5 Improvements in surface measurements

In the preceding sections, a technique for the measurement of surface profiles was presented, using the arrangement of the Ronchi test, whose theoretical interpretation was also worked out. As explained, a Ronchi ruling is used in order to sample the wavefront, two slope patterns being obtained by placing the ruling lines in two orthogonal positions. The patterns registered through a CCD camera have been termed ronchigrams, and a method for extracting from them the local normal to the surface has been presented. As each sampling point on the wavefront yields a measured value for the normal to the surface, therefore increasing the number of sampling points on the wavefront will lead to improvements in the surface reconstructions. Increasing the accuracy of slope measurements would also improve the measurements performed.

In the following Section three techniques for improving the measured surface reconstructions are discussed: increasing the frequency of the ruling, the use of phase-shifting techniques applied to the Ronchi test, and the use of a microstepping arrangement. Microstepping involves repetitive and precise displacements of the ruling achieved through linear encoder motors, and combining the results of a number (typically ten) of ronchigrams to perform as data coming from a single ronchigram. The number of sampled points on the wavefront will be two orders of magnitude larger. While Sections 5.1 and 5.3 deal with improvements in the sampling of the surface, Section 5.2 is oriented to enhancing the accuracy of the measured values.

5.1 Ruling spacing techniques

The key element in the measurement process is obviously the one that samples the wavefront: the Ronchi ruling. An initial approach to improvements in the measurement technique will involve studying modifications to the ruling itself.

Some solutions have been proposed over the years in order to improve the results of the Ronchi test using special ruling designs. Design of different null Ronchi tests has been one of the most active research lines in the field. The stripes on a null test are given a particular shape in order to yield straight fringes in the observation plane. Detection of errors on the surface being tested is then reduced to observing deviations from the expected straight lines. Null tests have been proposed for aspherical surfaces [Malacara 1974], parabolic surfaces [Mobsby 1974], convex optical surfaces [Szulc 1997], and even for Hartmann deflectometry setups, yielding a matrix of straight rows and columns of dots on the observation plane [Cordero 1990]. Although null tests allow better detection of surface deformations, they are suited for one particular kind of surface whose parameters are known, thus allowing the shape of the stripes on the ruling to be calculated. This makes them a very useful technique in the testing of large mirrors in astronomical telescopes, but impractical for the kind of setup we are proposing, which is expected to measure different sample geometries.

Improvements in the visibility of the shadows caused by the Ronchi test have also been obtained by varying the size of the light and dark stripes on the ruling [Murty 1973]. This type of ruling may be produced with preset ratios between light and dark stripes by using double-exposure techniques [Patorski 1980]. Notice that both null tests

and rulings with variable opening ratios are approaches which might improve the visibility of the lines in the ronchigrams, but which do not improve the sampling of the wavefront, as the same data points are collected under better visibility conditions. However, the visibility of the ronchigram lines is not a problem in our experimental setup, and does not need to be enhanced.

As the number of sampling points on the wavefront is directly related to the frequency of the ruling being placed in its path, increasing the number of lines per unit length on the ruling seems to be the simplest of the solutions in order to improve the sampling on the surface, as it might raise the number of sampled points at will. Incidentally, it must be recalled that a spatially and temporally coherent source together with a ruling with a square-wave transmittance profile is used in our setup. The incident wavefront will be diffracted to some extent, and diffractive effects will become more important as the frequency of the ruling is increased. This may lead the geometrical model on which our surface reconstruction procedure relies to be no longer valid.

In Section 3.2 it was shown how, when using low frequency rulings, the diffractive and geometrical theories of the Ronchi test were equivalent. It was supposed that the shear introduced by diffraction at the edges of the test could be ignored if its value was lower than T/8 for the first diffracted order (Eq.3.2.11). The geometrical model was shown to be valid if the shear of the first diffracted order fulfilled

$$\frac{\lambda f'}{2T} < \frac{T}{8} \Rightarrow T > \sqrt{4\lambda f'} \qquad (5.1.1)$$

Therefore, given a set of experimental conditions, such as the focal length of the objective and the wavelength of the source, there is a theoretical limit in the ruling period for the geometrical model to remain valid. In our case λ=635nm and f'=50mm, so periods longer than 0.356mm are required. Table 5.1 shows the ruling frequencies and periods available in our laboratory, the amount of shear for the first diffracted order, and the ratio of the first order diffractive shear and the grating period in each case. The final ratio shows the importance of the diffractive effects of each particular ruling - for instance, diffractive effects in the 200lpi ruling introduce a shear that is almost as large as the ruling period, showing that the information on the zero diffracted order overlaps with the first diffracted order. The geometrical theory cannot account for the interferential effects bound to appear under these experimental conditions. As the described method for developing surface profiles relies on ray propagation, it becomes invalid under these experimental conditions. The use of high frequency rulings over

300lpi yields such diffractive effects that even phase reconstructions obtained through the diffractive theory are severely degraded [Wan 1993].

Table 5.1.1: Diffractive effects using the available Ronchi rulings.

Frequency (lpi)	Period(mm)	1ST order shear (mm)	Shear/Period
50	0.508	0.0312	0.061
100	0.254	0.0625	0.246
200	0.127	0.125	0.984
300	0.0847	0.187	2.214

Fig.5.1.1 presents experimental ronchigrams for the three lower-frequency rulings of Table 5.1.1. All measurements have been carried out using an spherical surface of R=149.7mm (which will be measured as sample P175A in Section 6), placed at a distance of 192.4mm from the Ronchi ruling. The three presented ronchigrams have been acquired keeping the same position and sample surface, so the only difference between them is the frequency of the Ronchi ruling used. In all ronchigrams, the intensity signal measured along the central pixel row (pixel 256 along Y axis) is also provided in arbitrary units (a.u.), in order to show the numerical data values depicted by the figures. Experimental results for rulings with 50lpi, 100lpi and 200lpi frequencies are presented from top to bottom. The visibility of the bright lines may be seen to severely degrade as the ruling frequency is increased, showing to which extent diffractive effects may affect our measurements.

