf{u} \in \mathcal{E}_\mathbf{n}^+$, so that $(\tau_{-\mathbf{p}}(I * G))^\Phi(\mathbf{u}) = \theta$ and $(\tau_{-\mathbf{p}}(I * Z))^\Phi(\mathbf{u}) = \theta - \pi/2$. If (i) holds then $\tau_{-\mathbf{p}}(I * G)$ has constant phase 0 or π on $\mathcal{E}$, $\tau_{-\mathbf{p}}(I * Z)$ has constant phase $\pm\pi/2$ on $\mathcal{E}_\mathbf{n}^+$, and so it is odd-symmetric; as $(I * G)^\mathcal{A} \geq (I * Z)^\mathcal{A}$, by Proposition 3.14, $\tau_{-\mathbf{p}}(Q) = \tau_{-\mathbf{p}}(I * G)^2 + \tau_{-\mathbf{p}}(I * Z)^2$ has a real positive Fourier transform. If (ii) holds then $\tau_{-\mathbf{p}}(I * Z)$ has constant phase 0 or π on $\mathcal{E}$, $\tau_{-\mathbf{p}}(I * G)$ has constant phase $\pm\pi/2$ on $\mathcal{E}_\mathbf{n}^+$, and so it is odd-symmetric; as $(I * Z)^\mathcal{A} \geq (I * G)^\mathcal{A}$, by Proposition 3.14, $\tau_{-\mathbf{p}}(Q) = \tau_{-\mathbf{p}}(I * Z)^2 + \tau_{-\mathbf{p}}(I * G)^2$ has a real positive Fourier transform. Thus in both cases $\tau_{-\mathbf{p}}(Q)$ has constant zero Fourier phase. As Q is continuous, $\widehat{Q}$ is integrable by Proposition 3.5. The same argument as in the proof of Proposition 4.5 (in fact, an application of Corollary 3.4) shows then that $Q(\mathbf{p}) > Q(\mathbf{x})$ for $\mathbf{x} \neq \mathbf{p}$. ∎

For example, if G and Z are the Gabor cosine and sine functions, as $G^\mathcal{A} \geq Z^\mathcal{A}$, the edge detector will detect perfect line edges, which have a constant phase 0. For signals having constant phase $\pm\pi/2$, the detection of the corresponding edge is not guaranteed.

Another way of using two filters is to build two edge detectors, one for line edges using the even-symmetric filter, and one for step edges using the odd-symmetric filter, and to apply both detectors in parallel to the image. We show below that for a compound edge formed by the superposition of two signals with constant phases 0 and $\pi/2$ respectively, this leads to edge duplication: the two detectors localize edges on both sides of the true edge location:

PROPOSITION 4.11. *Let $I(x_t, x_n) = P(x_n)$, where P is neither even-symmetric nor odd-symmetric, and $P^\Phi(\nu) \in [0, \pi/2]$ for all $\nu > 0$. Consider two integrable functions G and Z, with $g = G_{/\mathbf{n}}$ and $z = Z_{/\mathbf{n}}$, such that*

(*i*) *$P * g$ and $P * z$ are continuous;*

(ii) *g is even-symmetric and has constant zero Fourier phase;*

(iii) *z is odd-symmetric and has constant Fourier phase* $-\pi/2$*;*

(iv) $\widehat{P}\widehat{g}$, $\widehat{P}\widehat{z}$, $\xi\widehat{P}\widehat{g}$, *and* $\xi\widehat{P}\widehat{z}$ *are integrable, in other words*

$$\int_{\mathbb{R}}(1+|x|)|\widehat{P}(x)|(|\widehat{g}(x)|+|\widehat{z}(x)|)\,dx<\infty.$$

Then $I*G$ *and* $I*Z$ *are derivable in* x_n, *and*

$$\frac{\partial[(I*G)^2]}{\partial x_n}(x_t,0)\cdot\frac{\partial[(I*Z)^2]}{\partial x_n}(x_t,0)<0.$$

PROOF. Clearly $(I*G)(x_t,x_n)=(P*g)(x_n)$ and $(I*Z)(x_t,x_n)=(P*z)(x_n)$ (see (4.8) and (4.9)). Since $P*g$ and $P*z$ are continuous and their Fourier transforms $\widehat{P}\widehat{g}$ and $\widehat{P}\widehat{z}$ are integrable, we can apply Lemma 3.2, and so $P*g=(\widehat{P}\widehat{g})^{\vee}$ and $P*z=(\widehat{P}\widehat{z})^{\vee}$.

Let $P_e=(P+P_\rho)/2$ and $P_o=(P-P_\rho)/2$ be the even-symmetric and odd-symmetric parts of P, in other words $P=P_e+P_o$, where P_e is even-symmetric and P_o is odd-symmetric. Since P is neither even-symmetric nor odd-symmetric, P_e and P_o are non-zero. As $P^{\Phi}(\nu)\in[0,\pi/2]$ for all $\nu>0$, we deduce that $P_e^{\Phi}(\nu)=0$ and $P_o^{\Phi}(\nu)=\pi/2$ for all $\nu>0$. Since the Fourier transform is linear and commutes with the reflection ρ, we have $P_e*g=(\widehat{P}_e\widehat{g})^{\vee}$, $P_o*g=(\widehat{P}_o\widehat{g})^{\vee}$, $P_e*z=(\widehat{P}_e\widehat{z})^{\vee}$ and $P_o*z=(\widehat{P}_o\widehat{z})^{\vee}$. Moreover, as $\xi\widehat{P}\widehat{g}$ and $\xi\widehat{P}\widehat{z}$ are integrable, the same will be true for $\xi\widehat{P}_e\widehat{g}$, $\xi\widehat{P}_o\widehat{g}$, $\xi\widehat{P}_e\widehat{z}$, and $\xi\widehat{P}_o\widehat{z}$; by the L^1 Fourier derivative formula, P_e*g, P_o*g, P_e*z, and P_o*z will be derivable, with $(P_e*g)'=2\pi i\,(\xi\widehat{P}_e\widehat{g})^{\vee}$, $(P_o*g)'=2\pi i\,(\xi\widehat{P}_o\widehat{g})^{\vee}$, $(P_e*z)'=2\pi i\,(\xi\widehat{P}_e\widehat{z})^{\vee}$ and $(P_o*z)'=2\pi i\,(\xi\widehat{P}_o\widehat{z})^{\vee}$.

As P_e and g are even-symmetric while P_o and z are odd-symmetric, we deduce that P_o*g, P_e*z, $(P_e*g)'$, and $(P_o*z)'$ are odd-symmetric; in particular

$$(P_o*g)(0)=(P_e*z)(0)=(P_e*g)'(0)=(P_o*z)'(0)=0.$$

We obtain then:

$$\begin{aligned}
&[(P*g)^2]'(0)=2(P*g)(0)(P*g)'(0)\\
&=2\big[(P_o*g)(0)+(P_e*g)(0)\big]\big[(P_o*g)'(0)+(P_e*g)'(0)\big]=2(P_e*g)(0)(P_o*g)'(0)\\
\text{and}\quad&[(P*z)^2]'(0)=2(P*z)(0)(P*z)'(0)\\
&=2\big[(P_o*z)(0)+(P_e*z)(0)\big]\big[(P_o*z)'(0)+(P_e*z)'(0)\big]=2(P_o*z)(0)(P_e*z)'(0).
\end{aligned}\tag{4.27}$$

Now (3.11) gives:

$$\begin{aligned}
(P_e*g)(0)&=\int_{\mathcal{E}}\widehat{P}_e(\nu)\widehat{g}(\nu)\,d\nu,\\
(P_o*g)'(0)&=\int_{\mathcal{E}}2\pi i\,\nu\widehat{P}_o(\nu)\widehat{g}(\nu)\,d\nu,\\
(P_o*z)(0)&=\int_{\mathcal{E}}\widehat{P}_o(\nu)\widehat{z}(\nu)\,d\nu,\\
\text{and}\qquad(P_e*z)'(0)&=\int_{\mathcal{E}}2\pi i\,\nu\widehat{P}_e(\nu)\widehat{z}(\nu)\,d\nu.
\end{aligned}\tag{4.28}$$

Now since $P_e^{\Phi}(\nu)=g^{\Phi}(\nu)=0$, $P_o^{\Phi}(\nu)=\pi/2$, and $z^{\Phi}(\nu)=-\pi/2$ for all $\nu>0$, we have thus for all $\nu\neq0$: $\mathrm{sgn}(\widehat{P}_e(\nu))=\mathrm{sgn}(\widehat{g}(\nu))=1$, $\mathrm{sgn}(\widehat{P}_o(\nu))=i\,\mathrm{sgn}(\nu)$, and $\mathrm{sgn}(\widehat{z}(\nu))=-i\,\mathrm{sgn}(\nu)$, so that:

$$\begin{aligned}
\mathrm{sgn}\big(\widehat{P}_e(\nu)\widehat{g}(\nu)\big)&=1\cdot1=1,\\
\mathrm{sgn}\big(2\pi i\,\nu\widehat{P}_o(\nu)\widehat{g}(\nu)\big)&=i\cdot\mathrm{sgn}(\nu)\cdot i\,\mathrm{sgn}(\nu)\cdot1=-1,\\
\mathrm{sgn}\big(\widehat{P}_o(\nu)\widehat{z}(\nu)\big)&=i\,\mathrm{sgn}(\nu)\cdot[-i\,\mathrm{sgn}(\nu)]=1,\\
\text{and}\qquad\mathrm{sgn}\big(2\pi i\,\nu\widehat{P}_e(\nu)\widehat{z}(\nu)\big)&=i\cdot\mathrm{sgn}(\nu)\cdot1\cdot[-i\,\mathrm{sgn}(\nu)]=1.
\end{aligned}$$

Combining this with (4.28) gives:

$$\operatorname{sgn}\big((P_e * g)(0)\big) = 1, \quad \operatorname{sgn}\big((P_o * g)'(0)\big) = -1, \quad \operatorname{sgn}\big((P_o * z)(0)\big) = 1, \quad \text{and} \quad \operatorname{sgn}\big((P_e * z)'(0)\big) = 1.$$

Hence (4.27) gives

$$\begin{aligned}
&\operatorname{sgn}\big([(P * g)^2]'(0) \cdot [(P * z)^2]'(0)\big) \\
=&\operatorname{sgn}\big((P_e * g)(0)\big) \cdot \operatorname{sgn}\big((P_o * g)'(0)\big) \cdot \operatorname{sgn}\big((P_o * z)(0)\big) \cdot \operatorname{sgn}\big((P_e * z)'(0)\big) \\
=&1 \cdot (-1) \cdot 1 \cdot 1 = -1.
\end{aligned}$$

Since $(I*G)(x_t, x_n) = (P*g)(x_n)$ and $(I*Z)(x_t, x_n) = (P*z)(x_n)$, we have $\partial[(I*G)^2]/\partial x_n(x_t, 0) = [(P*g)^2]'(0)$ and $\partial[(I*Z)^2]/\partial x_n(x_t, 0) = [(P*z)^2]'(0)$. Therefore the result follows. ∎

Thus at the line $x_n = 0$ given by the profile P, one of $|I*G|$ and $|I*Z|$ is strictly increasing in x_n, while the other is strictly decreasing in x_n. Therefore the maxima of $|I*G|$ and $|I*Z|$ in the normal direction $\mathbf{n}$ lie on both sides of the true line $x_n = 0$.

4.4. Orientation selectivity

Many previous studies of quadratic models for edge detection, in particular of the phase congruence model [8,9,13,14,15,16,17,18,22] assumed one-dimensional signals and filters. Here we will consider two-dimensional signals and the problems associated with the choice of the filter orientation. Up to now, we have supposed that the filters C and S have a fixed orientation $\mathbf{n}$, and all one-dimensional edge profiles were chosen to have their normal orientation parallel to it. We will now examine what happens when the orientation of the filters is allowed to vary.

We take thus $d = 2$; for every real number θ modulo 2π, let R_θ be the rotation of angle θ on $\mathcal{E}$:

$$(x_t, x_n) \mapsto (x_n \sin\theta + x_t \cos\theta, x_n \cos\theta - x_t \sin\theta), \tag{4.29}$$

and write $\mathbf{t}_\theta$ and $\mathbf{n}_\theta$ for the unit vectors resulting from the rotation R_θ applied to $\mathbf{t}$ and $\mathbf{n}$ respectively, in other words

$$\begin{aligned}
\mathbf{t}_\theta &= R_\theta(\mathbf{t}) = \cos\theta\, \mathbf{t} - \sin\theta\, \mathbf{n}, \\
\mathbf{n}_\theta &= R_\theta(\mathbf{n}) = \sin\theta\, \mathbf{t} + \cos\theta\, \mathbf{n}.
\end{aligned} \tag{4.30}$$

We can also apply R_θ to filters and signals, and so for every function $G : \mathcal{E} \to \mathbb{C}$ we define $G_\theta = R_\theta(G)$ by

$$G_\theta\big(R_\theta(\mathbf{x})\big) = G(\mathbf{x}),$$

in other words

$$G_\theta(\mathbf{x}) = G\big(R_\theta^{-1}(\mathbf{x})\big) = G\big(R_{-\theta}(\mathbf{x})\big), \tag{4.31}$$

that is:

$$G_\theta(x_t, x_n) = G(-x_n \sin\theta + x_t \cos\theta, x_n \cos\theta + x_t \sin\theta). \tag{4.32}$$

From (4.5) and (4.31) we obtain for all $x \in \mathbb{R}$:

$$[G_\theta]_{/\mathbf{n}}(x) = \int_R [G_\theta](x\cdot\mathbf{n} + y\cdot\mathbf{t})\, dy = \int_R G\big(R_{-\theta}(x\cdot\mathbf{n} + y\cdot\mathbf{t})\big)\, dy = \int_R G(x\cdot\mathbf{n}_{-\theta} + y\cdot\mathbf{t}_{-\theta})\, dy = G_{/\mathbf{n}_{-\theta}}(x),$$

so that

$$[G_\theta]_{/\mathbf{n}} = G_{/\mathbf{n}_{-\theta}}. \tag{4.33}$$

We know that the rotation R_θ commutes with all algebraic operations on functions, that it distributes the convolution, namely

$$(G * H)_\theta = G_\theta * H_\theta,$$

and that it commutes with the Fourier transform, in other words

$$(G_\theta)^\wedge = (\widehat{G})_\theta;$$

we write thus $\widehat{G}_\theta$ for the Fourier transform of G rotated by θ. Note that $(\mathrm{sgn}_{\mathbf{n}})_\theta = \mathrm{sgn}_{\mathbf{n}_\theta}$, so that given C and S in $\mathbf{n}$-quadrature, C_θ and S_θ will be in $\mathbf{n}_\theta$-quadrature.

