Journal of Optoelectronics and Advanced Materials Vol. 6, No. 3, September 2004, p. 841 - 855

STATISTICAL PROCESSING OF STRAIGHT EQUISPACED FRINGE PATTERNS

V. Nascov*, A. Dobroiu, D. Apostol, V. Damian

National Institute for Laser, Plasma and Radiation Physics, PO Box MG-36, R-77125 Bucharest, Romania

The paper is dealing with statistical processing of digitally recorded straight equispaced fringe patterns. It is determined the highest degree of accuracy that can be achieved in estimating fringe parameters by statistical processing under given statistical fluctuation conditions affecting the recorded image.

(Received August 21, 2003; accepted June 22, 2004)

Keywords: Fringes, Fringe pattern, Digital image, Statistics, Least squares fitting

1. Introduction

Straight equispaced fringe patterns occur in many interference cases. These are the simplest fringe patterns, but have a great importance because any other fringe patterns of any complexity may be decomposed in terms of these simple patterns. Furthermore, processing of these fringe patterns is very important in optical metrology [1,2,3].

There are many methods for fringe processing. The most modern is considered the phase shifting interferometry, a method that has to process at least three phase shifted fringe patterns to retrieve the phase map of an object wavefront relative to a reference wavefront. The statistical properties of the errors of this method are well described in Refs. 4-8.

Since the fringe patterns that we concern ourselves with have an *a priori* known phase map (linear ramp), we don't have to determine the phase map, but only it's gradient and phase offset, which must be constant parameters. For this purpose we use least squares fitting (LSF) to compute the basic parameters of the fringe patterns as statistical quantities [9,10]. LSF, a method widely used for data processing, is well described in the appendix using a matrix language well suited for computer programming. We also have implemented this method in computer programs that we used to process fringe patterns and to study the related errors. We do not give details on these computer programs, since the main purpose of this paper is to evaluate the statistical errors for the LSF computed fringe parameters.

Because these fringe patterns are simple (periodic), the Fourier transform method [11] is optimal to process them and may be faster than LSF, which is a very general method. However, we consider that LSF is very important as a classical method and the main point of the present work resides in the fact that we could obtain analytical expressions for the statistical errors of LSF. We can say that the errors found on LSF may serve as reference errors to compare the errors of any other methods used to process fringe patterns. Moreover we think that LSF is one of the most accurate statistical methods.

2. Fringe pattern processing

2.1. Periodic harmonic fringes

The simplest fringe pattern has a harmonic distribution of intensity along a direction x or over a plane (x, y):

* Corresponding author: nv@email.ro

1D fringe pattern: $$\mathrm{I}(x) = I_0 + I_1 \cdot \cos(2\pi\nu x + \varphi) \tag{1a}$$

2D fringe pattern: $$\mathrm{I}(x, y) = I_0 + I_1 \cdot \cos\left(2\pi(\nu_x \cdot x + \nu_y \cdot y) + \varphi\right) \tag{1b}$$

I_0 is the background intensity, I_1 is the intensity of the fundamental harmonic, ν is the spatial frequency of the fringes having two components for 2D fringe pattern, and φ is a phase parameter that specifies the global positioning of the fringe pattern. Such fringe patterns occur by the interference of two uniform plane wavefronts.

Fringe pattern acquisition by a CCD leads to a sequence of sampled and digitized values of the intensity distribution:

1D: $$I_{\mathrm{k}} = I(k \cdot \delta x) = I_0 + I_1 \cdot \cos\left(\frac{2\pi}{N} N_f k + \varphi\right); \qquad N_f = N \cdot \nu \cdot \delta x, \quad k = 0,1,\ldots,N-1 \tag{2a}$$

2D: $$I_{\mathrm{mn}} = I(m \cdot \delta x, n \cdot \delta y) = I_0 + I_1 \cdot \cos\left(2\pi \cdot \left(m \cdot \frac{N_{fx}}{N_x} + n \cdot \frac{N_{fy}}{N_y}\right) + \varphi\right); \tag{2b}$$
$$N_{fx} = N_x \cdot \nu_x \cdot \delta x, \quad N_{fy} = N_y \cdot \nu_y \cdot \delta y, \quad m = 0,1,\ldots,N_x - 1, \quad n = 0,1,\ldots,N_y - 1$$

where δx and δy denote the sampling steps, which are the spacing between two adjacent CCD pixels, on two orthogonal directions. Instead of the spatial frequency we prefer further using the adimensional parameter N_f that is the number of fringes in the fringe pattern. It has also two components in the case of 2D fringe pattern.

We should mention that actually the sampling performed on the fringe pattern by the CCD is not just the simple sampling of the functions (1), because the CCD pixels have finite size and they output the average light intensity over their active area. We describe this effect for the 1D case assuming a linear data acquisition system. According to the theory of linear systems, the output $\mathrm{J}(x)$ can be written as a convolution of the input $\mathrm{I}(x)$ with the impulse response function $\mathrm{g}(x)$ of the acquisition system:

$$\mathrm{J}(x) = (\mathrm{I} * \mathrm{g})(x) = \int_{-\infty}^{\infty} \mathrm{I}(x - x') * \mathrm{g}(x')dx' \tag{3}$$