In order to reduce the diffractive effects caused by the high frequency terms present at the edges of the ruling, sinusoidal transmittance rulings may be used replacing the square wave transmittance of the standard Ronchi test technique. However, the use of this kind of rulings severely reduces the contrast between the lines from light and dark stripes of the test, increasing the uncertainty in the determination of the central pixel of each line. In addition, it has recently been shown that only quasi-sinusoidal gratings are available experimentally. Real world sinusoidal rulings have non-zero harmonic terms from their nominal frequency that cause diffractive effects to appear in the final intensity pattern. Such non-zero terms have even been quantified for a quasi-sinusoidal test obtained by exposing an interference pattern onto a photographic plate [Hibino 1997].

Fig.5.1.1: Experimental ronchigrams for the three lower-frequency rulings of Table 5.1.1. They have been acquired for a same sample at a same distance, using different ruling frequencies; (a) Ronchigram,50lpi ruling; (b) Intensity profile of the central pixel row, 50lpi ruling; (c) Ronchigram, 100lpi ruling; (d) Intensity profile of the central pixel row, 100 lpi ruling; (e) Ronchigram,200lpi ruling; (f) Intensity profile of the central pixel row, 200 lpi ruling.

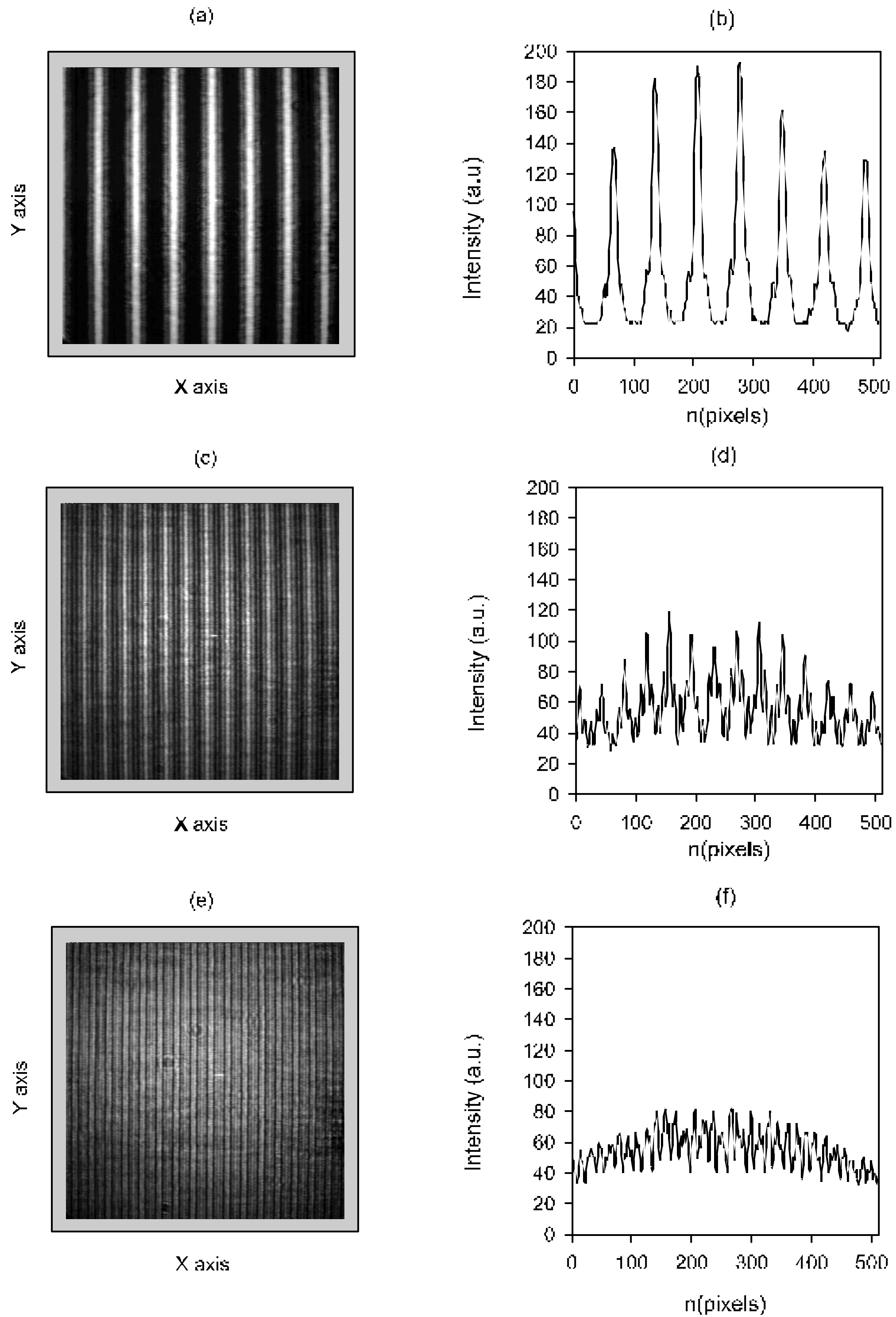
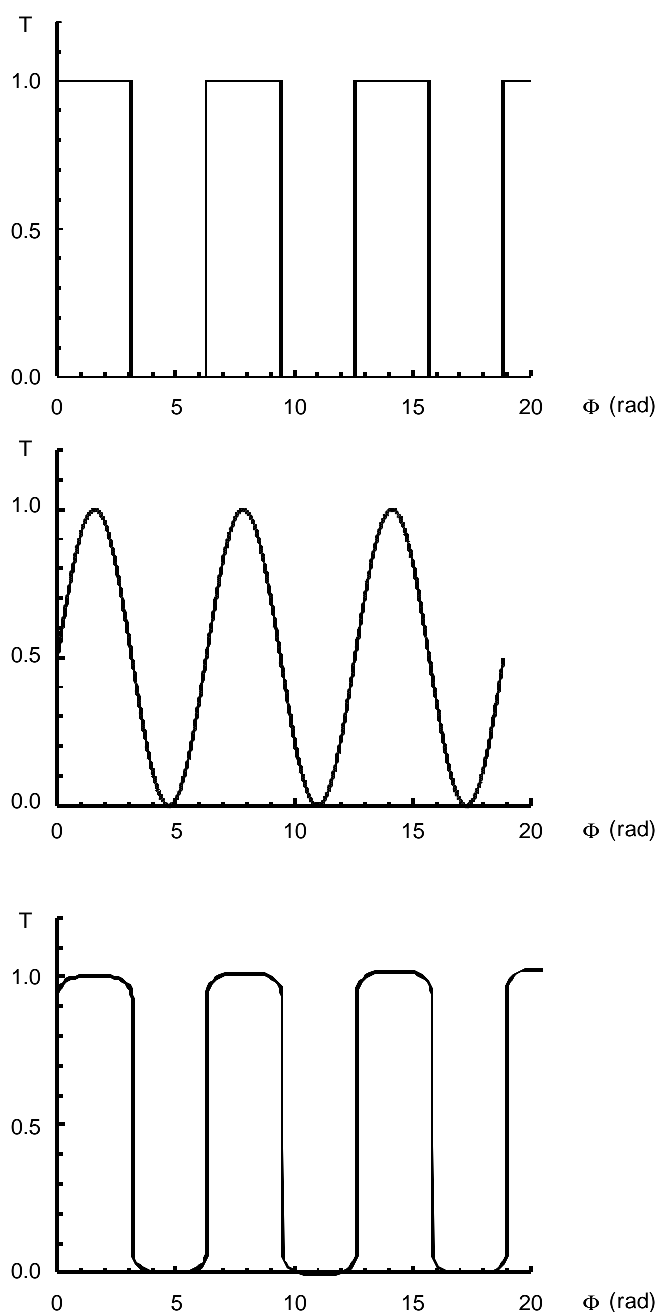