We will now consider the behaviour of the rotated filters C_θ, S_θ, and $F_\theta = C_\theta + i\, S_\theta$ on an image I forming a one-dimensional profile P in the normal direction $\mathbf{n}$, that is $I(x_t, x_n) = P(x_n)$ (cfr. (4.8)); by (4.9) and (4.33) we have

$$(I * F_\theta)(x_t, x_n) = (P * [F_\theta]_{/\mathbf{n}})(x_n) = (P * F_{/\mathbf{n}_{-\theta}})(x_n). \tag{4.34}$$

Now (4.6) together with (4.30, 4.31, 4.32) gives for every $\nu \in \mathbb{R}$:

$$\big([F_\theta]_{/\mathbf{n}}\big)^\wedge(\nu) = \big(F_{/\mathbf{n}_{-\theta}}\big)^\wedge(\nu) = \widehat{F}_\theta(\nu \cdot \mathbf{n}) = \widehat{F}(\nu \cdot \mathbf{n}_{-\theta}) = \widehat{F}(-\nu \sin\theta, \nu \cos\theta). \tag{4.35}$$

The same can be written with C or S instead of F. We obtain thus the following result:

PROPOSITION 4.12. *Let $d = 2$ and I be given by $I(x_t, x_n) = P(x_n)$ for a one-dimensional profile P. Then for $|\theta| < \pi/2$, the phase spectrum of the one-dimensional profile of $I * F_\theta$ (resp., $I * C_\theta$, $I * S_\theta$) does not depend on θ. In particular if P is in $L^1 + L^2$ and has local phase $P^\Phi(\nu, p)$ constant at point p for all $\nu > 0$, then the energy function will have an absolute maximum at the line $x_n = p$.*

PROOF. Indeed $\mathbf{n}_{-\theta} \in \mathcal{E}^+_{\mathbf{n}}$ for $|\theta| < \pi/2$, so that by Proposition 4.3, $C_{/\mathbf{n}_{-\theta}}$, $S_{/\mathbf{n}_{-\theta}}$, and $F_{/\mathbf{n}_{-\theta}}$ satisfy Requirements 1, 2', and 3 for $d = 1$. From Proposition 4.2 and (4.35) we deduce that the phases of $P * C_{/\mathbf{n}_{-\theta}}$, $P * S_{/\mathbf{n}_{-\theta}}$, and $P * F_{/\mathbf{n}_{-\theta}}$ do not depend on θ. By Proposition 4.5, the constancy of the local phase $P^\Phi(\nu, p)$ at p for $\nu > 0$ leads to an absolute maximum of $|P * F_{/\mathbf{n}_{-\theta}}|^2$ at p, in other words by (4.34), the energy function $|I * F_\theta|^2$ has an absolute maximum at the line $x_n = p$. ∎

For $|\theta| > \pi/2$, we have $\theta = \eta \pm \pi$ with $|\eta| < \pi/2$, and the fact that C is even-symmetric and S is odd-symmetric implies that

$$C_{\eta\pm\pi} = C_\eta \qquad \text{and} \qquad S_{\eta\pm\pi} = -S_\eta.$$

Thus, since the energy function takes the sum of squares of convolutions by C_θ and S_θ, Proposition 4.12 remains essentially true for $|\theta| > \pi/2$. Note that for $\theta = \pm\pi/2$ we have $F_{/\mathbf{n}_{-\theta}} = F_{/\pm\mathbf{t}}$, which is identically zero by (4.7).

The concrete meaning of this result is that a one-dimensional feature can be correctly localized even when the normal orientation of the filters does not match that of the feature, provided that they are not perpendicular. In practice, as θ, the angle between the normal orientations of the feature and of the filters, tends to $\pm\pi/2$, the amplitude of $F_{/\mathbf{n}_{-\theta}}$ will diminish, and quantization errors will prevent the localization of maxima of the energy function.

After the localization of the edge, the next problem is the determination of its orientation. Traditional approaches based on the convolution of the image with a single mask rotated into several orientations, select at every point the orientation for which the absolute value of the convolution is the highest. The early rationale behind this procedure was that the grey-level profile of the mask was chosen to represent a local template of an ideal (step or line) edge, and edge detection could thus be achieved as a form of template matching: the higher the correlation with a template with a certain orientation, the higher the likelihood of having there such an edge template with that orientation.

Since we are using convolution kernels specified by analytic properties, in particular in the Fourier domain, and do not consider them as templates for an edge profile, the approach derived from template matching is not guaranteed to work properly. We will show it with three simple examples involving some peculiar filters applied to a two-dimensional Heaviside step edge given by

$$I(x_t, x_n) = P(x_n) = \begin{cases} 0 & \text{if } x_n < 0, \\ 1 & \text{if } x_n > 0, \end{cases} \tag{4.36}$$

as illustrated in Figure 8.

Let us consider first a filter F whose support is restricted to the normal direction, having no extent in the tangential direction. The filter can be considered as a generalized function $(x_t, x_n) \mapsto \delta(x_t) \cdot f(x_n)$, where δ is the Dirac impulse and f is an integrable function $\mathbb{R} \to \mathbb{R}$. Take an angle θ such that $|\theta| < \pi/2$; the rotated filter F_θ takes then the form $F_\theta(x\mathbf{t}_\theta + y\mathbf{n}_\theta) = \delta(x) \cdot f(y)$, so that the convolution $I * F_\theta$ gives at every point $\mathbf{p} = (p_t, p_n)$:

$$(I * F_\theta)(\mathbf{p}) = \int_{\mathbb{R}^2} I(\mathbf{p} - x\mathbf{t}_\theta - y\mathbf{n}_\theta) F_\theta(x\mathbf{t}_\theta + y\mathbf{n}_\theta)\, dxdy = \int_{\mathbb{R}^2} I(\mathbf{p} - x\mathbf{t}_\theta - y\mathbf{n}_\theta)\delta(x) f(y)\, dxdy$$
$$= \int_{\mathbb{R}} I(\mathbf{p} - y\mathbf{n}_\theta) f(y)\, dy = \int_{\mathbb{R}} I(p_t - y\sin\theta, p_n - y\cos\theta) f(y)\, dy = \int_{\mathbb{R}} P(p_n - y\cos\theta) f(y)\, dy.$$

Since the profile P is given by the Heaviside step function (4.36), we have $P(p_n - y\cos\theta) = 1$ for $y < p_n/\cos\theta$, and 0 for $y > p_n/\cos\theta$, so that we get

$$(I * F_\theta)(p_t, p_n) = \int_{-\infty}^{p_n/\cos\theta} f(y)\, dy. \tag{4.37}$$

Geometrically speaking, this convolution by F_θ is constructed as follows: at every point $\mathbf{p} = (p_t, p_n)$, draw a line making an angle θ with the normal direction; we integrate on this line the reflected function f_ρ multipled by the edge profile, which means in fact that we integrate f_ρ on the portion of this line lying on the right side of the Heaviside step edge. This is illustrated in Figure 8.

There is no reason for having a maximum of the absolute value of (4.37) for $\theta = 0$. Indeed for $p_n = 0$ the value of (4.37) does not depend on θ. Furthermore, we show below how for $p_n \neq 0$ it is possible to have for the absolute value of (4.37) an absolute maximum at $\theta \neq 0$, but no local maximum at $\theta = 0$. We illustrate in Figure 8 two possible profiles for the function f, one even-symmetric and the other odd-symmetric. We make the simple assumption that the zero-crossings of f are $-a, a$ in the even-symmetric case, and $-a, 0, a$ in the odd-symmetric case; here $f(x) \geq 0$ for $-a \leq x \leq 0$ and $f(x) < 0$ for $x < -a$. Let

$$\alpha = \int_{-a}^{0} f(y)\, dy \qquad \text{and} \qquad \beta = -\int_{-\infty}^{-a} f(y)\, dy,$$

so that $\alpha, \beta > 0$. Take p_n such that $-a < p_n < 0$. As $|\theta|$ decreases from $\pi/2$ to $\arccos[p_n/(-a)]$, $p_n/\cos\theta$ increases from $-\infty$ to $-a$, so that (4.37) decreases from 0 to $-\beta$; as $|\theta|$ decreases further from $\arccos[p_n/(-a)]$ to 0, $p_n/\cos\theta$ increases from $-a$ to p_n, so that (4.37) increases to $-\beta + \int_{-a}^{p_n} f$, remaining smaller than $\alpha - \beta$. This means thus that the $\big|(I * F_\theta)(p_t, p_n)\big|$ has a maximum at $\theta = \pm\arccos[p_n/(-a)]$ and an extremum at $\theta = 0$; this extremum is a minimum if $\int_{-a}^{p_n} f < \beta$, in particular if $\alpha < \beta$, but it is a maximum otherwise; it will be an absolute maximum (i.e., greater than the maximum at $\theta = \pm\arccos[p_n/(-a)]$) if $\int_{-a}^{p_n} f > 2\beta$, which requires that $\alpha > 2\beta$.

Therefore it is possible for $\big|(I * F_\theta)(p_t, p_n)\big|$, as a function of θ to have an absolute minimum at $\theta = 0$ for some $p_n \neq 0$, while for $p_n = 0$ it will be constant. Note that (4.37) as a function of p_n and θ is discontinuous at $\theta = \pm\pi/2$; as θ increases from 0 to $\pi/2$, its evolution w.r.t. p_n becomes faster: the energy profile around $p_n = 0$ becomes steeper. We have thus two reasons for requiring the filter to have some width in the tangential direction: continuity of the convolution w.r.t. position and orientation, and orientation selectivity, in other words the possibility to determine the edge orientation as the one giving the highest result for the energy function. In practice, if the filter has a wide support in the normal direction but a narrow one in the tangential direction, orientation selectivity will not be achieved, and for $\theta \neq 0$ the edge in the filtered image will be sharper than for $\theta = 0$. In fact, it has been verified experimentally [31] that orientation selectivity increases with the ratio of tangential width over normal width, and that it is even necessary to take the width in the tangential direction equal to three times the width in the normal direction.

Let us now consider a second example with a step edge detector using a separable odd-symmetric filter defined as the product of a Gabor sine function in the normal direction and a Gabor cosine function in the tangential direction:

$$F(x_t, x_n) = G_\sigma(x_t)\cos(2\pi\alpha x_t) \cdot G_\sigma(x_n)\sin(2\pi\alpha x_n), \tag{4.38}$$

where G_σ is the Gaussian (cfr. (3.21)). We show in Figure 9 the sign and zero-crossings of F. Since the Gaussian has a Gaussian-type Fourier transform, it is easy to check that F has constant phase $-\pi/2$ on $\mathcal{E}_{\mathbf{n}}^{+}$, in other words that $\operatorname{sgn}(\widehat{F}(u_t, u_n)) = -i\operatorname{sgn}(u_n)$. Using (4.32) and the rotational symmetry of the two-dimensional Gaussian $G_\sigma(x_t)G_\sigma(x_n)$, we deduce that

$$\begin{aligned} F_{\pi/4}(x_t, x_n) &= G_\sigma(x_t)G_\sigma(x_n)\cos\big(2\pi\alpha\big[-x_n\sin\frac{\pi}{4} + x_t\cos\frac{\pi}{4}\big]\big)\sin\big(2\pi\alpha\big[x_n\cos\frac{\pi}{4} + x_t\sin\frac{\pi}{4}\big]\big) \\ &= G_\sigma(x_t)G_\sigma(x_n)\cos\big(2\pi\alpha[-x_n + x_t]/\sqrt{2}\big)\sin\big(2\pi\alpha[x_n + x_t]/\sqrt{2}\big) \\ &= \frac{1}{2}G_\sigma(x_t)G_\sigma(x_n)\big(\sin(2\pi\alpha\sqrt{2}x_t) + \sin(2\pi\alpha\sqrt{2}x_n)\big). \end{aligned} \tag{4.39}$$

Since $\widehat{G}_\sigma(t) = \exp[-2(\pi\sigma t)^2]$ (cfr. formula 7.4.6 of [61]), we obtain that

$$\begin{aligned} F_{/\mathbf{n}}(x_n) &= \int_{\mathbb{R}} F(x_t, x_n)\,dx_t = G_\sigma(x_n)\sin(2\pi\alpha x_n)\int_{\mathbb{R}} G_\sigma(x_t)\cos(2\pi\alpha x_t)\,dx_t \\ &= \widehat{G}_\sigma(\alpha)G_\sigma(x_n)\sin(2\pi\alpha x_n) = \exp[-2(\pi\sigma\alpha)^2]G_\sigma(x_n)\sin(2\pi\alpha x_n) \end{aligned} \tag{4.40}$$

and

$$\begin{aligned} [F_{\pi/4}]_{/\mathbf{n}}(x_n) &= \int_{\mathbb{R}} F_{\pi/4}(x_t, x_n)\,dx_t \\ &= \frac{1}{2}G_\sigma(x_n)\Big(\int_{\mathbb{R}} G_\sigma(x_t)\sin(2\pi\alpha\sqrt{2}x_t)\,dx_t + \sin(2\pi\alpha\sqrt{2}x_n)\int_{\mathbb{R}} G_\sigma(x_t)\,dx_t\Big) \\ &= \frac{1}{2}G_\sigma(x_n)\big(0 + \sin(2\pi\alpha\sqrt{2}x_n)\widehat{G}_\sigma(0)\big) = \frac{1}{2}G_\sigma(x_n)\sin(2\pi\alpha\sqrt{2}x_n). \end{aligned} \tag{4.41}$$