Let's consider a rectangular impulse response function:

$$\mathrm{g}(x) = \begin{cases} 1/\delta x, & |x| \le \varepsilon \cdot \delta x/2 \\ 0, & |x| > \varepsilon \cdot \delta x/2 \end{cases} \tag{4}$$

that signifies light averaging over the active area of a CCD pixel, whose size equals a fraction ε of the spacing δx between two adjacent pixels. With this impulse response function we obtain the following sequence of N samples:

$$\mathrm{J}_{\mathrm{k}} = \mathrm{J}(k \cdot \delta x) = J_0 + J_1 \cdot \cos\left(\frac{2\pi}{N} N_f k + \varphi\right); \qquad k = 0,1,\ldots,N-1$$
$$J_0 = \varepsilon \cdot I_0, \quad J_1 = \varepsilon \cdot \mathrm{sinc}(\varepsilon\pi N_f / N) \cdot I_1, \quad N_f = N \cdot \nu \cdot \delta x \tag{5}$$

Therefore the contrast in the acquired data of the fringe pattern is reduced, but very slightly if the CCD resolution N is large enough. Of course, beside light averaging due to the finite size of pixels, there are many other effects, such as light diffraction. All effects superpose and give a more complicated impulse response function, but for any linear acquisition system, the acquired data remains harmonic like the fringe pattern and the parameters (N_f, φ) remains unaffected, in spite of the actual impulse response function. Further we shall use the formulae (2) to describe the fringe patterns, taking into account that the fringe contrast never reaches 1 (always $I_1 < I_0$).

Let's show how LSF is applied to determine the set of 4 parameters $\mathbf{A}=(I_0, I_1, N_f, \varphi)$ in (2a), respectively 5 parameters $\mathbf{A}=(I_0, I_1, N_{fx}, N_{fy}, \varphi)$ in (2b). These parameters are statistical quantities and the LSF computer program for fringe processing gives their statistical means and variances. Their variances are due to the fluctuations of the acquired data, that we suppose to have Gaussian distribution with 0 mean and σ_I^2 variance. In fact only making this assumption is the LSF correctly applicable.

The LSF applicable on (2) to determine **A** is nonlinear (see appendix), because the relations (2) cannot be written in the form of a linear combination of independent functions having the parameters **A** as coefficients, as the equation (A1) shows. Nonlinear LSF needs a set of approximate starting values for **A** to iteratively compute high precision final results. It's easy to get the starting values using course precision methods, for example applying Fast Fourier Transform to a subset of the sequence (2). A 10% precision has been proved to be enough to successfully start the iterations. Usually 5-10 iterations lead to the final result.

The formula (A7) allows us to calculate the covariance matrix of the parameters **A**, that is, their statistical errors. This formula is applicable as well for nonlinear LSF, since it expresses the first order approximation of the error propagation law. For the 1D fringe pattern we calculate firstly a 4-component vector of functions that are the partial derivatives of the function I(x) with respect to each of the **A** parameters:

$$\mathbf{D}(x,\mathbf{A})=\left(\frac{\partial}{\partial I_0}I(x,\mathbf{A}),\frac{\partial}{\partial I_1}I(x,\mathbf{A}),\frac{\partial}{\partial N_f}I(x,\mathbf{A}),\frac{\partial}{\partial \varphi}I(x,\mathbf{A})\right) \tag{6}$$

Then we have to compute the 4.4 matrix of the statistical weights $\mathbf{P_A}$:

$$\mathbf{P}_{\mathbf{A},ij}=\frac{1}{\sigma_I^2}\cdot\sum_{k=0}^{N-1}\mathbf{D}_i(x_k,\mathbf{A})\cdot\mathbf{D}_j(x_k,\mathbf{A})\approx\frac{1}{\sigma_I^2}\cdot\int_0^N\mathbf{D}_i(k\cdot\delta x,\mathbf{A})\cdot\mathbf{D}_j(k\cdot\delta x,\mathbf{A})\cdot dk,\quad i,j=1,2,3,4 \tag{7}$$

Assuming a very large number of samples ($N\to\infty$), we have approximated the 16 sums above with integrals, which we could calculate analytically. However some of them are still complicated, but using the limit of very large fringe number $N_f\to\infty$ they simplify very much. Next one needs only inversing the $\mathbf{P_A}$ matrix in order to get the covariance matrix σ_{A}^2. The covariance matrix for the parameters of 2D fringe patterns are calculated similarly, except that the sums and integrals become double according to the (x, y) coordinates, and the amount of computation is much greater. As described above we have found analytical expressions for the statistical errors of the parameters **A** using an asymptotic approximation valid for large number of fringes N_f and large number of samples N:

1D:
$$\mathbf{A}=\begin{bmatrix}I_0\\I_1\\N_f\\\varphi\end{bmatrix},\qquad \sigma_{\mathbf{A}}^2=\sigma_I^2\cdot\begin{bmatrix}1/N & 0 & 0 & 0\\ 0 & 2/N & 0 & 0\\ 0 & 0 & 6/\pi^2I_1^2N & -6/\pi I_1^2N\\ 0 & 0 & -6/\pi I_1^2N & 8/I_1^2N\end{bmatrix} \tag{8a}$$

2D:
$$\mathbf{A}=\begin{bmatrix}I_0\\I_1\\N_{fx}\\N_{fy}\\\varphi\end{bmatrix},\qquad \sigma_{\mathbf{A}}^2=\sigma_I^2\cdot\begin{bmatrix}1/N & 0 & 0 & 0 & 0\\ 0 & 2/N & 0 & 0 & 0\\ 0 & 0 & 6/\pi^2I_1^2N & 0 & -6/\pi I_1^2N\\ 0 & 0 & 0 & 6/\pi^2I_1^2N & -6/\pi I_1^2N\\ 0 & 0 & -6/\pi I_1^2N & -6/\pi I_1^2N & 14/I_1^2N\end{bmatrix},\quad N=N_x\cdot N_y \tag{8b}$$

The correlations between (N_f, φ) are irrelevant for the statistical errors assessment, only the diagonal elements of the covariance matrix are sufficient. All the parameters have variance inversely proportional to N, consequent to an universal statistical law.