Fig. 5.1.2: (a) Square wave transmittance function; (b) Sinusoidal transmittance function; (c) Apodized square wave transmittance function.



Rulings with an intermediate transmittance function were also tested, with enough contrast to allow a precise determination of the shadow centers without lowering the dynamic range of the test, but without the straight edges present in the square wave transmittance function that cause the important diffractive effects (Fig. 5.1.2). This type of apodized square wave transmittance function was obtained using a

computer program allowing different exposure times for each column of pixels, which were then photographed using a conventional Nikon camera with a Vivitar Macro f/2.2 50mm objective, using different exposure times. Different development conditions were also tested in order to improve the performance of the rulings. Although diffractive effects were reduced to some extent, they were not removed from the shadow pattern, so the expected increase of the ruling frequency was not achieved. Furthermore, a new problem arose as it was seen to be extremely difficult to obtain straight fringe patterns from photographs of the curved PC screen. This additional problem, together with the developed tests' failure to eliminate the undesired diffractive effects, led us to abandon this approach.

5.2 Phase-shifting techniques

Under this general term a very successful system for improving measurements made using interferometric techniques has been developed. The system is now widely applied, and all top-performance interferometric commercial equipment incorporates such techniques in order to improve its accuracy up to two orders of magnitude ($\lambda/100$). Its application to interferometry has been so successful that many related fields in optical testing have tried to incorporate such strategies [Omura 1988] [Pfeifer 1995].

As phase-shifting techniques were originally applied to interferometric testing, we will first present a short description of how classical phase-shifting schemes are used in order to extract phase information from fringe patterns obtained through interferometric techniques. After presenting the original technique, the following sections will discuss in detail the aspects to be considered when applying phase-shifting techniques to ronchigrams.

5.2.1.- Phase shifting interferometry

Interferometric fringe patterns have traditionally been recorded on photographic plates, extracting the data needed in order to achieve a wavefront reconstruction from them using digitizing tablets, on which the maxima of the fringe patterns were manually marked, and an order number was assigned to each fringe in an overall count. This approach yields a measurement precision of $\lambda/2$, as this would be the distance between consecutive maxima or minima. However, solid-state detectors can extract

intermediate intensity values from the fringe pattern, as the whole sinusoidal intensity pattern can be digitally recorded. Phase-shifting strategies exploit this amount of unused information when working with fringe maxima or minima, enhancing the performance of interferometric techniques by at least two orders of magnitude.

In order to illustrate our short description of the fundamentals of phase-shifting interferometry techniques, a particular example of a simple theoretical measurement application using phase-shifting techniques will be developed in this work. One-dimensional intensity patterns and algorithms will be used in this example, as they fully reveal the main steps involved in a measurement process where phase-shifting techniques are used. A variety of alternatives to the procedure presented are workable, and will where possible be described in the course of the text. A broad review of phase-shifting interferometry techniques may be found in [Greivenkamp 1992].