Now (4.34) and (4.36) give

$$(I * F)(x_t, x_n) = (P * F_{/\mathbf{n}})(x_n) = \int_{-\infty}^{x_n} F_{/\mathbf{n}}(t)\, dt$$

and similarly

$$(I * [F_{\pi/4}])(x_t, x_n) = (P * [F_{\pi/4}]_{/\mathbf{n}})(x_n) = \int_{-\infty}^{x_n} [F_{\pi/4}]_{/\mathbf{n}}(t)\, dt.$$

From formula 7.4.7 of [61] we have

$$\int_{-\infty}^{0} G_\sigma(t) \sin 2\pi f t\, dt = -\frac{1}{\sqrt{\pi}} \exp[-2(\sigma\pi f)^2] \int_0^{\sqrt{2}\sigma\pi f} e^{s^2}\, ds,$$

so that (4.40) and (4.41) give

$$\begin{aligned}
(I * F)(x_t, 0) &= \int_{-\infty}^{0} F_{/\mathbf{n}}(t)\, dt = \int_{-\infty}^{0} \exp[-2(\pi\sigma\alpha)^2] G_\sigma(t) \sin(2\pi\alpha t)\, dt \\
&= -\frac{1}{\sqrt{\pi}} \exp[-(2\pi\sigma\alpha)^2] \int_0^{\sqrt{2}\sigma\pi\alpha} e^{s^2}\, ds, \qquad \text{and} \\
(I * [F_{\pi/4}])(x_t, 0) &= \int_{-\infty}^{0} [F_{\pi/4}]_{/\mathbf{n}}(t)\, dt = \int_{-\infty}^{0} \frac{1}{2} G_\sigma(t) \sin(2\pi\alpha\sqrt{2}t)\, dt \\
&= -\frac{1}{2\sqrt{\pi}} \exp[-(2\pi\sigma\alpha)^2] \int_0^{2\sigma\pi\alpha} e^{s^2}\, ds.
\end{aligned} \tag{4.42}$$

From the properties of Dawson's integral (see formulas 7.1.17 and 7.1.18 and Table 7.5 of [61]), the function $\int_0^x e^{s^2}\, ds$ tends to $e^{x^2}/(2x)$ for $x \to +\infty$. Thus for $\sigma\alpha$ large enough we have

$$\frac{1}{2} \int_0^{2\sigma\pi\alpha} e^{s^2}\, ds \approx \frac{\exp[4(\sigma\pi\alpha)^2]}{8\sigma\pi\alpha} \gg \frac{\exp[2(\sigma\pi\alpha)^2]}{2\sqrt{2}\sigma\pi\alpha} \approx \int_0^{\sqrt{2}\sigma\pi\alpha} e^{s^2}\, ds,$$

so that (4.42) gives

$$\big|(I * [F_{\pi/4}])(x_t, 0)\big| \gg \big|(I * F)(x_t, 0)\big|.$$

This means in practice that the larger we take σ and α, the more will show the tendency of the filter to detect edges making a an angle of 45 degrees with the normal orientation. If we refer to Figure 9, as σ and α increase, among the square regions enclosed by the zero-crossings around the origin, the number of those in which the absolute value of F is non-negligible will increase, so that the grey-level profile of F becomes dominated by an alternation of diagonal bands made from square regions of respectively positive and negative sign.

One could raise an objection against the latter example, that the Gabor cosine function is not a smoothing function like the Gaussian, but rather a feature detector; thus the filter being the product of feature detectors in the normal and tangential directions, should in fact detect features in the intermediate diagonal directions.

We now give a third example to show that even with a standard edge detector formed as the product of of a Gaussian in the tangential direction and a Gaussian derivative in the normal direction (cfr. [7]), the detection of the edge orientation is sometimes possible only in a close neighbourhood of the edge position. Let the filter F be defined by

$$F(x_t, x_n) = x_n G_\nu(x_n) \cdot G_\tau(x_t), \qquad \text{where} \quad \tau > \nu. \tag{4.43}$$

Here G_τ and G_ν are Gaussians of respective standard deviations τ and ν (cfr. (3.21)), and we have assumed (as in [31]) that the support of the filter is wider in the tangential direction than in the normal one. Take an angle θ with $|\theta| < \pi/2$. Let us define β, γ, δ by

$$\begin{aligned} \beta^2 &= \tau^2 \sin^2\theta + \nu^2 \cos^2\theta, \qquad \text{with} \quad \beta > 0, \\ \gamma &= (\tau^2 - \nu^2)\sin\theta\cos\theta, \\ \delta^2 &= \tau^2\cos^2\theta + \nu^2\sin^2\theta. \end{aligned} \tag{4.44}$$

It is easily verified that $\delta^2\beta^2 = \gamma^2 + \tau^2\nu^2$. A straightforward computation gives then

$$\begin{aligned} &\frac{(-x_n\sin\theta + x_t\cos\theta)^2}{\tau^2} + \frac{(x_n\cos\theta + x_t\sin\theta)^2}{\nu^2} = \frac{x_t\beta^2 + 2x_nx_t\gamma + x_n^2\delta^2}{\tau^2\nu^2} \\ =&\frac{x_t\beta^2 + 2x_nx_t\gamma + x_n^2(\gamma^2+\tau^2\nu^2)\beta^{-2}}{\tau^2\nu^2} = \frac{\left(x_t + x_n\gamma\beta^{-2}\right)^2\beta^2}{\tau^2\nu^2} + \frac{x_n^2}{\beta^2}, \end{aligned}$$

so that we obtain

$$\begin{aligned} &G_\tau(-x_n\sin\theta + x_t\cos\theta)G_\nu(x_n\cos\theta + x_t\sin\theta) \\ =&\frac{1}{\tau\sqrt{2\pi}}\exp\Big[-\frac{(-x_n\sin\theta + x_t\cos\theta)^2}{2\tau^2}\Big]\cdot\frac{1}{\nu\sqrt{2\pi}}\exp\Big[-\frac{(x_n\cos\theta + x_t\sin\theta)^2}{2\nu^2}\Big] \\ =&\frac{1}{2\pi\tau\nu}\exp\Big[-\frac{1}{2}\Big(\frac{(-x_n\sin\theta + x_t\cos\theta)^2}{\tau^2} + \frac{(x_n\cos\theta + x_t\sin\theta)^2}{\nu^2}\Big)\Big] \\ =&\frac{1}{2\pi\tau\nu}\exp\Big[-\frac{1}{2}\Big(\frac{(x_t + x_n\gamma\beta^{-2})^2\beta^2}{\tau^2\nu^2} + \frac{x_n^2}{\beta^2}\Big)\Big] \\ =&\frac{\beta}{\tau\nu\sqrt{2\pi}}\exp\Big[-(x_t + x_n\gamma\beta^{-2})^2\frac{\beta^2}{2\tau^2\nu^2}\Big]\cdot\frac{1}{\beta\sqrt{2\pi}}\exp\big[-\frac{x_n^2}{2\beta^2}\big] = G_{\tau\nu/\beta}(x_t + x_n\gamma\beta^{-2})G_\beta(x_n). \end{aligned}$$

Therefore we have by (4.32):

$$\begin{aligned} F_\theta(x_t, x_n) &= (x_n\cos\theta + x_t\sin\theta)G_\tau(-x_n\sin\theta + x_t\cos\theta)G_\nu(x_n\cos\theta + x_t\sin\theta) \\ &= (x_n\cos\theta + x_t\sin\theta)G_{\tau\nu/\beta}(x_t + x_n\gamma\beta^{-2})G_\beta(x_n). \end{aligned} \tag{4.45}$$

It is easily verified from (4.44) that $\beta^2\cos\theta - \gamma\sin\theta = \nu^2\cos\theta$, so that

$$\begin{aligned} x_n\cos\theta + x_t\sin\theta &= x_n\cos\theta - x_n\gamma\beta^{-2}\sin\theta + x_t\sin\theta + x_n\gamma\beta^{-2}\sin\theta \\ &= x_n\beta^{-2}(\beta^2\cos\theta - \gamma\sin\theta) + (x_t + x_n\gamma\beta^{-2})\sin\theta \\ &= x_n\beta^{-2}\nu^2\cos\theta + \sin\theta(x_t + x_n\gamma\beta^{-2}). \end{aligned}$$

Thus (4.45) gives

$$\begin{aligned} F_\theta(x_t, x_n) = &\, x_n\beta^{-2}\nu^2\cos\theta\, G_\beta(x_n)\cdot G_{\tau\nu/\beta}(x_t + x_n\gamma\beta^{-2}) \\ &+ G_\beta(x_n)\sin\theta\cdot(x_t + x_n\gamma\beta^{-2})G_{\tau\nu/\beta}(x_t + x_n\gamma\beta^{-2}). \end{aligned}$$

Then we obtain

$$\begin{aligned} [F_\theta]_{/\mathbf{n}}(x_n) = &\int_{\mathbb{R}} F_\theta(x_t, x_n)\, dx_t \\ =&x_n\beta^{-2}\nu^2\cos\theta\, G_\beta(x_n)\int_{\mathbb{R}} G_{\tau\nu/\beta}(x_t + x_n\gamma\beta^{-2})\, dx_t \\ &+ G_\beta(x_n)\sin\theta\int_{\mathbb{R}}(x_t + x_n\gamma\beta^{-2})G_{\tau\nu/\beta}(x_t + x_n\gamma\beta^{-2})\, dx_t \\ =&x_n\beta^{-2}\nu^2\cos\theta G_\beta(x_n)\cdot 1 + G_\beta(x_n)\sin\theta\cdot 0 = \frac{\nu^2}{\beta^2}x_n\cos\theta\, G_\beta(x_n). \end{aligned} \tag{4.46}$$

From (4.34) and (4.36) we get

$$
\begin{aligned}
(I * F_\theta)(x_t, p) = (P * [F_\theta]_{/\mathbf{n}})(p) &= \int_{-\infty}^{p} [F_\theta]_{/\mathbf{n}}(x_n)\, dx_n \\
&= \frac{\nu^2}{\beta^2}\cos\theta \int_{-\infty}^{p} x_n\, G_\beta(x_n)\, dx_n = \frac{\nu^2}{\beta^2}\cos\theta\cdot(-\beta^2)G_\beta(p) \qquad (4.47)\\
&= -\frac{\nu^2\cos\theta}{\beta\sqrt{2\pi}}\exp[-p^2/2\beta^2].
\end{aligned}
$$

The energy function at point (x_t, p) for the filter F_θ is the square of this expression, which is a function of p and θ:

$$
G(p,\theta) = \frac{\nu^4\cos^2\theta}{2\pi\beta^2}\exp[-p^2/\beta^2],
$$

where $\beta = \left(\tau^2\sin^2\theta + \nu^2\cos^2\theta\right)^{1/2}$ (see (4.44)). Note that G is symmetric in both p and θ. Let $\beta' = d\beta/d\theta$; then we have

$$
\beta\beta' = \frac{1}{2}\frac{d}{d\theta}(\beta^2) = \frac{1}{2}\frac{d}{d\theta}(\tau^2\sin^2\theta + \nu^2\cos^2\theta) = (\tau^2-\nu^2)\cos^2\theta.
$$

This gives thus

$$
\begin{aligned}
\frac{\partial}{\partial\theta}G(p,\theta) &= \frac{\nu^4\cos\theta}{\pi\beta^6}\exp[-p^2/\beta^2]\big(-\beta^4\sin\theta - \beta^3\beta'\cos\theta + p^2\beta\beta'\cos\theta\big)\\
&= \frac{\nu^4\cos\theta\sin\theta}{\pi\beta^6}\exp[-p^2/\beta^2]\big(-\tau^4 + (\tau^2+p^2)(\tau^2-\nu^2)\cos^2\theta\big),
\end{aligned}
$$

whose sign (for $|\theta| < \pi/2$) is given by

$$
\sin\theta\big(-\tau^4 + (\tau^2+p^2)(\tau^2-\nu^2)\cos^2\theta\big). \qquad (4.48)
$$

This expression is antisymmetric in θ, while it is symmetric in p and increasing with $|p|$. Let us fix $|p|$; we have then two cases:

(*a*) $|p| \leq \tau\nu/\sqrt{\tau^2-\nu^2}$.
Then $(\tau^2+p^2)(\tau^2-\nu^2) \leq \tau^4$, and so for every $\theta \neq 0$ we have

$$
-\tau^4 + (\tau^2+p^2)(\tau^2-\nu^2)\cos^2\theta < -\tau^4 + (\tau^2+p^2)(\tau^2-\nu^2) \leq 0,
$$

so that $G(p,\theta)$ increases for $\theta < 0$, reaches a maximum at $\theta = 0$, and then decreases for $\theta > 0$.

(*b*) $|p| > \tau\nu/\sqrt{\tau^2-\nu^2}$.
Then $(\tau^2+p^2)(\tau^2-\nu^2) > \tau^4$. Let

$$
\theta_p = \arccos\Big(\frac{\tau^2}{\sqrt{(\tau^2+p^2)(\tau^2-\nu^2)}}\Big). \qquad (4.49)
$$

We have $0 < \theta_p < \pi/2$, and θ_p increases from 0 to $\pi/2$ as $|p|$ increases from $\tau\nu/\sqrt{\tau^2-\nu^2}$ to ∞. Then (4.48) has the sign of θ for $0 < |\theta| < \theta_p$, vanishes for $\theta = 0$ or $\pm\theta_p$, and has the sign of $-\theta$ for $\theta_p < |\theta| < \pi/2$. Hence $G(p,\theta)$ increases for $-\pi/2 < \theta < -\theta_p$, reaches a maximum at $\theta = -\theta_p$, decreases for $-\theta_p < \theta < 0$, reaches a local minimum at $\theta = 0$, increases again for $0 < \theta < \theta_p$, reaches again a maximum at $\theta = \theta_p$, and finally decreases for $\theta_p < \theta < \pi/2$.

This shows that the correct orientation of the step edge is obtained only at points whose distance to the edge position does not exceed $\tau\nu/\sqrt{\tau^2-\nu^2}$. We illustrate this fact in Figure 10.