It is now easy to find out the probability density function of the statistical fluctuations of **A** providing that the intensity fluctuations of all the pixels in the fringe pattern are statistically independent and each of them has the Gaussian distribution function f_G of δ_I variable and $\mu = 0$, $\sigma = \sigma_I$ parameters:

$$f(\delta_I) = \frac{1}{\sigma_I \cdot \sqrt{2\pi}} \cdot e^{-\frac{\delta_I^2}{2\sigma_I^2}} = f_G(\delta_I; 0, \sigma_I) \tag{9}$$

We see that according to relation (A15) there is a linear relationship between the variations of the parameters **A** and that of the intensity I:

$$\boldsymbol{\delta}_{\mathbf{A}} = \mathbf{B} \cdot \boldsymbol{\delta}_{\mathbf{I}}, \quad \boldsymbol{\delta}_{\mathbf{A},k} = \sum_{i=0}^{N-1} \mathbf{B}_{ki}\, \boldsymbol{\delta}_{\mathbf{I},i}, \quad k = 1,2,3,4,(5) \tag{10}$$

δ_I in relation (9) should not be mistaken for $\boldsymbol{\delta}_{\mathbf{I}}$ in relation (10). The former quantity is a scalar representing the intensity variation of a pixel in the fringe pattern, while the latter denotes the intensity variations all over the fringe pattern.

It is known that the probability density function of the sum of two independent random quantities equals the convolution of their probability density functions. Hence each of the parameter in the set of **A** has a probability density function that is the convolution of N Gaussian distribution function of different widths σ:

$$f(\boldsymbol{\delta}_{\mathbf{A},k}) = f_G(\delta_I; 0, \mathbf{B}_{k0} \cdot \sigma_I) * f_G(\delta_I; 0, \mathbf{B}_{k1} \cdot \sigma_I) * \ldots * f_G(\delta_I; 0, \mathbf{B}_{k,N-1} \cdot \sigma_I) \tag{11}$$

It is also known that convolving two Gaussian functions leads to another Gaussian function, according to the following rule:

$$f_G(x; \mu_1, \sigma_1) * f_G(x; \mu_2, \sigma_2) = f_G(x; \mu_1 + \mu_2, \sqrt{\sigma_1^2 + \sigma_2^2}) \tag{12}$$

Therefore each component $\boldsymbol{\delta}_{\mathbf{A},k}$ has the distribution $f_G(\boldsymbol{\delta}_{\mathbf{A},k};\ 0, \boldsymbol{\sigma}_{\mathbf{A},kk})$ and moreover, the fluctuations of the whole set of **A** have the multivariate normal distribution [12]:

$$f(\boldsymbol{\delta}_{\mathbf{A}}) = \left(2\pi \cdot \det(\boldsymbol{\sigma}_{\mathbf{A}}^2)\right)^{-\frac{r}{2}} \cdot e^{-\frac{1}{2}\boldsymbol{\delta}_{\mathbf{A}}^T \cdot (\boldsymbol{\sigma}_{\mathbf{A}}^2)^{-1} \cdot \boldsymbol{\delta}_{\mathbf{A}}}, \quad r = 4(5) \tag{13}$$

r being the number of parameters **A**, equal to 4 or 5 whether the fringe pattern is 1D or 2D. To obtain the covariance matrix with more precision, one needs to perform very difficult analytical calculation, which leads to very complicated results. That's why we restrict ourselves to the simple approximate results (8). In order to check the level of accuracy of these analytical formulae, we have processed a large number of computer simulated fringe pattern and we have performed a statistical analysis on the results to compare the actual errors with the theoretically predicted ones in Eq. (8).