A general fringe pattern from an interferometric measurement may be expressed as

$$I(x) = I_0(x) \cdot \left[1 + \gamma(x) \cdot \cos(\Phi(x))\right] \qquad (5.2.1)$$

where $I_0(x)$ is the dc intensity at each data position, $\gamma(x)$ is the modulation of the measured fringe pattern and $\Phi(x)$ is the wavefront phase to be measured. These three values may be different at each data position, but are assumed to be stable in time. Although there are indeed spatial phase-shifting techniques that may be applied to fringe patterns varying in time along each register [Creath 1996], we will restrict ourselves to temporally stable interferograms.

In order to completely describe the measured fringe pattern, three unknowns must thus be determined at each point ($\Phi(x)$, $\gamma(x)$, and $I_0(x)$); obviously, at least three data values at each pixel position are needed. These new values may be obtained by introducing a known phase difference (a known "phase shift") in the reference arm of the interferometer and recording the resulting interference patterns, because the recorded intensity will have a known intensity profile, given by Eq. 5.2.1. The most usual way of applying this phase shift is through piezoelectric transducers (PZT), which displace the reference surface of the interferometer. However, many techniques involving Zeeman lasers, tilted plates, moving diffraction gratings, or Bragg cells have been described and applied [Creath 1988].

If the phase shift is introduced in steps, and the interferogram is recorded at constant phase values, we are dealing with a phase-stepping technique. If the phase is

varied continuously while the interferogram is being recorded, we are using an integrating bucket technique. The intensity pattern for the *i* measurement of the phase-stepping technique is described as

$$I_i(x) = I_o(x) \cdot \left[1 + \gamma(x) \cdot \cos(\Phi(x) + \alpha_i)\right] \qquad (5.2.2)$$

with α_i being the phase increment introduced in the i measurement. If the integrating bucket technique is applied, the intensity pattern is

$$I_i(x) = I_o(x) \cdot \left[1 + \gamma(x) \cdot \mathrm{sinc}\left(\frac{1}{2}\Delta\right) \cdot \cos\left(\Phi(x) + \alpha_i\right)\right] \qquad (5.2.3)$$

where Δ is the interval of phase variation, along which the signal is integrated. Practical reasons, such as transient oscillations of the PZT transducer after the movement of the reference surface, tend to make this smooth variation of the phase over time more practical, despite the additional modulation present in the intensity pattern. In order to present phase-shifting techniques, stepping techniques will be assumed in the following pages, as they are the techniques that are closest to those which could be applied in our experimental setup. Furthermore, the two techniques are equivalent except for the modulation term present in the signal.

Fig. 5.2.1 presents the theoretical intensity profile that is obtained when using interferometric techniques to measure a spherical wavefront with a radius of curvature of ten waves, with a flat wavefront (obtained, for example, using a flat mirror surface) used as a reference. An arbitrary one-dimensional area has been selected for the example, with its limits placed at $x=\pm 6.26\lambda$, so the maximum height of the wavefront along Y axis will be 2.202 waves. Values of $I_o(x)=2$, $\gamma(x)=0.707$ were used in Eq. 5.2.2 for all the sampled points.

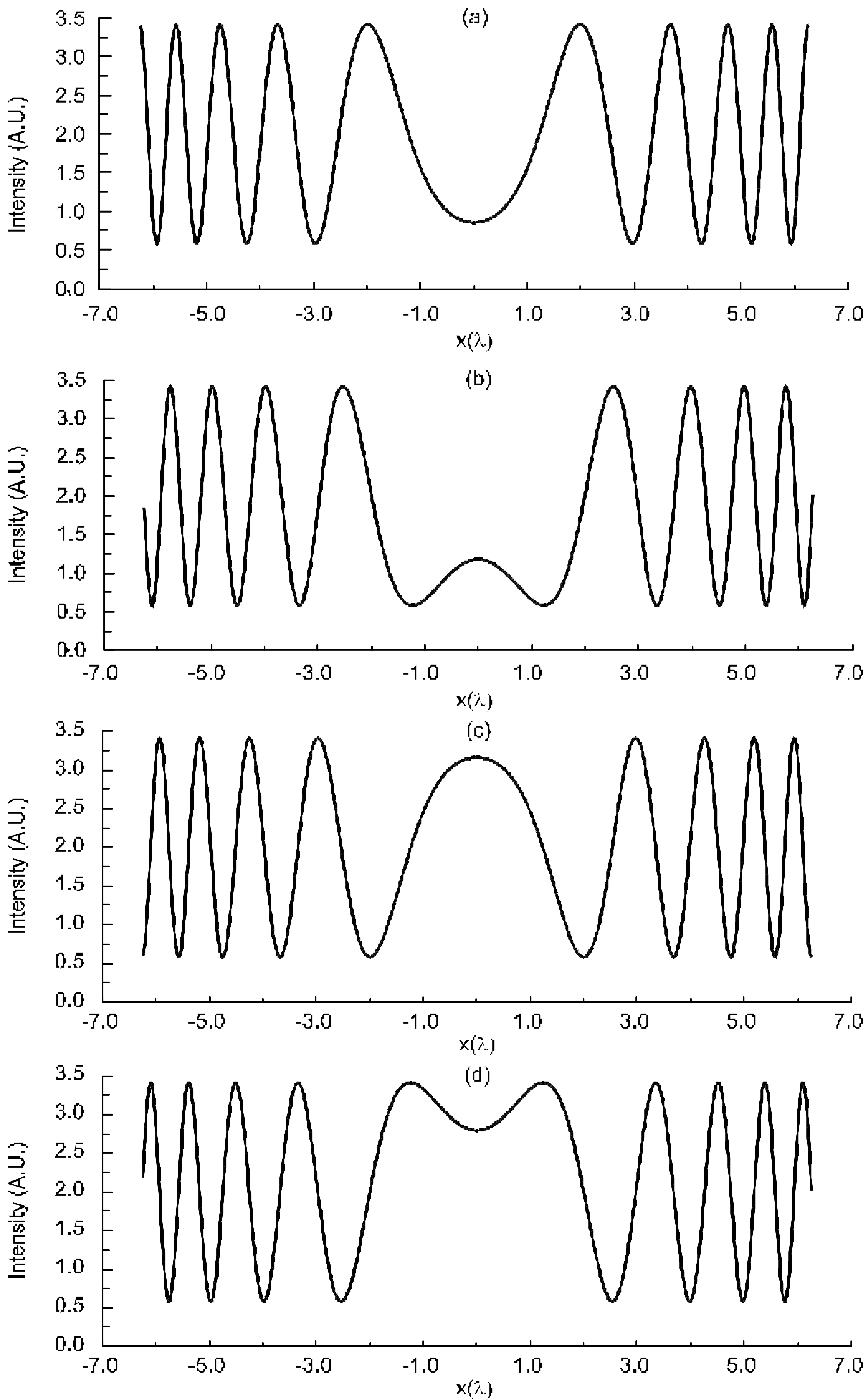
Four phase increments between consecutive recordings were assumed to have been achieved through phase-stepping displacement of the flat reference mirror in the amounts:

$$\Delta x_T = \left[1 + \frac{0.02}{\pi}\right] \cdot \frac{\lambda}{8} \qquad (5.2.4)$$

therefore, the total displacement of the flat reference mirror would be

$$\Delta x_T = \left[1 + \frac{0.02}{\pi}\right] \cdot \frac{\lambda}{2} \qquad (5.2.5)$$

Fig. 5.2.1: Theoretical signal for a spherical wavefront with a radius of curvature of 10 waves. The four calculated phase-shifted intensity patterns are presented (see text).



and the phase displacement between interferograms would be

$$\Delta\alpha_i = \frac{\pi}{2} + 0.01 \qquad (5.2.6)$$

The reason for this choice will become clear later. Minima and maxima of the signal obviously correspond to dark and bright interference fringes of variable width. Four phase-shifted intensity patterns are assumed, so a set of four different intensity values are known at each pixel. A phase reconstruction algorithm will recover the values for the wavefront phase, which will allow the local height of the wavefront to be reconstructed at each sampled point.

However, notice that only three unknowns are present in Eq.5.2.2, as α_i is determined, and four intensity patterns have been registered. Phase shifting algorithms usually work with oversampled signals to which least squares methods are applied. A number of algorithms have been proposed, from the simplest three-step algorithm (which does not need least squares methods) to complex error-compensating algorithms, insensitive to bias modulation [Surrel 1997], including general approaches to the design of algorithms tailored to compensate the most important errors in the considered setup [Surrel 1996] [Phillion 1997], or the analysis of the Fourier components of the window function, yielding algorithms with eight sampling points [DeGroot 1995] [Schmit 1996]. General self-correcting algorithms that calculate the wavefront phase together with the phase increments through least-squares fitting have also been reported [Kong 1995].

In our simple example, the phase will be recovered from the signals in Fig. 5.2.1 through the classic Carré algorithm, which requires four recorded interferograms with a constant phase increment between them, but does not apply any least-squares method, as the fourth signal is used in order to determine the phase step applied. If the recorded intensities (not writing the functionality on the position in order to make the expressions more readable) are

$$I_1 = I_o \cdot [1 + \gamma \cdot \cos(\Phi)] \qquad (5.2.7)$$

$$I_2 = I_o \cdot [1 + \gamma \cos(\Phi + \Delta\alpha)] \qquad (5.2.8)$$

$$I_3 = I_o \cdot [1 + \gamma \cos(\Phi + 2\Delta\alpha)] \qquad (5.2.9)$$

$$I_4 = I_o \cdot [1 + \gamma \cos(\Phi + 3\Delta\alpha)] \qquad (5.2.10)$$

the expressions for the determination of the phase, the modulation and the phase increment are

$$\tan\Phi = \frac{\sqrt{[3(I_2 - I_3) - (I_1 - I_4)] \cdot [(I_2 - I_3) + (I_1 - I_4)]}}{I_2 + I_3 - I_1 - I_4} \quad (5.2.11)$$

$$\gamma = \frac{1}{2I_o\sqrt{2}}\sqrt{[(I_2 - I_3) + (I_1 - I_4)]^2 + [(I_2 + I_3) - (I_1 + I_4)]^2} \quad (5.2.12)$$

$$\tan\frac{\alpha}{2} = \sqrt{\frac{3(I_2 - I_3) - (I_1 - I_4)}{I_2 - I_3 + I_1 - I_4}} \quad (5.2.13)$$

Additional care must be taken in the phase reconstruction process in order to remove the ambiguities introduced by the intrinsic nondetermination present in the arctangent function, which only reports phase values from $-\frac{\pi}{2}$ to $\frac{\pi}{2}$ rad. In order to achieve a full phase reconstruction from 0 to 2π radians, the signs for the sine and cosine must be calculated, so as to establish the appropriate quadrant for the computed phase value. In the Carré technique [Creath 1988]