Note that the whole argument relies on the fact that $\tau > \nu$. Thus, following [31] and the result of the first example, we took a wider extent of the filter in the tangential direction in order to improve orientation selectivity, but we remark that precisely this wider extent restricts orientation selectivity to the neighbourhood of the edge position. We will thus seek criteria in order to guarantee correct orientation selection at the edge position only, knowing from Proposition 4.12 that the edge position can be found even when its orientation is not known. We introduce the following:

REQUIREMENT 7. *When $d = 2$, for every $\nu > 0$, $\widehat{C}(\nu\mathbf{n}_\theta)$ as a function of θ is strictly increasing for $-\pi/2 \le \theta < 0$, has a maximum at $\theta = 0$, and is strictly decreasing for $0 < \theta \le \pi/2$.*

Note that by Requirement 4 (that C and S are symmetric in x_t), we have $\widehat{C}(\nu\mathbf{n}_\theta) = \widehat{C}(\nu\mathbf{n}_{-\theta})$, so that the statement of Requirement 7 concerning $-\pi/2 \le \theta < 0$ is redundant; moreover, we know that $\widehat{C}(\nu\mathbf{n}_{\pi/2}) = \widehat{C}(\nu\mathbf{t}) = 0$, so that we can also omit the case where $\theta = \pi/2$. We obtain then the following result:

PROPOSITION 4.13. *Let $d = 2$ and I be given by $I(x_t, x_n) = P(x_n)$ for a one-dimensional profile P in $L^1 + L^2$. Let $p \in \mathbb{R}$ with local phase $P^\Phi(\nu, p)$ constant for all $\nu > 0$. Then for $x_t \in \mathbb{R}$ and $|\theta| \le \pi/2$, $|I * F_\theta|^2(x_t, p)$, the energy function at (x_t, p), is strictly increasing in θ for $-\pi/2 \le \theta < 0$, has a maximum at $\theta = 0$, and is and strictly decreasing in θ for $0 < \theta \le \pi/2$.*

PROOF. By Proposition 4.3, $C_{/\mathbf{n}_{-\theta}}$, $S_{/\mathbf{n}_{-\theta}}$, and $F_{/\mathbf{n}_{-\theta}}$ satisfy Requirements 1, 2', and 3 for $d = 1$. By (4.34) and (4.35), the one-dimensional profile of $I * F_\theta$ is given by $P * [F_\theta]_{/\mathbf{n}}$, with the Fourier transform of $[F_\theta]_{/\mathbf{n}}$ being given by $\widehat{F}(\nu\mathbf{n}_{-\theta})$. By Proposition 4.5, in particular (4.10), the corresponding energy function at point (x_t, p) is

$$\begin{aligned}
&4 \iint_{\mathbb{R}^+\times\mathbb{R}^+} P^{\mathcal{A}}(u)[F_\theta]_{/\mathbf{n}}{}^{\mathcal{A}}(u) P^{\mathcal{A}}(v)[F_\theta]_{/\mathbf{n}}{}^{\mathcal{A}}(v)\, \cos\big[P^\Phi(u,p) - P^\Phi(v,p)\big]\, dudv \\
=&4 \iint_{\mathbb{R}^+\times\mathbb{R}^+} P^{\mathcal{A}}(u) A(u\mathbf{n}_{-\theta}) P^{\mathcal{A}}(v) A(v\mathbf{n}_{-\theta})\, dudv,
\end{aligned}$$

where $A = C^{\mathcal{A}}$ (see (4.1)). The integrand is positive, strictly increasing in θ for $-\pi/2 \le \theta < 0$, has a maximum at $\theta = 0$, and is strictly decreasing in θ for $0 < \theta \le \pi/2$. ∎

In [Fousse] Requirement 7 was studied in the case where the filter F is separable, in other words $C(x_t, x_n) = b(x_t)\cdot c(x_n)$ and $S(x_t, x_n) = b(x_t)\cdot s(x_n)$, where b, c, and s are continuous and integrable, b and c are even-symmetric, s is odd-symmetric, $\widehat{b}$ and $\widehat{c}$ have real non-negative values, and (c, s) is a quadrature pair (see (4.2)). It was shown that this led to very strong conditions on $\widehat{b}$ and $\widehat{c}$. We have the following (with $\mathbb{R}^+$ being the set of reals $x \ge 0$):

LEMMA 4.14. *Let β and γ be derivable functions $\mathbb{R}^+ \to \mathbb{R}^+$; assume that γ is not monotonously increasing. Then the following three statements are equivalent:*

(*i*) *For every $u > 0$, $\gamma(u\cos\theta)\beta(u\sin\theta)$ is monotonously decreasing on $\theta \in [0, \pi/2]$.*

(*ii*) *For every $x, y > 0$ such that $\gamma(x) > 0$ and $\beta(y) > 0$, we have:*

$$\frac{\beta'(y)}{y\beta(y)} \le \frac{\gamma'(x)}{x\gamma(x)}.$$

(*iii*) *β is monotonously decreasing, and there is some $K > 0$ such that*

(a) *For every* $y > 0$ *having* $\beta(y) > 0$,

$$\frac{|\beta'(y)|}{y\beta(y)} \geq K.$$

(b) *For every* $x > 0$ *having* $\gamma(x) > 0$ *and* $\gamma'(x) < 0$,

$$\frac{|\gamma'(x)|}{x\gamma(x)} \leq K.$$

Proof. Note that for every $x > 0$, if $\gamma(x) = 0$, as γ has only non-negative values, γ has a minimum at x and so $\gamma'(x) = 0$. Similarly for every $y > 0$, if $\beta(y) = 0$, then $\beta'(y) = 0$.

We show first the equivalence between (i) and (ii). By the finite increment theorem, (i) is equivalent to:

$$\forall u > 0,\ \forall \theta,\ 0 < \theta < \pi/2, \quad \frac{\partial}{\partial \theta}\big(\gamma(u\cos\theta)\beta(u\sin\theta)\big) \leq 0.$$

Now

$$\frac{\partial}{\partial \theta}\big(\gamma(u\cos\theta)\beta(u\sin\theta)\big) = -u\sin\theta\gamma'(u\cos\theta)\beta(u\sin\theta) + u\cos\theta\gamma(u\cos\theta)\beta'(u\sin\theta),$$

so that the condition becomes

$$u\cos\theta\gamma(u\cos\theta)\beta'(u\sin\theta) \leq u\sin\theta\gamma'(u\cos\theta)\beta(u\sin\theta) \qquad \text{for} \quad u > 0 \quad \text{and} \quad 0 < \theta < \pi/2.$$

Now the set of pairs $(u\cos\theta, u\sin\theta)$ for $u > 0$ and $0 < \theta < \pi/2$ coincides with the set of pairs (x, y) for $x, y > 0$. Thus (i) is equivalent to

$$\forall x, y > 0, \quad x\gamma(x)\beta'(y) \leq y\gamma'(x)\beta(y). \tag{4.50}$$

Two special cases arise:

— $\gamma(x) = 0$, so that $\gamma'(x) = 0$.

— $\beta(y) = 0$, so that $\beta'(y) = 0$.

In both cases the inequality of (4.50) is trivially verified as $0 \leq 0$. We have thus only to consider the remaining case where $\gamma(x), \beta(y) > 0$, and dividing each member of the inequality $x\gamma(x)\beta'(y) \leq y\gamma'(x)\beta(y)$ by the positive factor $x\gamma(x)y\beta(y)$, (4.50) becomes (ii).

We show next that (iii) implies (ii). If (iii) is verified, as β is decreasing, for every $y > 0$ we have $\beta'(y) \leq 0$; if $\beta(y) > 0$, then (a) gives

$$-\frac{\beta'(y)}{y\beta(y)} = \frac{|\beta'(y)|}{y\beta(y)} \geq K.$$

Taking $x > 0$ with $\gamma(x) > 0$, either $\gamma'(x) \geq 0$ and so

$$-\frac{\gamma'(x)}{x\gamma(x)} \leq 0 \leq K,$$

or $\gamma'(x) < 0$ and (b) gives

$$-\frac{\gamma'(x)}{x\gamma(x)} = \frac{|\gamma'(x)|}{x\gamma(x)} \leq K.$$

Combining both inequalities, we get

$$-\frac{\beta'(y)}{y\beta(y)} \geq K \geq -\frac{\gamma'(x)}{x\gamma(x)},$$

which gives (ii) by a change of sign.

We show finally that (ii) implies (iii). Since γ is not monotonously increasing, there exist a, b with $0 \leq a < b$ and $\gamma(a) > \gamma(b)$. By the finite increment theorem, there is thus some $c > 0$ (with $a < c < b$) such that $\gamma'(c) < 0$; we must then have $\gamma(c) > 0$. By (ii), for every $y > 0$ such that $\beta(y) > 0$, we have:

$$\frac{\beta'(y)}{y\beta(y)} \leq \frac{\gamma'(c)}{c\gamma(c)} < 0,$$

in other words $\beta'(y) < 0$; on the other hand for $\beta(y) = 0$ we have $\beta'(y) = 0$. Thus β is monotonously decreasing. Let

$$K = \sup\{-\frac{\gamma'(z)}{z\gamma(z)} \mid z > 0 \text{ and } \gamma(z) > 0\}.$$

By (ii) we have for every $y > 0$ with $\beta(y) > 0$:

$$-\frac{|\beta'(y)|}{y\beta(y)} = \frac{\beta'(y)}{y\beta(y)} \leq \inf\{\frac{\gamma'(z)}{z\gamma(z)} \mid z > 0 \text{ and } \gamma(z) > 0\} = -K,$$

so that (a) holds. On the other hand for $x > 0$ with $\gamma(x) > 0$ and $\gamma'(x) < 0$ we have trivially

$$-\frac{|\gamma'(x)|}{x\gamma(x)} = \frac{\gamma'(x)}{x\gamma(x)} \geq \inf\{\frac{\gamma'(z)}{z\gamma(z)} \mid z > 0 \text{ and } \gamma(z) > 0\} = -K,$$

so that (b) holds. ∎

Now, by the symmetry of b and c, Requirement 7 reduces to the fact that for $u > 0$, $\widehat{C}(u\mathbf{n}_\theta) = \widehat{b}(u\sin\theta)\widehat{c}(u\cos\theta)$ is decreasing in θ for $0 \leq \theta \leq \pi/2$. Taking β and γ the restriction of $\widehat{b}$ and $\widehat{c}$ to non-negative frequencies, we have thus precisely condition (i); under the relatively weak assumptions that β and γ are derivable and that γ is not monotonously increasing, it is equivalent to (ii) and to (iii). These conditions were found by Fousse in [62], where the equivalence between (i) and (ii) and the sufficiency of (iii) were shown. If we return to our above examples where orientation selectivity fails, in the first one we had $b = \delta$, the Dirac impulse, whose Fourier transform is constant 1, so that condition (ii) gives $\gamma'(x) \geq 0$ whenever $\gamma(x) > 0$, in other words $\gamma = \widehat{c}$ is increasing on $\mathbb{R}^+$, which contradicts our assumption (indeed, this is impossible for c in $L^1 + L^2$). On the other hand in the second example, b was a Gabor cosine function, whose Fourier transform is not decreasing on $\mathbb{R}^+$, contradiciting condition (iii).

In the case of filters with polar-separable Fourier transforms (see (4.3)), Requirement 7 is more easily expressed in terms of the angular function.

We mentioned above the observation by Perona that orientation selectivity is improved when the spatial extent of the edge detection filter is greater in the tangential direction than in the normal one. Requirement 7 provides a rationale for it. Indeed, the wider the spatial extent of F in the tangential direction, the narrower the spatial extent of its Fourier transform $\widehat{F}$ in that direction, and so the lower will be $\widehat{F}(u\cos\theta, u\sin\theta)$ for $u > 0$ and $\theta \neq 0$.

The traditional method for localizing edges with an edge-detecting filter F consists in two steps:

(1°) First compute at every point $\mathbf{p}$ the angle θ for which the energy function $|(I * F_\theta)(\mathbf{p})|^2$ is the greatest; write it $\theta(\mathbf{p})$.

(2°) Second select as edge position the set all points $\mathbf{p}$ such that $|(I * F_{\theta(\mathbf{p})})(\mathbf{p})|^2 \geq |(I * F_{\theta(\mathbf{q})})(\mathbf{q})|^2$ for all points $\mathbf{q}$ in a neighbourhood of $\mathbf{p}$ in the normal direction.

The neighbourhood in (2°) can be purely local, or have some extent depending on the width of the filter F (we will return to this question in the next subsection). This method, although introduced in an empirical framework based on template matching, is justified in the phase congruence model by Propositions 4.12 and 4.13. Indeed, assume as above an image I forming a one-dimensional edge profile P in the normal direction: $I(x_t, x_n) = P(x_n)$. Then we know that the phase of $|(I * F_{\theta(\mathbf{p})})|^2$ does not depend on $\theta(\mathbf{p})$ (provided that $\theta(\mathbf{p})$ is not oriented in the edge tangential direction), so that if $\mathbf{p}$ is not on the edge, we will have

$$|(I * F_{\theta(\mathbf{p})})(\mathbf{p})|^2 < |(I * F_{\theta(\mathbf{p})})(\mathbf{q})|^2 \leq |(I * F_{\theta(\mathbf{q})})(\mathbf{q})|^2$$

for some neighbouring point $\mathbf{q}$ in the normal direction; on the other hand if $\mathbf{p}$ is on the edge, then we know that $\theta(\mathbf{p})$ will be along the normal direction of the edge. Thus (2°) will eliminate points which do not lie on the edge, while (1°) will give the edge orientation on edge points.

Of courses, this argument based on an ideal one-dimensional edge is justified for straight edges, and does not hold when we have strongly curved edges (in other words edges whose radius of curvature is comparable to the spatial extent of the filter).

4.5. Authentication of edges, and scale-space behaviour of the energy function

A priori, any local maximum of the energy function in the normal direction could be selected as an edge point. But this could produce spurious edges, for example a point where the local phases of the image are only slightly less discordant than in its neighbourhood. Moreover, as we saw in Subsection 4.1 when we introduced three spatial constraints on the filters, it cannot be excluded that filters satisfying our seven requirements may respond to a pure edge (an ideal step, line, or roof) with an energy function having one global maximum at the edge location and also other local maxima which do not correspond to any perceptually meaningful feature in the image. We must thus find methods for avoiding spurious maxima in the energy function. In the first place, we can select the filters carefully in order to satisfy the three spatial constraints, as well as any others which might arise from other models of perfect edges. However we need also to eliminate some of the maxima obtained in the energy function. Two general criteria can be envisaged:

(i) The peak in the energy function must be high enough.