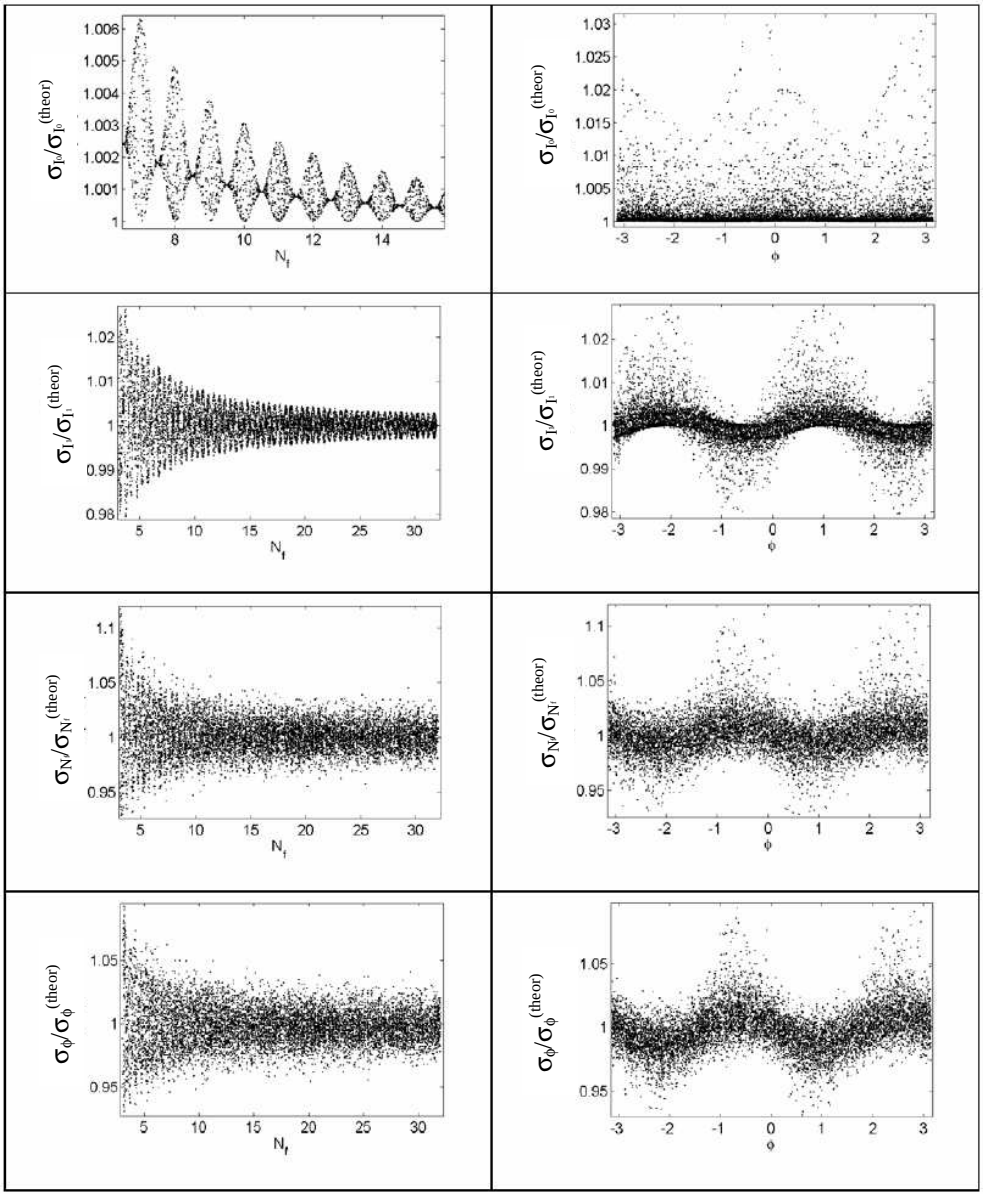
We simulate 10,000 1D fringe patterns of the form (2a), each of them with 256 samples. For all the fringe patterns in the set the parameters (I_0, I_1) were assigned the values (0.5, 0.4), which give a good fringe contrast and calibrate the intensity I(x) within the range 0÷1 that we suppose to be the data acquisition range (we take into account the previous mention that always $I_1 < I_0$). The parameter N_f was given uniform random values in the interval of 3÷32 fringes. We chose this interval because there is no sense to process fringe patterns with too few fringes ($N_f < 3$) and we limit the upper limit of fringe number to 32 to avoid aliasing associated with undersampling (we only consider well sampled fringes with sampling rate >8 samples/fringe spacing). The parameter φ was given uniform random values in the interval $-\pi \div \pi$. We add Gaussian noise of 0 mean and $\sigma_I = 0.05$ standard deviation. The noise due to usual 8-bit quantisation is very small, but the simulation took it into

account too. (Statistical mathemathics shows that by uniform quantisation with step q, the quantisation error is uniform in the interval $-q/2 \div q/2$ and has the variance $\sigma_q^2 = q^2/12$) .

In these conditions the relations (8a) give for the deviations of the parameters **A** the values: $\sigma_{\bar{A}}^{theor}$ =(0.0031, 0.0044, 0.0061, 0.022), equal to the squared root of the diagonal elements in the covariance matrix.

Let's denote with $\mathbf{A}_0$ the parameters given to the simulated fringe patterns, while $\overline{\mathbf{A}}$ and $\sigma_{\overline{\mathbf{A}}}$ denote their statistical mean values and standard deviations provided by the computer program (applying (A7) to the data I). We define the displacement between the exact values and the computed values of the fringe parameters: $\delta\mathbf{A} = \mathbf{A}_0 - \overline{\mathbf{A}}$ and we do statistics on the reduced parameters $\mathbf{y} = \delta\mathbf{A} / \sigma_{\overline{\mathbf{A}}}$ over the whole set of the 10,000 simulated fringe patterns. Here are the results for average and standard deviation: $\overline{\mathbf{y}}$ =(-0.63, -0.013, -0.0046, 0.0041), $\sigma_{\overline{\mathbf{y}}}$ =(0.997, 0.991, 0.998, 0.998). The mean values $\overline{\mathbf{y}}$ show that the estimations of the parameters $\overline{\mathbf{A}}$ are biased, but except the first component, these biases are quite small considering that they are expressed as fractions of $\sigma_{\overline{\mathbf{A}}}$.

Fig. 1. Results on processing a set of 10,000 harmonic fringe patterns.



The –0.63 bias of the I_0 intensity is mainly due to quantization. Quantization error is actually not a statistical error. It may be regarded as statistical only when the image has large variations. The DC component of an image is systematically biased by quantization. This bias is not greater than $\pm 1/N_q \sigma_{I_0}$, N_q being the number of quantization levels. If we repeat the whole procedure of statistics once again without quantizing the fringe patterns, we obtain the means $\bar{\mathbf{y}}$ =(0.0076, 0.0020, 0.0070, -0.0082) and the standard deviations $\sigma_{\bar{\mathbf{y}}}$ =(1.0075, 1.00077, 0.997, 0.997). This time all biases are very small and we can regard them as being only due to statistical fluctuations: $\bar{y} < \sigma_{\bar{y}} / \sqrt{N_p}$, where N_p=10,000 is the number of processed fringe patterns.