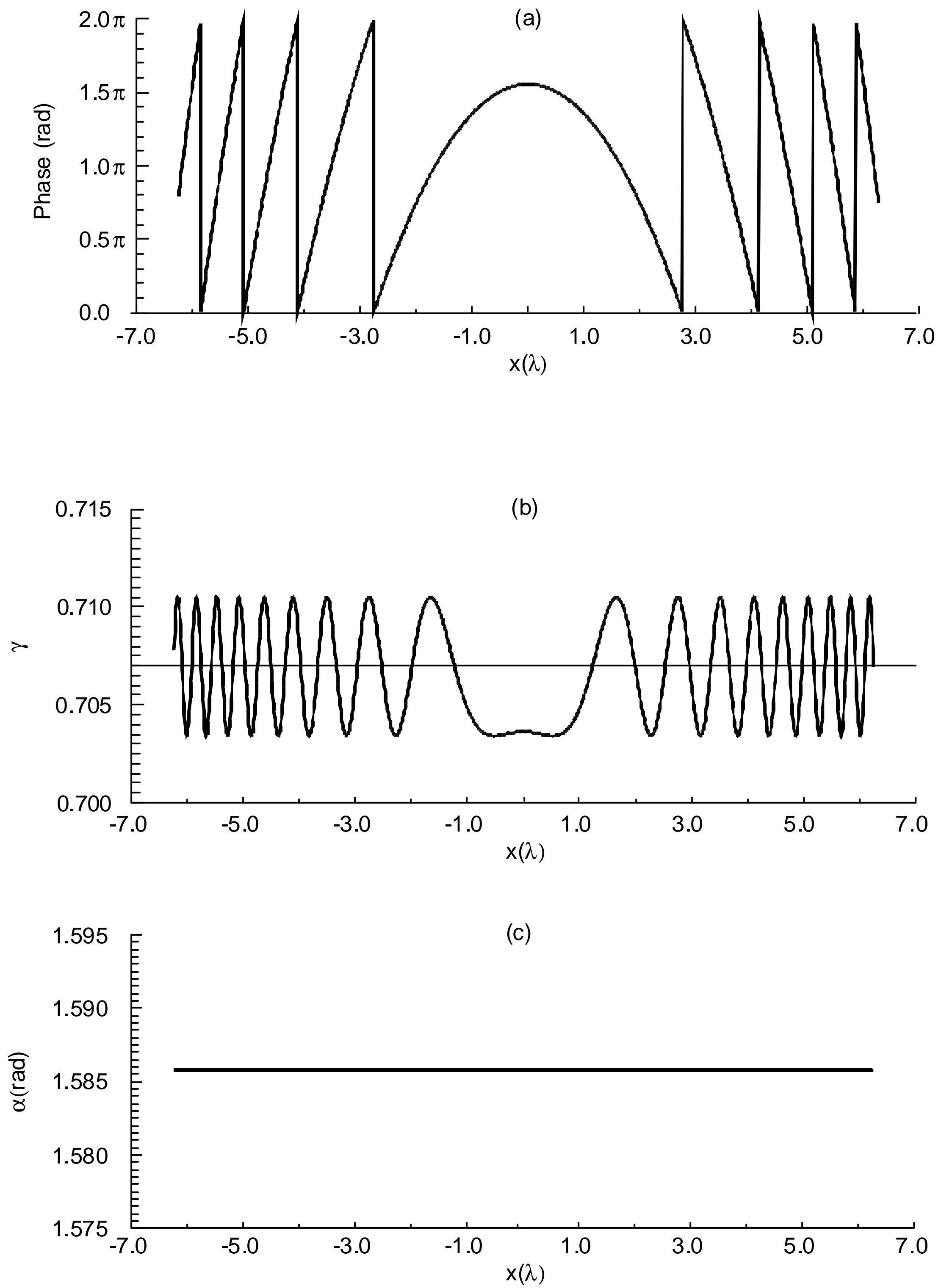
$$[I_2 - I_3] \propto \sin\Phi \quad (5.2.14)$$

$$[I_2 + I_3 - I_1 - I_4] \propto \cos\Phi \quad (5.2.15)$$

Fig. 5.2.2a presents the modulo 2π phase reconstruction, which contains all the information on the wavefront being measured. The calculated modulation (Fig. 5.2.2b) and phase increment (Fig. 5.2.2c) graphs are constant in this theoretical case, resulting in its exact value in the calculation of the phase increment ($\Delta\alpha = \frac{\pi}{2} + 0.01 \approx 1.581$ rad).

Calculated modulation, however, may be seen to oscillate closely around its theoretical value (γ=0.707), as the modulation of the signal at different points is not equivalent in the present sampling conditions. The modulation value used in the calculations is the straight line in the figure. This effect is caused by round-off errors in the calculation, so the measured intensities do not exactly follow Eq.5.2.7 to Eq.5.2.10. In real-world experiments the effect of having a finite number of intensity levels in the recordings introduces this effect more significantly, as intensity patterns are usually digitized using 8-bit or 10-bit frame-grabbers, which yield 256 and 1024 intensity levels respectively. In our example, 1000 intensity levels were used, which would nearly equal the 1024 intensity levels of a signal recorded using a 10-bit frame-grabber. The effect is fittingly termed quantization noise, and is one of the many sources of errors that limit the performance of phase-shifting techniques.

Fig.5.2.2.- Values obtained from the signals in Fig. 5.2.1: (a) Reconstructed phase modulo 2π; (b) Calculated modulation; (c) Calculated phase increment.