(ii) The peak in the energy function must be wide enough.

If we follow (i) and require high peaks, we must first calibrate the energy function in order to make it independent of the average image contrast: if the image grey-levels are all multiplied by a positive constant, the energy function should not change. We can for example use as measure of the energy at a point $\mathbf{p}$ the *phase congruence function*

$$\frac{|J(\mathbf{p})|}{\int_{\mathcal{E}} I^{\mathcal{A}} A} \tag{4.51}$$

introduced in (4.17), whose values range in the interval $[-1, 1]$ and is equal to 1 only when all phases $I^\Phi(\mathbf{u}, \mathbf{p})$ at $\mathbf{p}$ for all frequencies $\mathbf{u} \in \mathcal{E}^+_{\mathbf{n}}$ are equal to a constant. We can also take some variants such as

$$\frac{|J(\mathbf{p})|}{\|\widehat{I}\|_\infty \cdot \|\widehat{C}\|_1} \tag{4.52}$$

when I is integrable, or

$$\frac{|J(\mathbf{p})|}{\|\widehat{I}\|_2 \cdot \|\widehat{C}\|_2} = \frac{|J(\mathbf{p})|}{\|I\|_2 \cdot \|C\|_2} \tag{4.53}$$

when I is square-integrable; note that $\|\widehat{I}\|_\infty \cdot \|\widehat{C}\|_1$ and $\|\widehat{I}\|_2 \cdot \|\widehat{C}\|_2$ are both $\geq \int_{\mathcal{E}} I^{\mathcal{A}} A$, thanks to Hölder's inequality, so that the values of (4.52) and (4.53) range also in the interval $[-1, 1]$. Then we might select as edge points only maxima of the energy function for which one of the measures $(4.51, 4.52, 4.53)$ of phase congruence exceeds some threshold (say, 1/2).

On the other hand, requiring wide peaks as in (*ii*) does not necessitate a calibration of the energy function, but rather a standard width to compare the peaks with. We can consider that this must be the width of the grey-level profile of F in the normal direction, because $|F|^2$ is the energy function for the input signal given by a Dirac impulse. This width can be measured as

$$\frac{\|\xi_{\mathbf{n}} F\|_1}{\|F\|_1} = \frac{\int_{\mathcal{E}} |x_n F(x_t, x_n)| \, dx_t dx_n}{\int_{\mathcal{E}} |F(x_t, x_n)| \, dx_t dx_n}$$

or

$$\frac{\|\xi_{\mathbf{n}} F\|_2}{\|F\|_2} = \left(\frac{\int_{\mathcal{E}} x_n^2 |F(x_t, x_n)|^2 \, dx_t dx_n}{\int_{\mathcal{E}} |F(x_t, x_n)|^2 \, dx_t dx_n} \right)^{\frac{1}{2}},$$

or the least w such that for $|x_n| > w$ we have $|F(x_t, x_n)| < \varepsilon \|F\|_p$ (where $p = 1$, 2, or ∞), or such that $\int_{|x_n|>w} |F|^p < \varepsilon \int_{\mathcal{E}} |F|^p$ (where $p = 1$ or 2), with ε being chosen very small (say, 1/100). We can also take as standard input signal a one-dimensional line constant in the tangential direction and making a Dirac impulse in the normal one, and so we replace F by $F_{/\mathbf{n}}$ in the above formulas.

Then we might select as edge points those for which the energy function is greater than that of all points at distance at most kw of it in the normal direction, where k is some threshold smaller than 1 (say, 1/2). In other words we admit edges separated by a distance between kw and w, but for closer edges one of them must be condidered as spurious. We might also require of an edge point $\mathbf{p}$ that for all points $\mathbf{q}$ in the normal direction w.r.t. $\mathbf{p}$ and at distance less than kw from it, the ratio of energies $E(\mathbf{p})/E(\mathbf{q})$ must increase with the distance between $\mathbf{p}$ and $\mathbf{q}$ according to some function of that distance:

$$\frac{E(\mathbf{p})}{E(\mathbf{q})} \geq \psi\big(d(\mathbf{p}, \mathbf{q})\big);$$

this function ψ could be selected from the what happens with the Dirac impulse (or Dirac impulse line) as input signal, in other words the grey-level profile of $|F|^2$ (or $|F_{/\mathbf{n}})$ around the origin. The effect of this stronger criterion would be to consider a close succession of edges as texture or noise, which should not appear in the edge map.

The two approaches could be combined, so that we might require peaks to be both wide and high, or that the product of the peak height and width should exceed a given threshold. These considerations are speculative, we have no mathematical result justifying them, and the criteria suggested above should be experimented with natural images. Morphological operators [63] could also be applied in order to eliminate spurious peaks, and the watershed transformation could be used instead of non-maxima deletion in order to produce closed contours.

We said in Section 2 that each feature corresponds to a certain scale, that a change of scale can lead to modifications of the edges; this was illustrated with Figure 5. Now it should be clear that the scale corresponding to an edge is proportional to the above-mentioned width w of the filter F detecting it. Consider for example the slanted ridge profile shown in Figure 5. When the filter F is very narrow (say, w is 1/40-th the width of the ridge), the energy function will have four peaks corresponding to the four Mach bands; these peaks will be well separated and will not interfere. As the filter width increases (say, w going up to to 1/6-th the width of the ridge), the two energy peaks at the extremities of each step will interfere, and will at a certain scale merge into a single peak; thus two steps are detected. Increasing further the filter width (say, until w becomes larger than the width of the ridge), the two energy peaks located at each step will interfere and finally merge into a single peak; thus a single bar edge is detected.

In this example we see that as the scale (i.e., the width) of the filter increases, features can merge into other ones whose nature is different; we had here no new feature arising at a certain scale from nothing, or a feature dividing into several ones at a wider scale. This is the principle of *causality*, and we could require that features should abide to it. Thus any feature existing at a certain scale which does not arise from one at a smaller scale must be considered as spurious.

Note that points where the Fourier phases of the image are equal to a constant lead to an absolute maximum of the energy function at all scales. This is true either if we scale the filter F (taking at scale s the filter F_s defined by $F_s(\mathbf{x}) = F(\mathbf{x}/s)$), or if we smooth the image with a function W having constant zero phase (say, a Gaussian of increasing scale), because the phases of $I * W$ are the same as those of I (cfr. [Ronse]).

Kube and Perona [17] have shown in the case of one-dimensional signals that quadratic edge detectors using a Hilbert transform pair of filters are non-causal in scale space; this was illustrated in concrete examples with the pair of filters consisting of a Gaussian derivative and its Hilbert transform. For quadratic operators where one of the filters is the derivative of the other, only for a certain class of filters which includes the Gaussian and its derivatives, is the causality property verified. Thus the phase congruence model introduces spurious non-causal features which are not detected by quadratic edge detectors using Gaussian derivatives. A typical pattern of causality failure, where a new feature arises from nothing as scale increases, is shown in Figure 11 (see also Figures 2 and 3 of [17]). We see here that a purely local maximum arises in the energy function, which could be eliminated by taking as edge points only regional maxima of the energy function, as explained above. It would be interesting to know if the two above-mentioned approaches for authenticating edges could allow the elimination of all causality failures.

5. Related questions and conclusion

We will discuss here miscellaneous problems concerning the phase congruence model, in particular: (*i*) how to adapt it to a digital framework and digitize the filters specified by our requirements in the Euclidean framework; (*ii*) possible ways to extend it towards the detection of bi-directional features (corners, end-stopped edges, or strongly curved edges); (*iii*) applications to other vision tasks. We end then with the conclusion.

5.1. *Digitization of the filters*

Any practical implementation must assume that images and filters are sampled, and that filters have bounded support. The classical signal processing approach to digitization, based on the Shannon sampling theorem, is unsuitable for the analysis of visual features, because it assumes unnecessarily that the signal can be band-limited, gives a secondary role to the spatial localization of masks, which is however crucial to to the spatial localization of features, and does not guarantee that our mathematical results will extend to the digital case. It is better to take the approach suggested by Hummel and Lowe [64], which we describe here in general mathematical terms.

We already have the Euclidean space $\mathcal{E} = \mathbb{R}^d$; consider now the corresponding digital space $\mathcal{D} = \mathbb{Z}^d$. Let $\mathcal{I}(\mathcal{E})$, $\mathcal{I}(\mathcal{D})$, $\mathcal{F}(\mathcal{E})$, $\mathcal{F}(\mathcal{D})$ be the families of respectively Euclidean images, digital images, Euclidean filters, and digital filters. Theory gives us a Euclidean filter $F = C + i\,S$; practice gives us a digital input image I; we must be able to apply F to I, and for this we must digitize F. The basic idea underlying the method of Hummel and Lowe is that the sampling of filters must be considered as corresponding to an extrapolation of digital images into Euclidean ones. Thus the filter sampling $\Sigma : \mathcal{F}(\mathcal{E}) \to \mathcal{F}(\mathcal{D})$ corresponds to the image extrapolation $\Xi : \mathcal{I}(\mathcal{D}) \to \mathcal{I}(\mathcal{E})$ in such a way that for every $F \in \mathcal{F}(\mathcal{E})$ and $I \in \mathcal{I}(\mathcal{D})$ we have

$$\int_{\mathcal{D}} I \cdot \Sigma[F] = \int_{\mathcal{E}} \Xi[I] \cdot F. \tag{4.54}$$

Note that in this equation, integration on $\mathcal{E}$ is done w.r.t. the Lebesgue measure, while integration on $\mathcal{D}$ is done w.r.t. the discrete measure, in other words $\int_{\mathcal{D}} I \cdot \Sigma[F]$ must be read as $\sum_{\mathbf{z}\in\mathcal{D}} I(\mathbf{z}) \cdot \Sigma[F](\mathbf{z})$. Assuming that Ξ and Σ commute with the reflection ρ and with all translations by points in $\mathcal{D}$, (4.54) gives:

$$\forall \mathbf{p} \in \mathcal{D}, \qquad \big(I * \Sigma[F]\big)(\mathbf{p}) = \big(\Xi[I] * F\big)(\mathbf{p}). \tag{4.55}$$

Here the first convolution is made on $\mathcal{D}$ (with the integral becoming a series), and the second one on $\mathcal{E}$. Given an even-symmetric bounded integrable function $W : \mathcal{E} \to \mathbb{R}$, we define Ξ and Σ by

$$\begin{aligned} \Xi[I] &= \sum_{\mathbf{z}\in\mathcal{D}} I(\mathbf{z})\tau_{\mathbf{z}}(W) : \ \mathbf{x} \mapsto \sum_{\mathbf{z}\in\mathcal{D}} I(\mathbf{z}) W(\mathbf{x}-\mathbf{z}) \\ \text{and} \qquad \Sigma[F] &: \ \mathbf{z} \mapsto (W * I)(\mathbf{z}) = \int_{\mathcal{E}} d\mathbf{x}\, W(\mathbf{z}-\mathbf{x}) F(\mathbf{x}). \end{aligned} \tag{4.56}$$

It is easily checked that Ξ and Σ commute with the reflection and with all translations by points in $\mathcal{D}$, and (4.55) is verified as follows:

$$\begin{aligned} \big(I * \Sigma[F]\big)(\mathbf{p}) &= \int_{\mathcal{E}} d\mathbf{x} \sum_{\mathbf{z}\in\mathcal{D}} I(\mathbf{z}) W(\mathbf{p}-\mathbf{z}-\mathbf{x}) F(\mathbf{x}) \\ &= \sum_{\mathbf{z}\in\mathcal{D}} \int_{\mathcal{E}} d\mathbf{x}\, I(\mathbf{z}) W(\mathbf{p}-\mathbf{z}-\mathbf{x}) F(\mathbf{x}) = \big(\Xi[I] * F\big)(\mathbf{p}). \end{aligned}$$

This equality holds indeed if we assume I in $\ell^1 + \ell^2$ (i.e., to be the sum of a summable image and a square-summable one). This was the choice for Ξ and Σ in [64], where several examples of functions W were considered, and the advantages of this new sampling method over the classical one was experimentally demonstrated.

Note that by (4.55) the digital image $I * \Sigma[F]$ resulting from the application of the sampled filter $\Sigma[F]$ to the original digital image I, is equal to the classical sampling (in other words, the digital trace) of the Euclidean image $\Xi[I] * F$ resulting from the application of the original Euclidean filter F to the extrapolated image $\Xi[I]$. Now the mathematical properties of the phase congruence model can be applied to $\Xi[I] * F$, of which $I * \Sigma[F]$ is a sampling. We can thus expect that some properties of $\Xi[I] * F$ will be inherited by $I * \Sigma[F]$. For example if I is a digital edge profile, with a proper choice of W, $\Xi[I]$ will be a Euclidean edge profile, so that $|\Xi[I] * F|$ will have a peak at the edge location $\mathbf{x}$. Now assuming I to be in $\ell^1 + \ell^2$, $\Xi[I]$ will be in $L^1 + L^2$, so that $\Xi[I] * F$ will be uniformly continuous (by Proposition 4.2), so that we know that at the digital point $\mathbf{z}$ closest to the peak $\mathbf{x}$ of $|\Xi[I] * F|$, we will have

$$\big|\big(I * \Sigma[F]\big)(\mathbf{z})\big| - \big|\big(\Xi[I] * F\big)(\mathbf{x})\big| = \big|\big(\Xi[I] * F\big)(\mathbf{z})\big| - \big|\big(\Xi[I] * F\big)(\mathbf{x})\big| \leq \psi(\mathbf{z} - \mathbf{x}),$$

where ψ is the modulus of continuity of $\Xi[I] * F$, so that this difference can become arbitrary small, provided that the digitization step is taken small enough. Thus we can expect a peak in $|I * \Sigma[F]|$ at some digital point close to $\mathbf{x}$.

Of courses, it is possible to make a theory of phase congruence for digital images. Note that the Fourier spectrum of digital signals is periodic, so that the distinction of "positive" and "negative" frequencies becomes arbitrary.