All the deviations $\sigma_{\bar{\mathbf{y}}}$ are close to 1, proving the fact that the computed values of the statistical errors on a single fringe pattern processing estimate very well the actual fluctuations that have been observed over the set of 10,000 simulations. Moreover, we have proved that the histograms of the statistical quantity **y** fit well with Gaussian distribution functions of parameters $(\bar{y}, \sigma_{\bar{y}})$, very close to standard normal distribution functions, which have (0, 1) parameters.

The Fig. 1 shows the result of the 10,000 simulations in form of graphs that show in what manner the number of fringes N_f and the phase parameter φ influence the numerically computed standard deviations of the parameters **A**. The top left plot has the narrower range from 6 to 16 fringes for the abscissa to give more details. Thus it is easy to see that the errors become minim when the fringe number N_f takes integer values. That's because the harmonic functions are orthogonal over ranges of integer periods. The covariance matrix (8) is calculated just assuming exact orthogonality between the functions sine and cosine involved in the sums (7).

We conclude that the formulas (8) estimate very well the actual statistical errors, even if the number of fringes is not large and there are only 256 sampling points. We should mention that for N_f =3 fringes, the difference between the theoretical and computed evaluation of the deviation for the parameter N_f is only 3%.

2.2. Non-harmonic periodic fringes

Often fringe patterns may be periodic, but not harmonic so that their intensity profile can be written as a Fourier series:

$$I(x) = I_0 + \sum_{k=1}^{\infty} l_k \cdot \cos(2k\pi\nu x + \varphi_k) \tag{14}$$

There are two main reasons why the fringe patterns to be processed have high-order harmonics:

- The non-harmonic profile is specific to certain interference conditions, for example in multiple beam interference.
- By acquiring data with a non-linear device the output signal becomes higher order harmonics.

Often we may need to process fringe patterns with unknown profile and determine the set of parameters **A**=(I_0, I_1, N_f, φ) as if the fringe were harmonic. Let's analyze if there is any sense in doing this.

If the number of fringes is very large, then the harmonic functions cos $k \cdot (2\pi\nu x + \varphi)$ build up an orthogonal system over the domain of the fringe pattern. If the Fourier series (14) were finite, then due to the orthogonality of the harmonics, the accuracy of this series would increases only by adding new terms in the series, not by adjusting the former ones. Thus, the parameters (I_0, I_1) computed with the algorithm for harmonic fringes will exactly match the 0th and the first orders of harmonic intensities in the non-harmonic fringe pattern. Also, the last two parameters (N_f, φ) estimate correctly the number of fringes and the phase offset parameter in the non-harmonic fringe pattern and they have the same accuracy as the fringes would have harmonic profile with the intensity I_1.

If the fringe number is not very large, then the orthogonality of the harmonic functions $\cos k \cdot (2\pi\nu x + \varphi)$ is distorted and the parameters (I_0, I_1) do not match exactly the orders 0 and 1 of the harmonic intensities in the non-harmonic fringe pattern, nor do the last two parameters (N_f, φ) estimate correctly the number of fringes and the phase parameter. All parameters (I_0, I_1, N_f, φ) computed with the assumption of harmonic fringes are biased estimations of the corresponding parameters in the non-harmonic fringe pattern, and they tend to the exact values as the number of fringes increases (surely, the sampling rate must be increased correspondingly to avoid aliasing).

To see the conditions suitable for good results on processing non-harmonic fringes with the computer algorithm for harmonic fringes, we use a statistical analysis on the results of processing a large set (10,000) of computer simulated fringe patterns.

We simulate 10,000 1D Fabry Perot fringe patterns, each of them with 256 samples. For all the fringe patterns in the set the parameters (I_0, I_1) were assigned the values (0.25, 0.15), so that the intensity is confined within the range 0÷1. Fig. 2 displays the actual fringe profile of the simulated fringes. The parameter N_f was given uniform random values in the interval of 3÷32 fringes and the parameter φ was given uniform random values in the interval $-\pi \div \pi$. We add Gaussian noise with 0 mean and σ_I =0.05 standard deviation. In these conditions the relations (8a) gives for the deviations of the parameters **A** the values: $\sigma_{\bar{A}}^{theor}$ =(0.0031, 0.0044, 0.016, 0.059).