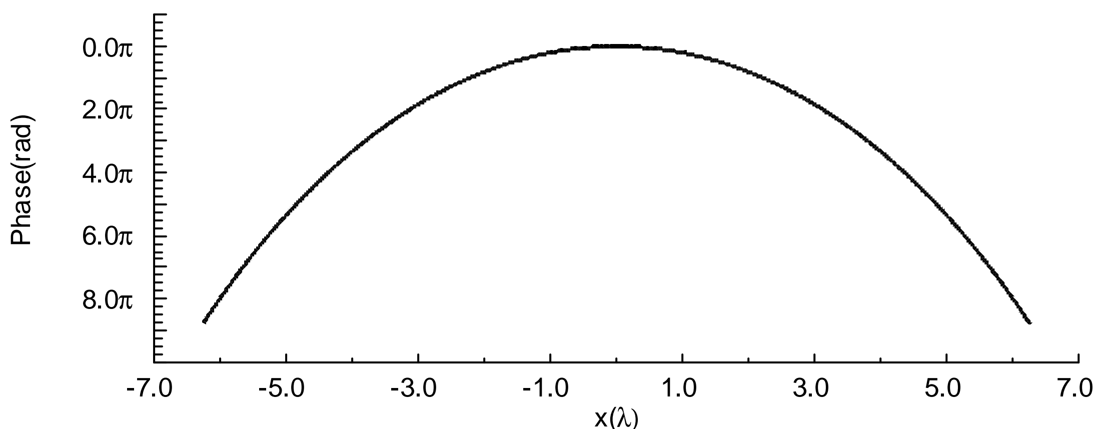


The procedure explained up to this point only accounts for phase variations modulo 2π, which cause the discontinuities in the measured phase, as the real phase will have the values:

$$\Phi(x) = \{\Phi(x)\}_{MOD.2\pi} + 2n\pi \qquad (5.2.16)$$

with an integer n that cannot be determined with the techniques discussed so far. Phase unwrapping techniques are used to determine the appropriate value for n at each sampled point and yield an absolute phase value. In a theoretical example such as ours, the procedure is as simple as looking for phase discontinuities above a certain threshold value. The final reconstructed wavefront is shown in Fig. 5.2.3. However, in real-world applications things are not quite as simple and complex strategies have been reported in order to suitably unwrap the phase under noisy environments [Huntley 1989] [Quiroga 1994]. When large amounts of aberration are present, aliasing effects will appear if fewer than two samples per period are measured; this limit frequency is the well-known Nyquist frequency. In some special cases, such as aspheric surface testing, some techniques have been presented that allow wavefront reconstructions with fringes well over the Nyquist frequency of the detector [Greivenkamp 1987].

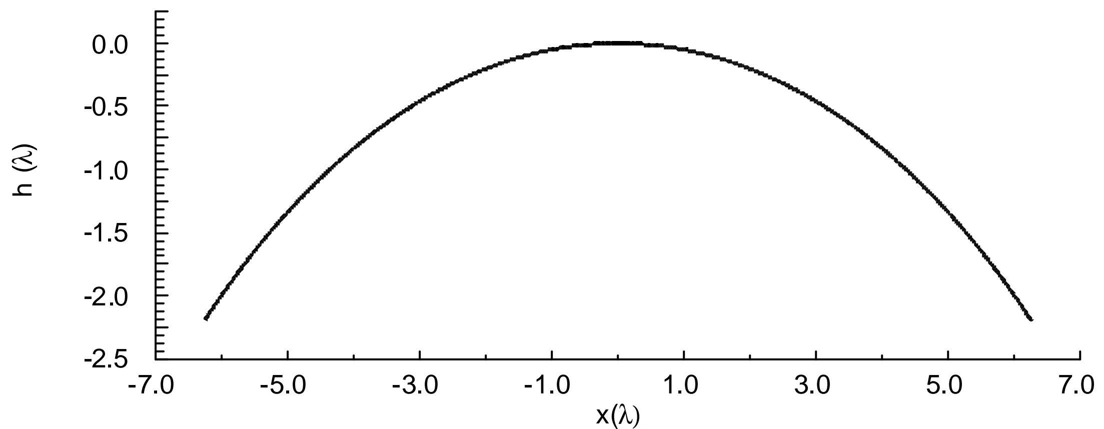
Fig.5.2.3.- Unwrapped phase of the signals in Fig. 5.2.1.



The final step will involve the transformation of the phase measurements into height measurements, using the relationships established by Eqs.5.2.4 and 5.2.6. The results are presented in Fig. 5.2.4, where only a linear transformation of the Y axis has been performed from Fig. 5.2.3. An arbitrary origin was fixed at the maximum height of Fig. 5.2.4. It may also be seen how the theoretical maximum height (2.202 waves) has been recovered from the intensity pattern data. In a real-world application, this final

shape would be fitted to a set of polynomials. Zernicke polynomials are commonly used, because of their orthogonality, in order to study what aberrations are present in the measured wavefront, and to what extent [Malacara 1992b].

Fig. 5.2.4.- Wavefront reconstruction from the signals in Fig. 5.2.1.



Real-world applications, however, present other problems apart from that of phase unwrapping. Systematic and random errors may affect measurement accuracy: the most important of these [Joenathan 1994] are the miscalibration of the phase shifter [Ai 1987] and the nonlinearities of the detector [Kinnstaetter 1988]. The effects of these two errors, and of other sources of error, such as random fluctuations of the measured phase, mechanical vibrations, or light source instabilities on the precision of the technique have been studied using linear approximations for phase-stepping and integrating-bucket techniques. This approach permits studies on how the ten most widely used classical algorithms are affected by these error sources [Van Wingerden 1991].