5.2. Bi-directional features

We suppose again that $d = 2$, that is $\mathcal{E} = \mathbb{R}^2$. Up to now we have assumed features with a significant event in the grey-levels along a direction $\mathbf{n}$, but without such grey-level events in the perpendicular direction $\mathbf{t}$. We will now briefly consider features having perceptually significant grey-level changes in two directions given by unit vectors $\mathbf{n}_1$ and $\mathbf{n}_2$, which are not necessarily perpendicular. We will first consider the case where $\mathbf{n}_1$ and $\mathbf{n}_2$ are perpendicular and coincide with the canonical basis vectors $\mathbf{e}_1$ and $\mathbf{e}_2$ of $\mathcal{E}$; the general case where $\mathbf{n}_1$ and $\mathbf{n}_2$ are not perpendicular will be dealt with later, thanks to a change of basis of $\mathcal{E}$. Let us start with some examples of ideal bi-directional features:

(a) *Corners.* An ideal corner (see Figure 12 (a)) has its grey-levels given by:

$$I(x_1, x_2) = \begin{cases} a & \text{if } x_1 > 0 \text{ and } x_2 > 0, \\ b & \text{if } x_1 < 0 \text{ or } x_2 < 0, \end{cases}$$

where a and b are constants. Recall the Heaviside step function defined by

$$H(x) = \begin{cases} 0 & \text{if } x < 0, \\ 1 & \text{if } x > 0; \end{cases} \tag{4.57}$$

then we have

$$I(x_1, x_2) = b + (a - b)H(x_1)H(x_2).$$

Thus we have, up to a constant, the product of a step edge in x_1 and one in x_2.

(b) *T-junctions.* An ideal T-junction (see Figure 12 (b)) can be modeled as

$$I(x_1, x_2) = \begin{cases} a & \text{if } x_1 > 0 \text{ and } x_2 > 0, \\ b & \text{if } x_1 < 0 \text{ and } x_2 > 0, \\ c & \text{if } x_2 < 0, \end{cases}$$

with a, b, and c being constants. This can be written as

$$I(x_1, x_2) = c + H(x_2)\big[(b - c) + (a - b)H(x_1)\big].$$

We get again, up to a constant, the product of a step edge in x_1 and one in x_2.

(c) *Line terminations.* We can model a line termination (see Figure 12 (c)) as an image containing a line edge on one side, and nothing on the other side; it is thus given by the following grey-level function:

$$I(x_1,x_2) = a + E(x_1)H(x_2) = \begin{cases} a & \text{if } x_2 < 0, \\ a + E(x_1) & \text{if } x_2 > 0, \end{cases}$$

where a is a constant and E is the line edge profile in the normal direction. We have here, up to a constant, the product of a feature in x_1 and of a step edge in x_2.

(d) *X-junctions.* An ideal X-junction (see Figure 12 (d)) can be modeled as the grey-level function

$$I(x_1,x_2) = \begin{cases} a & \text{if } x_1 > 0 \text{ and } x_2 > 0, \\ b & \text{if } x_1 < 0 \text{ and } x_2 > 0, \\ c & \text{if } x_1 > 0 \text{ and } x_2 < 0, \\ d & \text{if } x_1 < 0 \text{ and } x_2 < 0, \end{cases}$$

with a, b, c, and d being constants. We have two cases:

($d1$) $a + d - b - c = 0$. Then we have

$$I(x_1,x_2) = d + (c-d)H(x_1) + (b-d)H(x_2).$$

Here we have, up to a constant, the sum of a step edge in x_1 and one in x_2.

($d2$) $a + d - b - c \neq 0$. Let us define the two step edges

$$S_1(x) = (b-d) + (a+d-b-c)H(x) = \begin{cases} b-d & \text{if } x < 0, \\ a-c & \text{if } x > 0, \end{cases}$$

$$S_2(x) = (c-d) + (a+d-b-c)H(x) = \begin{cases} c-d & \text{if } x < 0, \\ a-b & \text{if } x > 0; \end{cases}$$

Then we can write:

$$I(x_1,x_2) = \frac{1}{a+d-b-c}\big[ad - bc + S_1(x_1)S_2(x_2)\big].$$

We get, up to a constant, the product of a step edge in x_1 and one in x_2.

(e) *Peaks.* An ideal peak (see Figure 12 (e)) can be considered as the product of a line in x_1 and a line in x_2.

We see that in all these examples, except ($d1$), the image can be decomposed as the sum of a constant signal and of the product E of an ideal feature E_1 in x_1 and another one E_2 in x_2. The constant signal contributes to the zero frequency in the Fourier plane. According to the phase congruence model, there are two constant phase ϕ_1 and ϕ_2 such that

$$E_1^{\Phi}(\nu) = \operatorname{sgn}(\nu)\phi_1 \qquad \text{and} \qquad E_2^{\Phi}(\nu) = \operatorname{sgn}(\nu)\phi_2.$$

Note that if E_1 or E_2 involves a constant signal (the "dc level"), this contributes to the frequency zero. We get then

$$E^{\Phi}(\nu_1,\nu_2) = E_1^{\Phi}(\nu_1) + E_2^{\Phi}(\nu_2) = \operatorname{sgn}(\nu_1)\phi_1 + \operatorname{sgn}(\nu_2)\phi_2 = \begin{cases} \phi_1 + \phi_2 & \text{if } \nu_1 > 0 \text{ and } \nu_2 > 0, \\ \phi_1 - \phi_2 & \text{if } \nu_1 > 0 \text{ and } \nu_2 < 0, \\ -\phi_1 + \phi_2 & \text{if } \nu_1 < 0 \text{ and } \nu_2 > 0, \\ -\phi_1 - \phi_2 & \text{if } \nu_1 < 0 \text{ and } \nu_2 < 0. \end{cases}$$

The Fourier component along the axes $\nu_1 = 0$ and $\nu_2 = 0$ corresponds to the constant base signal (or "dc level") involved in E_1, E_2, or E, and we will neglect it. Thus, instead of having a constant phase in a half-plane $\mathcal{E}^+_{\mathbf{n}}$, and the opposite constant phase in the opposite half-plane $\mathcal{E}^-_{\mathbf{n}}$, we will have constant phases in all quadrants of the plane.

From the point of view of filters, taking (C_1, S_1) and (C_2, S_2) two pairs of one-dimensional filters satisfying Requirements 1, 2', and 3, the edge in the one-dimensional signal E_1 will produce a peak in the energy function

$$(E_1 * C_1)^2 + (E_1 * S_1)^2 = \big|(E_1 * C_1) \pm i\,(E_1 * S_1)\big|^2,$$

while the edge in E_2 will also produce a peak in the energy function

$$(E_2 * C_2)^2 + (E_2 * S_2)^2 = \big|(E_2 * C_2) \pm i\,(E_2 * S_2)\big|^2,$$

so that the bi-directional feature in E will produce a peak in the function

$$\begin{aligned}&\big[(E_1 * C_1)^2(x_1) + (E_1 * S_1)^2(x_1)\big] \cdot \big[(E_2 * C_2)^2(x_2) + (E_2 * S_2)^2(x_2)\big]\\ =&\Big|\big[E_1 * (C_1 \pm i\,S_1)\big](x_1) \cdot \big[E_2 * (C_2 \pm i\,S_2)\big](x_2)\Big|^2.\end{aligned}$$

This amounts to convolving E with the four filters $C_1(x_1)C_2(x_2)$, $S_1(x_1)C_2(x_2)$, $C_1(x_1)S_2(x_2)$, $S_1(x_1)S_2(x_2)$, and taking the sum of the squares of these convolutions. Note that these four two-dimensional filters have a null Fourier component along the axes $\nu_1 = 0$ and $\nu_2 = 0$, and so they do not respond to any component of the input image which is constant in direction x_1 or x_2.

We can generalize the above discussion into a phase congruence model for bidirectional edges and a quadratic filter approach to their detection. We write $\mathcal{Q}^{++}$, $\mathcal{Q}^{+-}$, $\mathcal{Q}^{-+}$, and $\mathcal{Q}^{--}$, for the four quadrants given repectively by the equations

$$\begin{cases} x_1 > 0, \\ x_2 > 0, \end{cases} \qquad \begin{cases} x_1 > 0, \\ x_2 < 0, \end{cases} \qquad \begin{cases} x_1 < 0, \\ x_2 > 0, \end{cases} \qquad \begin{cases} x_1 < 0, \\ x_2 < 0, \end{cases}$$

see Figure 13 (a). We write P^{++}, P^{+-}, P^{-+}, and P^{--}, for the characteristic functions of $\mathcal{Q}^{++}$, $\mathcal{Q}^{+-}$, $\mathcal{Q}^{-+}$, and $\mathcal{Q}^{--}$ respectively, that is:

$$\begin{aligned} P^{++} &= \mathrm{pos}_{\mathbf{e}_1} \cdot \mathrm{pos}_{\mathbf{e}_2}, \\ P^{+-} &= \mathrm{pos}_{\mathbf{e}_1} \cdot \mathrm{neg}_{\mathbf{e}_2}, \\ P^{-+} &= \mathrm{neg}_{\mathbf{e}_1} \cdot \mathrm{pos}_{\mathbf{e}_2}, \\ P^{--} &= \mathrm{neg}_{\mathbf{e}_1} \cdot \mathrm{neg}_{\mathbf{e}_2}. \end{aligned}$$

We will consider as a bidirectional edge at point $\mathbf{p}$ an image such that the local phases at $\mathbf{p}$ are maximally congruent for frequencies in each quadrant. Since Fourier phase is odd-symmetric, we have to check phase congruence only in two quadrants making an half-plane, say $\mathcal{Q}^{++}$ and $\mathcal{Q}^{+-}$. An ideal edge, as above, will have a constant phase in each of these two quadrants.

We take as convolution kernels four functions F_0, F_1, F_2, and F_3 satisfying Requirements 2' and 3 (namely, they are continuous and integrable), and the following variant of Requirement 1:

$F_0 \not\equiv 0$, F_0 and F_3 are even-symmetric, F_1 and F_2 are odd-symmetric, $\widehat{F_0}$ has real non-negative values, (F_0, F_1) and (F_2, F_3) are $\mathbf{e}_1$-quadrature pairs, while (F_0, F_2) and (F_1, F_3) are $\mathbf{e}_2$-quadrature pairs.

The signs of the Fourier transform of F_0, F_1, F_2, and F_3 in the four quadrants is shown in Figure 13 (b). Note that the four filters satisfy Propositions 4.1 and 4.2. Moreover, each of them has a zero response on a signal which is constant in direction $\mathbf{e}_1$ or $\mathbf{e}_2$, in particular to a one-dimensional feature oriented in one of these two directions. This corresponds to the fact that, due to the quadrature relations, their Fourier transforms vanish on the two axes $x_1 = 0$ and $x_2 = 0$. Let us write A for the Fourier transform of F_0 (cfr. (4.1)).

Let us remark that the pair (F_0, F_1) satisfies Requirement 1 for the "normal" direction $\mathbf{e}_1$, and similarly the pair (F_0, F_2) satisfies Requirement 1 for the "normal" direction $\mathbf{e}_2$; however they do not give an edge detector in these two directions, because they do not satisfy Requirement 6: they give a zero response to an image whose greylevels are constant in the "tangential" direction and vary according to the "normal" direction (as in (4.8)).

Given an image I, let us write J_0, J_1, J_2, and J_3 for the convolution of I by F_0, F_1, F_2, and F_3 respectively. Then I gives rise to the energy function

$$E = J_0^2 + J_1^2 + J_2^2 + J_3^2 = (I * F_0)^2 + (I * F_1)^2 + (I * F_2)^2 + (I * F_3)^2. \tag{5.1}$$

Edge points are localized at points where E forms a maximum in both directions $\mathbf{e}_1$ and $\mathbf{e}_2$. Let us show that this energy function measures phase congruence in both quadrants $\mathcal{Q}^{++}$ and $\mathcal{Q}^{+-}$, by giving an analogue of Proposition 4.5:

PROPOSITION 5.1. *Assume that I is in $L^1 + L^2$. Then*

$$\widehat{E} = 8\left[(P^{++}\,\widehat{I}A) * (P^{--}\,\widehat{I}A) + (P^{+-}\,\widehat{I}A) * (P^{-+}\,\widehat{I}A)\right],$$

and it is an integrable function. Furthermore:

(*i*) *If I^Φ is constant both on $\mathcal{Q}^{++}$ and on $\mathcal{Q}^{+-}$, then*

$$\widehat{E} = 8\left[(P^{++}\,I^{\mathcal{A}}A) * (P^{--}\,I^{\mathcal{A}}A) + (P^{+-}\,I^{\mathcal{A}}A) * (P^{-+}\,I^{\mathcal{A}}A)\right],$$

a real-valued non-negative function, and E^Φ is constant zero.