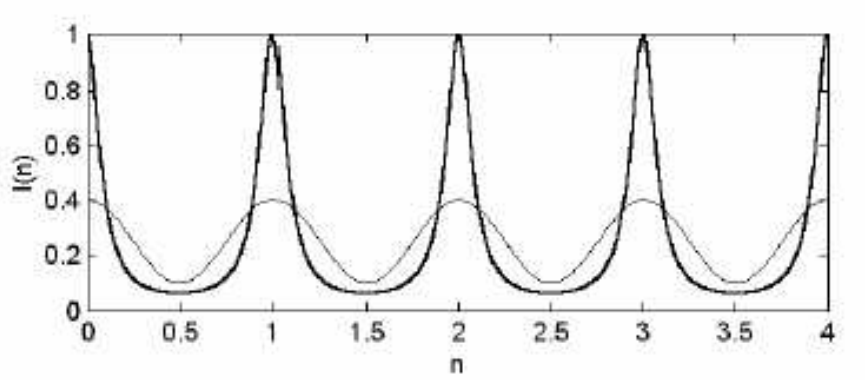
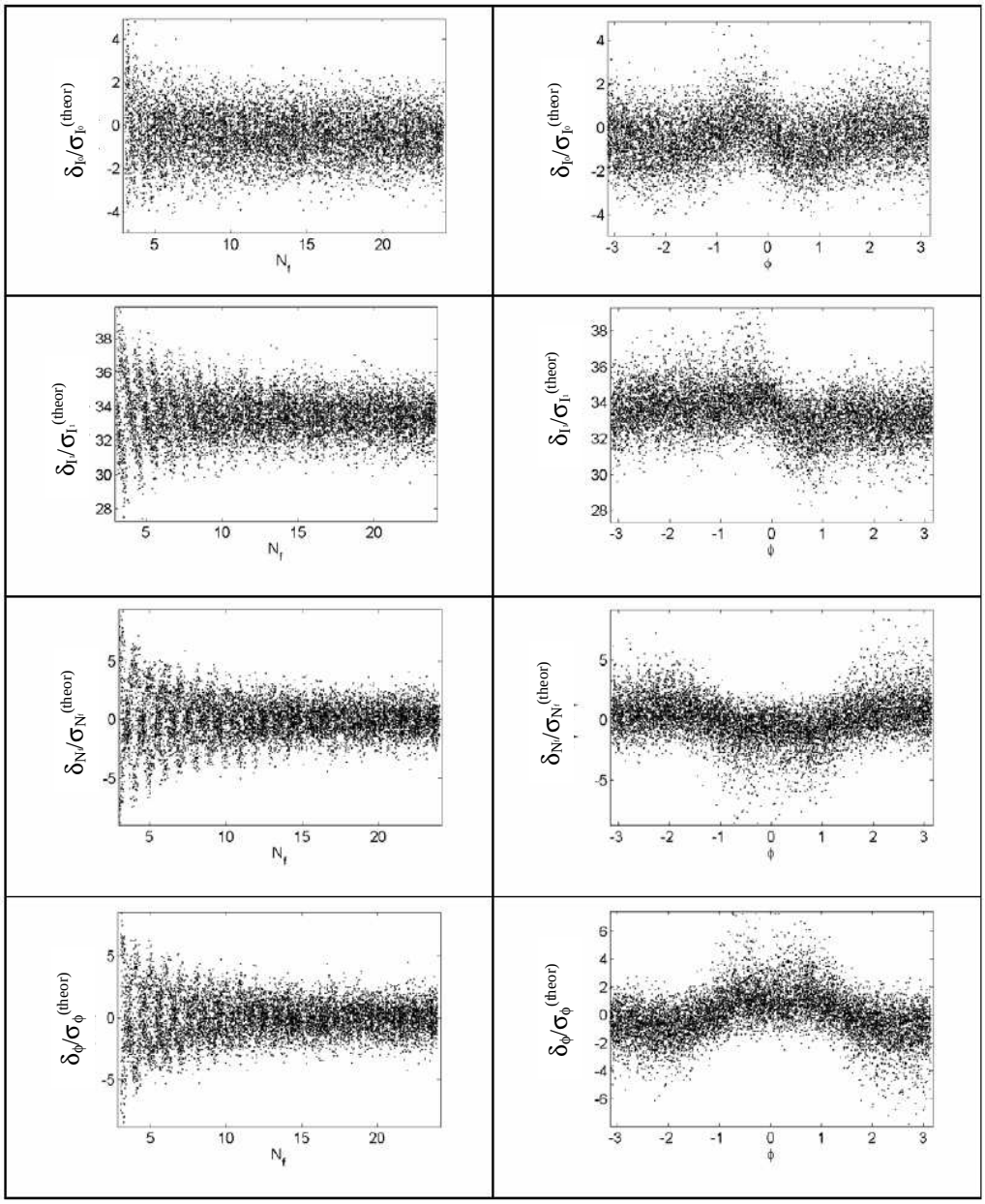


Fig. 2. Periodic Fabry -Pérot fringes.

The Fig. 3 shows the result of the 10,000 simulations in form of graphs that indicate how the number of fringes N_f and the phase parameter φ influence the displacements of the parameters **A**. We see that these influence becomes insignificant if N_f >8. This time we preferred to display the displacements $\delta A / \sigma_{\bar{A}}$ instead the normalized standard deviations $\delta A / \sigma_{\bar{A}}^{\text{theor}}$ as we did for harmonic fringes, because we expected greater variations in respect to parameter N_f, so that the histograms of the displacements (which are the actual statistical errors) don't fit with Gaussian distribution functions if the number of fringes is not large enough.

Next we repeat a statistic over a new set of 10,000 computer-simulated fringes, with the same parameters, except N_f that was given uniform random values in the range 8÷20 fringes. We have changed the range of N_f thinking that so the errors $\delta A / \sigma_{\bar{A}}$ will become a distribution closer to the standard normal distribution one. This time for the reduced parameter $y = \delta A / \sigma_{\bar{A}}$ we obtain the following results for average and standard deviation: $\bar{\mathbf{y}}$ =(-0.46, 33.4, 0.017, -0.0079), $\sigma_{\bar{y}}$ =(1.04, 1.12, 1.35, 1.29). In comparison with the harmonic fringes, the estimations of the parameters **A** are more biased, especially I_1 has a displacement of $33\sigma_{I_1}$. The parameters (N_f, φ) have small displacements. The deviations $\sigma_{\bar{y}}$ are greater than 1, showing that the computed statistic errors are underestimated. The histogram of the statistical quantity **y** for each component fits well with a Gaussian distribution.