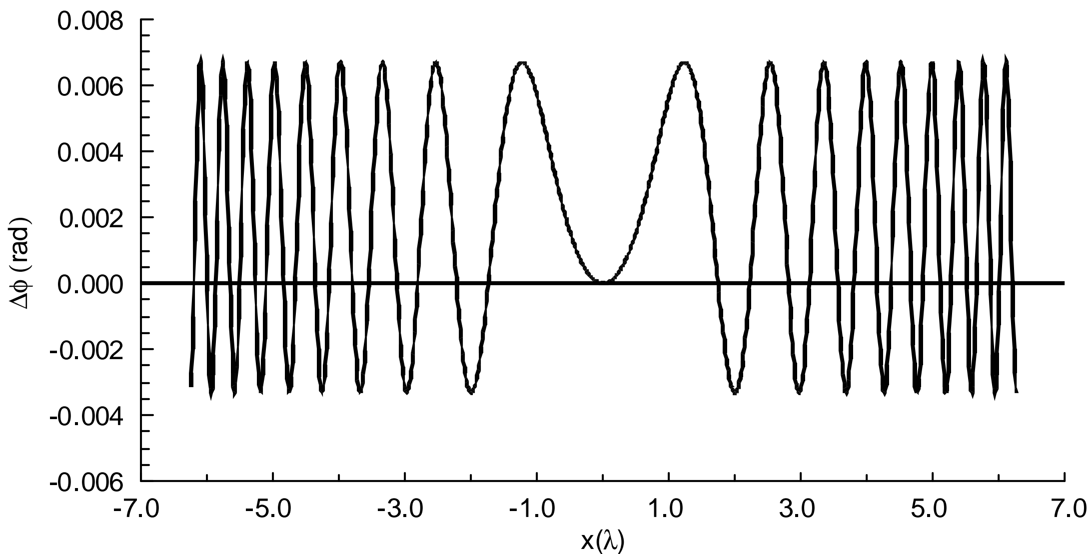
In order to give an elementary idea of the extent to which measurements may be degraded by experimental errors, a simple test was carried out, using the four-step algorithm [Creath 1988] to recover the wavefront phase from the intensity patterns in Fig. 5.2.1. The phase is recovered from the expression

$$\tan\Phi = \frac{I_4 - I_2}{I_1 - I_3} \qquad (5.2.17)$$

The four-step algorithm assumes that phase differences between phase-shifted signals are exactly $\pi/2$ rad. As the phase increment applied to the signals in Fig. 5.2.1. was $\Delta\alpha = \pi/2 + 0.01$ rad (Eq. 5.2.6), the situation is equivalent to applying the four-step

algorithm to a signal with a phase-shifter constant error of 0.01rad, accumulated at each step. Fig. 5.2.5 shows the differences between the recovered phases using the Carré and the four-step algorithms, once unwrapped and given the same origin. As the Carré algorithm requires constant phase increments but does not need a given phase value to be established, its values are assumed to be essentially correct, so the figure may be considered a plot of the error caused by the described phase-shift miscalibration. An asymmetrical error in the phase, dependent on the position, is introduced along the whole wavefront pattern; a slight modulation of the error function may also be appreciated. The example is also a simple demonstration of the influence of using one type of algorithm or another in the final measured value of the wavefront phase.

Fig. 5.2.5: Errors in phase recovery using the four-step algorithm with a miscalibrated phase shifter with the signals in Fig. 5.2.1.



5.2.2.- Phase-shifting techniques applied to the Ronchi test

The success of phase-shifting techniques led to attempts to take advantage of their benefits in optical testing techniques close to interferometry. As the Ronchi test has been understood as a kind of lateral shearing interferometer, phase measurement setups using the Ronchi test were quite a natural approach. In the Ronchi test, phase-shifting can easily be achieved through lateral displacements of the Ronchi ruling that are much less precise than the ones needed in interferometry; as the phase increments are in the micrometer range, no PZT transducers are needed. Furthermore, the Ronchi

test technique measures surface areas and height variations larger than the ones normally found in interferometry, although it lacks the accuracy of interferometric techniques. Phase-shifting techniques are particularly well-suited to removing this main drawback to the Ronchi test, as they have enhanced the accuracy of interferometric techniques by two orders of magnitude. Section 5.2.2.1 will describe and discuss some of the reported applications of phase-shifting techniques applied to the Ronchi test. Section 5.2.2.2 will deal with the problem arising from the fact that the intensity profiles of ronchigrams cannot be described using a simple sinusoidal expression like that of Eq.5.2.2.

5.2.2.1.- Review of reported applications

In the following section the reported applications of phase shifting techniques to the Ronchi test of which we are aware are presented in chronological order. Our aim is to present the ways in which other researchers have confronted the differences in the Ronchi test with interferometric techniques in order to implement phase-shifting systems.

The first report of a phase measurement using the Ronchi test which we know of is from [Yatagai 1984]. In this paper a synchronous phase-stepping measurement technique is used for obtaining quantitative measurements of the aberrations present in an aspherical surface. This technique obtains the intensity measurements in the same way as the techniques presented in Section 5.2.1, but the phase of the signal is retrieved using an algorithm coming from Fourier series analysis rather than from the least-squares techniques used in common phase shifting interferometry. The wavefront phase is recovered from

$$\Phi(x,y) = \mathrm{atan}\left(-\frac{\sum_{i=1}^{N} I_i(x,y)\,\mathrm{sen}(\delta_i)}{\sum_{i=1}^{N} I_i(x,y)\cos(\delta_i)} \right) = \mathrm{atan}\left(\frac{S(x,y)}{C(x,y)}\right) \qquad (5.2.18)$$

being $\delta_i = \frac{2\pi}{N} i$ the ith of the N applied phase steps [Greivenkamp 1992]. Yatagai uses a ruling with 0.2mm period in order to sample a wavefront coming from the sample lens under test. This lens receives an impinging wavefront consisting of a collimated beam obtained using a 40µm wide puntual source and a 40mm focal length aspherical lens.