(*ii*) *If for a given $\mathbf{p} \in \mathcal{E}$ we have $I^\Phi(\mathbf{u},\mathbf{p})$ constant both for $\mathbf{u} \in \mathcal{Q}^{++}$ and for $\mathbf{u} \in \mathcal{Q}^{+-}$, then $E^\Phi(\mathbf{u},\mathbf{p}) = 0$ for all $\mathbf{u} \in \mathcal{E}$, and $E(\mathbf{p}) > E(\mathbf{x})$ for all $\mathbf{x} \in \mathcal{E}$ such that $\mathbf{x} \neq \mathbf{p}$.*

(*iii*) *For every $\mathbf{p} \in \mathcal{E}$ we have*

$$E(\mathbf{p}) = 8 \iint_{[\mathcal{Q}^{++}\times\mathcal{Q}^{++}]\cup[\mathcal{Q}^{+-}\times\mathcal{Q}^{+-}]} I^{\mathcal{A}}(\mathbf{u})A(\mathbf{u})I^{\mathcal{A}}(\mathbf{v})A(\mathbf{v})\, \cos\left[I^\Phi(\mathbf{u},\mathbf{p}) - I^\Phi(\mathbf{v},\mathbf{p})\right] d\mathbf{u}d\mathbf{v}. \tag{5.2}$$

PROOF. Since F_0 and F_1 are in $\mathbf{e}_1$-quadrature, we have $(F_0 + i\,F_1)^\wedge = 2\,\mathrm{pos}_{\mathbf{e}_1} \cdot \widehat{F_0} = 2\,\mathrm{pos}_{\mathbf{e}_1} \cdot A$. Since (F_0, F_2) and (F_1, F_3) are $\mathbf{e}_2$-quadrature pairs, $F_0 + i\,F_1$ and $F_2 + i\,F_3$ are in $\mathbf{e}_2$-quadrature, so that for

$$F^{++} = (F_0 + i\,F_1) + i\,(F_2 + i\,F_3) = (F_0 - F_3) + i\,(F_1 + F_2)$$

and

$$F^{+-} = (F_0 + i\,F_1) - i\,(F_2 + i\,F_3) = (F_0 + F_3) + i\,(F_1 - F_2)$$

we have

$$(F^{++})^\wedge = 2\,\mathrm{pos}_{\mathbf{e}_2} \cdot (F_0 + i\,F_1)^\wedge = 4\,\mathrm{pos}_{\mathbf{e}_2} \cdot \mathrm{pos}_{\mathbf{e}_1} \cdot A = 4\,P^{++}A$$

and

$$(F^{+-})^\wedge = 2\,\mathrm{neg}_{\mathbf{e}_2} \cdot (F_0 + i\,F_1)^\wedge = 4\,\mathrm{neg}_{\mathbf{e}_2} \cdot \mathrm{pos}_{\mathbf{e}_1} \cdot A = 4\,P^{+-}A.$$

Let us write J^{++}, and J^{+-} for the convolution of I by F^{++}, and F^{+-} respectively. Thus

$$\begin{aligned} J^{++} &= (J_0 - J_3) + i\,(J_1 + J_2), \qquad (J^{++})^{\wedge} = 4\,P^{++}\widehat{I}A, \\ J^{+-} &= (J_0 + J_3) + i\,(J_1 - J_2), \qquad (J^{+-})^{\wedge} = 4\,P^{+-}\widehat{I}A. \end{aligned} \tag{5.3}$$

Now we can verify that

$$\begin{aligned} & J^{++}\overline{J^{++}} + J^{+-}\overline{J^{+-}} = |J^{++}|^2 + |J^{+-}|^2 \\ & = (J_0 - J_3)^2 + (J_1 + J_2)^2 + (J_0 + J_3)^2 + (J_1 - J_2)^2 = 2\left(J_0^2 + J_1^2 + J_2^2 + J_3^2\right) = 2E. \end{aligned} \tag{5.4}$$

Our result is then obtained by applying the same argument as in the proof of Proposition 4.5, replacing F and J first by F^{++} and J^{++}, next by F^{+-} and J^{+-}, and using (5.3); adding the formulas obtained for the partial energy functions $|J^{++}|^2$ and $|J^{+-}|^2$, we get those for the global energy function E thanks to (5.4). ∎

Thus bidirectional features correspond to points of maximum phase congruence within each quadrant instead of each half-plane (as was the case with unidirectional features). We may ask what are the analogues of Requirements 4 to 7. Requirement 4 becomes that:

F_0 is symmetric in both directions $\mathbf{e}_1$ and $\mathbf{e}_2$, in other words

$$\forall x_1, x_2 \in \mathbb{R}, \qquad F_0(x_1, x_2) = F_0(-x_1, x_2) = F_0(x_1, -x_2) = F_0(-x_1, -x_2).$$

From the quadrature relations between the filters, it follows that F_1 is odd-symmetric along $\mathbf{e}_1$ and even-symmetric along $\mathbf{e}_2$, F_2 is even-symmetric along $\mathbf{e}_1$ and odd-symmetric along $\mathbf{e}_2$, while F_3 is odd-symmetric in both directions $\mathbf{e}_1$ and $\mathbf{e}_2$; in summary, we can write:

$$\forall \eta, \zeta = \pm 1 \quad \text{and} \quad i, j = 0, 1, \qquad F_{2i+j}(\eta x_1, \zeta x_2) = \zeta^i \eta^j F_{2i+j}(x_1, x_2).$$

Although it corresponds to traditional intuition about edge detection, Requirement 4 was not used in any of our mathematical results, except for simplifying the statement of Requirement 7. We may expect the same for its present analogue.

Requirement 5 can be taken verbatim with the filters F_i intead of C and S. Requirement 6 may be translated as follows:

For every unit vector $\mathbf{v} \neq \pm\mathbf{e}_1, \pm\mathbf{e}_2$, $(F_0)_{/\mathbf{v}}$ is not identically zero, in other words $\widehat{F}_0(u \cdot \mathbf{v}) \neq 0$ for some $u \in \mathbb{R}$.

It follows from these analogues of Requirements 5 and 6 that for an edge forming a signal constant in a tangential direction and varying in the orthogonal normal direction according to a given one-dimensional profile, convolution with the filters will give a zero output if the normal direction is aligned either with $\mathbf{e}_1$ or with $\mathbf{e}_2$; on the other hand if this normal direction is oblique w.r.t. $\mathbf{e}_1$ and $\mathbf{e}_2$, then taking either F_0 or F_3 with either F_1 or F_2 (in fact filters in each pair give the same result, up to a ± 1 factor), the requirements of the one-dimensional phase congruence model will be satisfied, and so Proposition 4.12 will be verified.

We see no way of generalizing Requirement 7, because it applies to the case of an image forming a unidirectional grey-level profile. If we consider an ideal bidirectional edge as an image I given by $I(x_1, x_2) = P_1(x_1)P_2(x_2)$ for two one-dimensional functions P_1 and P_2 having constant Fourier phase for postive frequencies, the argument underlying Proposition 4.13 would require the condition that for every angle θ with $|\theta| \leq \pi/2$, if A_θ is the rotation by θ of $A = \widehat{F}_0$ (cfr. (4.31)), we must have

$$\Big| \iint_{\mathcal{Q}} A_\theta(u_1, u_2) \widehat{P}_1(u_1) \widehat{P}_2(u_2)\, du_1 du_2 \Big|$$

decreasing for θ going from 0 to $\pm\pi/2$ both for $\mathcal{Q} = \mathcal{Q}^{++}$ and $\mathcal{Q} = \mathcal{Q}^{+-}$.

We have up to now considered features aligned along the two axes given by the canonical basis vectors $\mathbf{e}_1$ and $\mathbf{e}_2$. We suppose now that they are aligned along two directions which are not necessarily perpendicular. We suppose that these two directions are given by two vectors $\mathbf{v}_1$ and $\mathbf{v}_2$ forming an angle $\theta > 0$, and such that $|\mathbf{v}_1| \cdot |\mathbf{v}_2| \cdot \sin\theta = 1$. Let λ be the linear transformation of $\mathbb{R}^2$ mapping $\mathbf{e}_1$ and $\mathbf{e}_2$ onto $\mathbf{v}_1$ and $\mathbf{v}_2$ respectively; its determinant $\det(\lambda)$ verifies

$$\det(\lambda) = |\mathbf{v}_1| \cdot |\mathbf{v}_2| \cdot \sin\theta = 1.$$

Write λ^{-1}, λ^T, and λ^{-T} for respectively the inverse, the transpose, and the inverse transpose of λ; then λ^{-T} maps $\mathbf{e}_1$ and $\mathbf{e}_2$ onto two vectors $\mathbf{w}_1$ and $\mathbf{w}_2$ respectively, which are uniquely defined by the relations

$$\mathbf{v}_i \cdot \mathbf{w}_j = \delta_{ij} = \begin{cases} 1 & \text{if } i = j, \\ 0 & \text{if } i \neq j, \end{cases} \qquad (i, j = 1, 2).$$

The basis $\{\mathbf{w}_1, \mathbf{w}_2\}$ is said to be *conjugate* to the basis $\{\mathbf{v}_1, \mathbf{v}_2\}$. In fact $\mathbf{w}_1$ and $\mathbf{w}_2$ are obtained repectively from $\mathbf{v}_2$ rotated by $-\pi/2$ and $\mathbf{v}_1$ rotated by $+\pi/2$ (see Figure 14 (a)).

Suppose that the image I has a bidirectional feature oriented along directions $\mathbf{v}_1$ and $\mathbf{v}_2$; then I is obtained by applying λ to an image J having a bidirectional feature oriented along $\mathbf{e}_1$ and $\mathbf{e}_2$, in other words

$$I(\lambda(\mathbf{x})) = J(\mathbf{x}) \qquad \text{or equivalently} \qquad I(\mathbf{x}) = J(\lambda^{-1}(\mathbf{x})) \tag{5.5}$$

for all $\mathbf{x} \in \mathbb{R}^2$. Similarly, applying λ to a filter F_i oriented along $\mathbf{e}_1$ and $\mathbf{e}_2$ gives a filter G_i defined by

$$G_i(\lambda(\mathbf{x})) = F_i(\mathbf{x}) \qquad \text{or equivalently} \qquad G_i(\mathbf{x}) = F_i(\lambda^{-1}(\mathbf{x})) \tag{5.6}$$

for all $\mathbf{x} \in \mathbb{R}^2$, oriented along directions $\mathbf{v}_1$ and $\mathbf{v}_2$. Since convolution commutes with linear transforms of $\mathbb{R}^2$ having determinant ± 1, applying λ to $J * F_i$ gives $I * G_i$. Thus for any combination of the filters F_i extracting the features in J, the same combination of the filters G_i will detect corresponding features in I. Since we used the energy function (5.1), we will consider here the energy function

$$E = (I * G_0)^2 + (I * G_1)^2 + (I * G_2)^2 + (I * G_3)^2,$$

where each G_i is given by (5.6). In the Fourier domain (5.6) gives

$$\widehat{G}_i(\lambda^{-T}(\mathbf{u})) = \widehat{F}_i(\mathbf{u}) \qquad \text{or equivalently} \qquad \widehat{G}_i(\mathbf{u}) = \widehat{F}_i(\lambda^T(\mathbf{u})) \tag{5.7}$$

for every $\mathbf{u} \in \mathbb{R}^2$. This means that $\widehat{G}_i$ results from applying λ^{-T} to $\widehat{F}_i$. Thus (G_0, G_1) and (G_2, G_3) are $\mathbf{w}_1$-quadrature pairs, while (G_0, G_2) and (G_1, G_3) are $\mathbf{w}_2$-quadrature pairs. The signs of the Fourier transforms of G_0, G_1, G_2, and G_3 in the four quadrants determined by $\mathbf{w}_1$ and $\mathbf{w}_2$ is shown in Figure 14 (b).

There are no experimental data on the efficiency of the approach introduced here for bidirectional features. Other models using bidirectional filters (also called "end-stopped" detectors) have been considered [27,28,29,30,31,32]. The problem of orientation selection is much more complicated here than in the unidirectional case, since we have to determine two directions instead of one, and we do not know how to generalize Requirement 7 and Proposition 4.13. Note also that there exist other types of features involving more than two directions, for example *Y-junctions* (see Figure 15); such a feature involves 3 directions. Moreover, strongly curved edges can be considered as multidirectional features, and it has been suggested in [27,28] that "end-stopped" bidirectional filters can be used to calculate curvature.

It might also be possible to detect bidirectional features with unidirectional filters, as in the standard phase congruence model of Section 4. Here we must abandon Requirement 4 (namely, that the filters are even-symmetric w.r.t. the tangential direction), in order to have a non-zero energy function on corners. Let us mention for example the use in [31] of "one-sided" filters where the extent of the filter in the tangential direction is truncated on one side. Here, taking into account Requirement 1, instead of truncating the filters C and S on one side along the tangential direction, we might modulate them by a cisoid function in that direction, in other words shift the Fourier spectrum of C towards one side, so that the support of the spectrum of $F = C + i\,S$ would be centered in one half-plane, say $x_t < 0$.

5.3. *Some applications, and conclusion*

The phase congruence model has been constructed to be used for the detection and localization of edges. Here the precision of the localization of edge points demands that the filters C and S have a narrow support in the spatial domain.

On the other hand this model can also be used for other tasks than feature localization, but in order to make measurements on whole regions, and this time with C and S having a narrow support in the Fourier domain. Let us mention the model proposed by [65] for the measurement of stereo disparity. We assume that $\widehat{C}$ and $\widehat{S}$ have their energy concentrated in two narrow strips around the lines $|u_1| = d$ for some distance $d > 0$, where u_1 represents the horizontal component of vector $\mathbf{u}$. Thus the spectrum of $F = C + i\,S$ is concentrated around the line $u_1 = d$. We have two images I_1 and I_2. Suppose first that there is a uniform horizontal disparity h between them, in other words $I_1(x_1, x_2) = I_2(x_1 + h, x_2)$ for all points (x_1, x_2); then we have $\widehat{I_1}(u_1, u_2) = \widehat{I_2}(u_1, u_2) \cdot \exp(2\pi i\, u_1 h)$ for all frequency vectors (u_1, u_2). Hence we get $\widehat{I_1}(u_1, u_2)\widehat{F}(u_1, u_2)$ and $\widehat{I_2}(u_1, u_2)\widehat{F}(u_1, u_2)$ both very weak for u_1 far from d, while for u_1 in the vicinity of d we will have $\widehat{I_1}(u_1, u_2)\widehat{F}(u_1, u_2) \approx \widehat{I_2}(u_1, u_2)\widehat{F}(u_1, u_2)\exp(2\pi i\, dh)$; therefore $\widehat{I_1}\widehat{F} \approx \widehat{I_2}\widehat{F}\exp(2\pi i\, dh)$, in other words $I_1 * F$ will be "close" to $\exp(2\pi i\, dh)(I_2 * F)$. Removing the assumption of uniform disparity, the local disparity between I_1 and I_2 at a point (p_1, p_2) will be the value h such that $I_1(x_1, x_2)$ is "close" to $I_2(x_1 + h, x_2)$ in the neighbourhood of of (p_1, p_2); it can be estimated as the argument of $(I_1 * F)(p_1, p_2)/(I_2 * F)(p_1, p_2)$.

Such an approach for computing stereo disparity can also be used for measuring region motion between two images. In [65] a three-dimensional model of disparity measurement using filters in quadrature is given for the integration of stereo and motion.