Fig. 3. Results on processing a set of 10,000 Fabry-Pérot fringe patterns.



We conclude that the LSF algorithm made for harmonic fringe patterns can be used to process strongly non-harmonic fringe patterns as well, being able to compute accurately the number of fringes and the phase offset (N_f, φ), if at least 12 well sampled fringes are present in the fringe pattern to be processed. It's obvious one can make a LSF routine that takes into account more parameters related to high order harmonics, but the amount of computation increases as the squared number of parameters.

2.3. Modulated fringes

Most fringe patterns do not have uniform contrast, they are modulated so that their intensity profile may be written with variable coefficients (I_0, I_1) and the fringes are said to be quasiperiodic:

$$I(x) = I_0(x) + I_1(x) \cdot \cos(2\pi\nu x + \varphi) \tag{15}$$

Obviously, the most general fringe profile may be written as a Fourier series (14) with variable intensities I_k.

Often we may use to process various quasiperiodic fringe patterns and determine the set of parameters $\mathbf{A}=(I_0, I_1, N_f, \varphi)$ as if the fringes were harmonic. Let's analyze if there is a reason for doing this.

In the Fourier spectrum of the function (15), the component of frequency ν is widened, its spectral profile being the spectrum of the function $I_1(x)$ (due to the convolution of the function $I_1(x)$ with the harmonic function). Similarly, the component of frequency 0 takes the spectrum of the function $I_0(x)$.

By least squares fitting of the function (15) with the harmonic function (2a) the computed parameters $\mathbf{A}=(I_0, I_1, N_f, \varphi)$ have the following meanings: $I0$, $I1$ become the average of the functions $I_0(x)$, $I_1(x)$ respectively, while N_f and φ estimate the number of fringes and the phase parameter in the formula (15), but all these values are biased estimations and they tend to be exact as the number of fringes increases.

To see the conditions suitable for good results on processing quasiperiodic fringes with the computer algorithm for harmonic fringes, we made statistics on the results of processing of a large set (10,000) of computer simulated fringe patterns.

We simulate 10,000 1D modulated fringe patterns, each of them with 256 samples, with such a profile as that of the interference pattern of two Gaussian beams. Fig. 4 displays the actual fringe profiles used for the simulated fringes. For all the fringe patterns in the set the parameters (I_0, I_1) were assigned the values (0.42, 0.334), so that the intensity lies in the range 0÷1. The parameter N_f was given uniform random values in the interval of 3÷24 fringes and the parameter φ was given uniform random values in the interval $-\pi \div \pi$. We add Gaussian noise of 0 mean and σ_I =0.05 standard deviation. In these conditions the relations (8a) give for the deviations of the parameters **A** the values: $\sigma_{\bar{A}}^{theor}$ =(0.0031, 0.0073, 0.016, 0.026).

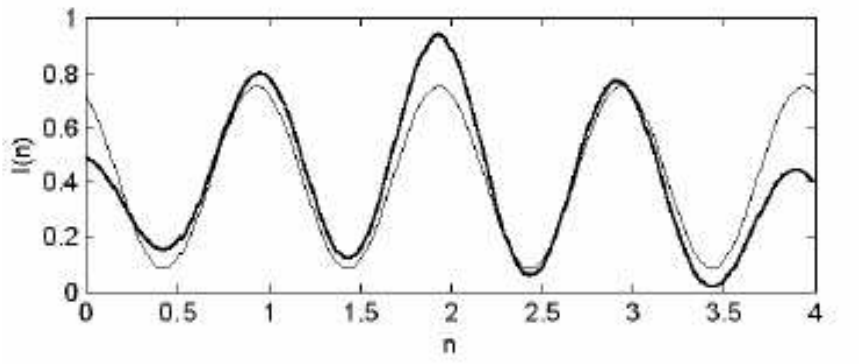
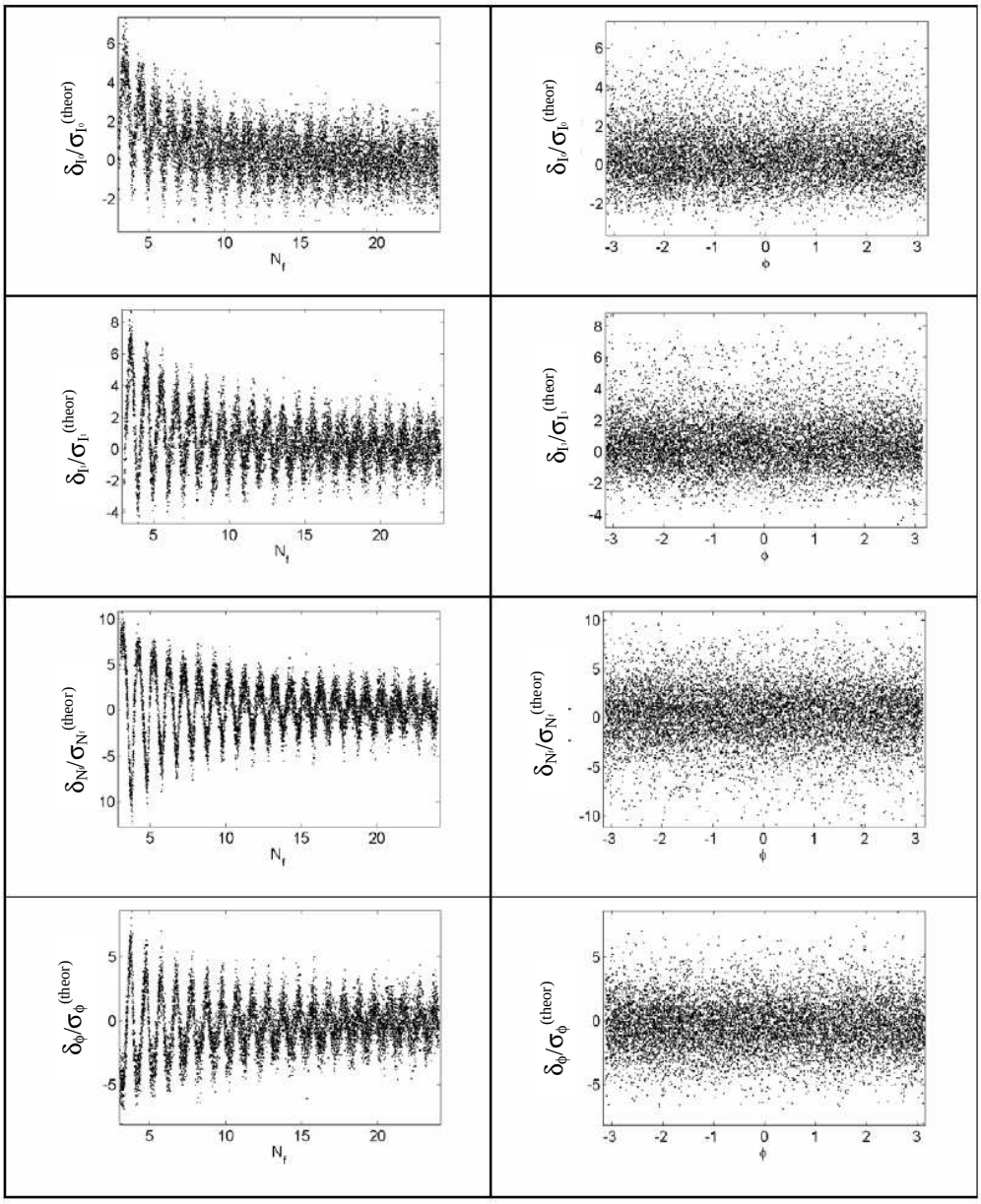