Let us now conclude. The phase congruence model for edge detection comprises several aspects:

(*i*) Edges are characterized as points of maximum Fourier phase congruence in the image. This type of definition is objective and scale-independent; in particular points where all Fourier phases coincide give an absolute maximum of the energy function at all scales. However there is no precise qualification of the notion of maximum phase congruence, since purely local maxima can be meaningless and do not respect the principle of causality in scale-space [17].

(ii) Edge detection is achieved by looking for maxima of the sum of squares of convolutions of the image with two filters C and S satisfying the requirements given in Section 4, or equivalently of the absolute value of the convolution of the image with the complex-valued filter $F = C + i\,S$. The use of two filters C and S, respectively even- and odd-symmetric, rather than a single one, is justified by the existence of several types of edges (see Figure 1), but one could envisage using more than two filters [18] if necessary. The Fourier amplitude and phase characteristics of C and S are justified from mathematical considerations, since interesting facts result from the nature of the Fourier spectrum of F (with positive values on $\mathcal{E}^+_{\mathbf{n}}$, and vanishing on $\mathcal{E}^-_{\mathbf{n}}$); in particular we could characterize orientation selectivity in this framework. Their phase characteristics are also justified from a physiological point of view [25]. However the equality of Fourier amplitudes of C and S can be challenged, since derivative pairs have some advantage over Hilbert transform pairs, for example causality in scale space [17].

Our study does not attempt to justify this model as the one corresponding to the functioning of human or animal visual detection of edges, nor does it claim any validity w.r.t. the photometric relevance of the edges that it gives. Furthermore the practical effectiveness of this model compared to other ones, in particular with methods using only one filter for each orientation, or using several filters separately, should be ascertained through experimental work. Probably this model would be most advantageously exploited in combination with radically different approaches, for example region-based segmentation (in particular watersheds).

More theoretical and practical studies are needed for the characterization and detection of multidirectional features and keypoints. We have only hinted at one possibility, without any justification of its validity.

The notion of Fourier phase congruence is not spatially localized, which contradicts the local nature of visual edges. A refined theory using concepts of localized Fourier phases, which could for example be based on wavelets, is needed.

References

1. J. Ross, R. Owens: Sketches: man-made and machine-made. *Preprint*, University of Western Australia.
2. B.K.P. Horn: Understanding image intensities. *Artificial Intelligence*, Vol. 8, pp. 201–231, 1977.
3. G.J. Klinker: *A Physical Approach to Color Image Undestanding*. A.K. Peters, Wellesley, MA.
4. D.H. Hubel, T.N. Wiesel: Functional architecture of macaque monkey visual cortex. *Proc. Royal Soc. London*, Ser. B, Vol. 198, no. 1130, pp. 1–59, Jul. 1977.
5. D. Lee: Coping with discontinuities in computer vision: their detection, classification, and measurement. *IEEE Trans. Pattern Analysis & Machine Intelligence*, Vol. PAMI-12, no. 4, pp. 321–344, Apr. 1990.
6. D. Marr, E. Hildreth: Theory of edge detection. *Proc. Royal Soc. London*, Ser. B, Vol. 207, pp. 187–217, 1980.
7. J. Canny: A computational approach to edge detection. *IEEE Trans. Pattern Analysis & Machine Intelligence*, Vol. PAMI-8, no. 6, pp. 679–698, Nov. 1986.

8. M.C. Morrone, R. Owens: Feature detection from local energy. *Pattern Recognition Letters*, Vol. 6, pp. 303–313, 1987.
9. M.C. Morrone, D.C. Burr: Feature detection in human vision: a phase-dependent energy model. *Proc. Royal Soc. London*, Ser. B, Vol. 235, pp. 221–245, 1988.
10. A.N. Tikhonov, V.Y. Arsenin: *Solutions of Ill-Posed Problems.* V.H. Winston & Sons, Washington (1977).
11. V. Torre, T.A. Poggio: On edge detection. *IEEE Trans. Pattern Analysis & Machine Intelligence*, Vol. PAMI-8, no. 2, pp. 147–163, Mar. 1986.
12. D. Marr: *Vision.* W.H. Freeman & Co., San Francisco, CA (1982).
13. R. Owens, S. Venkatesh, J. Ross: Edge detection is a projection. *Pattern Recognition Letters*, Vol. 9, pp. 233–244, 1989.
14. S. Venkatesh, R. Owens: On the classification of image features. *Pattern Recognition Letters*, Vol. 11, pp. 339-349, 1990.
15. C. Ronse: On idempotence and related requirements in edge detection. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, Vol. PAMI-15, no. 5, pp. 484–491, May 1993.
16. P. Kube: Properties of energy edge detectors. *Proc. IEEE Conf. Computer Vision & Pattern Recognition* (1992).
17. P. Kube, P. Perona: Scale-space properties of quadratic feature detectors. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, submitted.
18. P. Perona, J. Malik: Detecting and localizing edges composed of steps, peaks and roofs. *Third International Conference on Computer Vision* (1990), p. 52–57.
19. W.T. Freeman, E.H. Adelson: The design and use of steerable filters. *IEEE Trans. Pattern Analysis & Machine Intelligence*, Vol. PAMI-13, no. 9, pp. 891–906, Sept. 1991.
20. J.M.H. du Buff: Responses of simple cells: events, interferences, and ambiguities. *Biol. Cybernetics*, Vol. 68, pp. 321-333, 1993.
21. J.M.H. du Buff: Ramp edges, Mach bands, and the functional significance of the simple cell assembly. *Biol. Cybernetics*, Vol. 70, pp. 449-461, 1993.
22. M.C. Morrone, J. Ross, D.C. Burr, R. Owens: Mach bands are phase dependent. *Nature*, Vol. 324, pp. 250–253, 20 Nov. 1986.
23. J. Ross, M.C. Morrone, D.C. Burr: The conditions under which Mach bands are visible. *Vision Research*, Vol. 29, no. 6, pp. 699-715, 1989.
24. D.C. Burr, M.C. Morrone, D. Spinelli: Evidence for edge and bar detectors in human vision. *Vision Research*, Vol. 29, no. 4, pp. 419–431, 1989.
25. D.A. Pollen, S.F. Ronner: Phase relationships between adjacent simple cells in the visual cortex. *Science*, Vol. 212, pp. 1409–1411, 19 Jun. 1981.
26. J.A. Noble: Finding corners. *Image and Vision Computing*, Vol 6, no. 2, pp. 121–128, 1988.
27. A. Dobins, S.W. Zucker, M.S. Cynader: Endstopped neurons in the visual cortex as a substrate for calculating curvature. *Nature*, Vol. 329, pp. 438–441, 1 Oct. 1987.
28. A. Dobins, S.W. Zucker, M.S. Cynader: Endstopping and curvature. *Vision Research*, Vol. 29, no. 10, pp. 1371–1387, 1989.
29. F. Heitger, L. Rosenthaler, R. von der Heydt, E. Peterhans, O. Kubler: Simulation of neural contour mechanisms: from simple to end-stopped cells. *Vision Research*, Vol. 32, no. 5,

pp. 963–981, 1992.

30. L. Rosenthaler, F. Heitger, O. Kubler, R. von der Heydt: Detection of general edges and keypoints. *Proc. 2nd European Conf. Computer Vision* (1992), pp. 78–86.

31. P. Perona: Steerable-scalable kernels for edge detection and junction analysis. *Image & Vision Computing*, Vol. 10, no. 10, pp. 663–672, 1992.

32. B. Robbins, R. Owens: The 2D oriented energy model. *Preprint*, University of Western Australia, 1994.

33. E.H. Adelson, J.R. Bergen: Spatiotemporal energy models for the perception of motion. *J. Opt. Soc. Am.*, Ser. A, Vol. 2, no. 2, pp. 284–299, Feb. 1985.

34. A.B. Watson, A.J. Ahumada Jr.: Model of human visual-motion sensing. *J. Opt. Soc. Am.*, Ser. A, Vol. 2, no. 2, pp. 322–341, Feb. 1985.

35. J. Ling: *Analysis of Edge-Detection Algorithms for a Difficult Image Processing Application.* D.Phil. Thesis, Technical University Eindhoven (1989).

36. V.S. Nalwa, T.O. Binford: On detecting edges. *IEEE Trans. Pattern Analysis & Machine Intelligence*, Vol. PAMI-8, no. 6, pp. 699-714, Nov. 1986.

37. D. Marr: Early processing of visual information. *Phil. Trans. Royal Soc. London*, Ser. B, Vol. 275, pp. 483–519, 1976.

38. A. Herskovits, T.O. Binford: On boundary detection. *MIT AI Memo* no. 183, Jul. 1970.

39. B.K.P. Horn: *Robot Vision.* MIT Press (1986).

40. C. Ronse: A twofold model of edge detection. *PRLB Unpublished Working Document WD65*, Sep. 1990.

41. P. Brou, T.R. Sciascia, L. Linden, J.Y. Lettvin: The colors of things. *Scientific American*, Vol. 255, no. 3, pp. 80–77, Sep. 1986.

42. K. Koffka: *Principles of Gestalt Psychology.* Harcourt, Brace & Co., New York (1935).

43. M.S. Livingstone: Art, illusion, and the visual system. *Scientific American*, Vol. 258, no. 1, pp. 68–75, Jan. 1988.

44. M.S. Livingstone, D.H. Hubel: Psychophysical evidence for separate channels for the perception of form, color, movement, and depth. *Journal of Neuroscience*, Vol. 7, no. 11, pp. 3416–3468, 1987.

45. M.S. Livingstone, D.H. Hubel: Segregation of form, color, movement, and depth: anatomy, physiology, and perception. *Science*, Vol. 240, pp. 740–749, 6 May 1988.

46. E.M. Stein, G. Weiss: *Introduction to Fourier Analysis on Euclidean Spaces.* Princeton University Press, Princeton, NJ (1971).

47. W. Rudin: *Real and Complex Analysis.* McGraw-Hill (1970).

48. G.B. Folland: *Real Analysis — Modern Techniques and their Applications.* J. Wiley & Sons (1984).

49. M.E. Munroe: *Measure and Integration.* 2nd edition, Addison-Wesley, Reading, MA (1971).

50. I. Daubechies: The wavelet transform, time-frequency localization and signal analysis. *IEEE Trans. Information Theory*, Vol. 36, no. 5, pp. 961–1005, Sep. 1990.

51. C. Sadosky: *Interpolation of Operators and Singular Integrals — An Introduction to Harmonic Analysis.* Marcel Dekker Inc., New York, NY (1979).

52. E.M. Stein: *Singular Integrals and Differentiability Properties of Functions.* Princeton University Press, Princeton, NJ (1970).

53. S.K. Pichorides: On the best values of the constants in the theorems of M. Riesz, Zygmund and Kolmogorov. *Studia Mathematica*, T. 44, pp. 165–179, 1972.

54. D. Gabor: Theory of communication. *J. IEE*, Vol. 93, pp. 429–457, 1946.

55. R. Owens: Feature-free images. *Pattern Recognition Letters*, Vol. 15, pp. 35–44, 1994.

56. J.B. Garnett: *Bounded Analytic Functions.* Academic Press, New York, NY (1981).

57. M. Rosenblum, J. Rovnyak: *Topics in Hardy Classes and Univalent Functions.* Birkhäuser Verlag, Basel, Switzerland (1994).

58. S.A. Klein, D.M. Levi: Hyperacuity thresholds of 1 sec: theoretical predictions and empirical validation. *J. Opt. Soc. Am.*, Ser. A, Vol. 2, no. 7, pp. 1170–1190, Jul. 1985.

59. J.G. Daugman: Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters. *J. Opt. Soc. Am.*, Ser. A, Vol. 2, no. 7, pp. 1160–1169, Jul. 1985.

60. S. Marčelja: Mathematical description of the response of simple cortical cells. *J. Opt. Soc. Am.*, Vol. 70, pp. 1297–1300, 1980.

61. M. Abramowitz, I.A. Stegun: *Handbook of Mathematical Functions with Formulas, Graphs and Mathematicals Tables* (2nd. printing). Applied Mathematics Series, Vol. 55, National Bureau of Standards, Washington, DC (1964).

62. A. Fousse: *Applications de l'Analyse de Fourier à la Détection d'Arêtes dans les Images.* Mémoire de DEA de Mathématiques, Université Louis Pasteur, Strasbourg (1994).

63. J. Serra: *Image Analysis and Mathematical Morphology*, Academic Press, London (1982).

64. R. Hummel, D. Lowe: Computational considerations in convolution and feature-extraction in images. *From Pixels to Features*, J.C. Simon ed., Elsevier Science Publishers (North-Holland), Amsterdam (1989), pp. 91–102.

65. Ning Qian: Computing stereo disparity and motion with known binocular cell properties. *Neural Computation*, Vol. 6, pp. 390–404, 1994.

66. JJ. Clark: Authenticating edges produced by zero-crossing algorithms. *IEEE Trans. Pattern Analysis & Machine Intelligence*, Vol. PAMI-11, no. 1, pp. 43–57, 1989.

67. J.G. Daugman: Six formal properties of two-dimensional anisotropic visual filters: structural principles and frequency/orientation selectivity. *IEEE Trans. Systems, Man, & Cybernetics*, Vol. SMC-13, no. 5, pp. 882-887, 1983.

68. J.G. Daugman: Pattern and motion vision without Laplacian zero crossing. *J. Opt. Soc. Am.*, Ser. A, Vol. 5, no. 7, pp. 1142–1148, Jul. 1988.

69. E. De Micheli, B. Caprile, P. Ottonello, V. Torre: Localization and noise in edge detection. *IEEE Trans. Pattern Analysis & Machine Intelligence*, Vol. PAMI-11, no. 10, pp. 1106–1117, Oct. 1989.

70. M. Kisworo, S. Venkatesh, G. West: Modeling edges at subpixel accuracy using the local energy approach. *IEEE Trans. Pattern Analysis & Machine Intelligence*, Vol. PAMI-16, no. 4, pp. 405–410, Apr. 1994.

71. A.V. Oppenheim, J.S. Lim: The importance of phases in signals. *Proc. IEEE*, Vol. 69, no. 5, pp. 529–541, 1981.

72. C. Zetzsche, E. Barth: Fundamental limits of linear filters in the visual processing of two-dimensional signals. *Vision Research*, Vol. 30, no. 7, pp. 1111-1117, 1990.