Fig. 4. Modulated fringes.

The Fig. 5 shows the result of the 10,000 simulations in form of graphs that show how the number of fringes N_f and the phase parameter φ influence the displacements of the parameters **A**. We see that these influence becomes insignificant if N_f >12.

Fig. 5. Results on processing a set of 10,000 modulated fringe patterns.



We further made a statistical analysis over a new set of 10,000 computer-simulated fringes, with the same parameters, except for N_f that was given uniform random values in the range 12÷24 fringes, so that the distribution of the errors $\delta A/\sigma_{\bar{A}}$ becomes closer to the standard normal distribution. This time for the reduced parameter $\mathbf{y} = \delta\mathbf{A}/\sigma_{\bar{\mathbf{A}}}$ we obtain the following results for the average and deviation: $\bar{\mathbf{y}}$=(-0.14, 0.15, -0.89, 0.99), $\sigma_{\bar{\mathbf{y}}}$=(1.03, 1.15, 1.80, 1.44). In comparison with the harmonic fringes, the estimations of the parameters **A** are more biased, especially the parameters (N_f, φ) have displacements of about 1σ. The deviations $\sigma_{\bar{\mathbf{y}}}$ are greater than 1, thus the numerically computed statistical errors are underestimated up to 80% in the case of parameter N_f. The histogram of the statistical quantity **y** for each component fits well the Gaussian distribution.

We conclude that the LSF algorithm for harmonic fringe patterns is also suitable for rocessing quasiperiodic fringe patterns to compute accurately the number of fringes and the phase offset (N_f, φ), if at least 20 well sampled fringes are provided in the fringe pattern to be processed.

3. Conclusions

We develop analytical expressions for the statistical errors concerning the measurement of the parameters $\mathbf{A}=(I_0,\ I_1,\ N_f,\ \varphi)$ for the simplest fringe pattern (2a,b). These theoretical results obtained using an asymptotic approximation (valid for a large number of fringes) give the upper limit of accuracy the fringe parameters can be determined with this method.

Performing statistics on a set of 10,000 simulated fringe patterns, we have proved that the analytical formulae (8) estimate correctly the statistical errors for harmonic fringes if the fringe pattern has at least 3 fringes, which are well sampled (the sampling rate considered was >8 sample/period).

We stated the valid conditions for processing non-harmonic fringes with the LSF algorithm for harmonic fringes. Doing statistics on the results of processing a set of 10,000 simulated Fabry Perot fringes, we proved that we get the fringe parameters with high accuracy if there are at least 8 fringes in the fringe pattern. We should mention that although the numerically computed statistical errors are underestimated (and not always reliable), the theoretical statistical errors (8) always estimate very well the actual statistical errors if the number of fringes are greater than 8.

In the same manner we treated the processing of quasiperiodic fringe with the LSF algorithm made for harmonic fringes. By statistical processing of a set of 10,000 computer simulated fringe patterns, we proved that we can calculate the fringe parameters with good accuracy if there are at least 12 fringes in the fringe pattern. Also in this case, the numerically computed statistical errors are underestimated, but the theoretical results (8) always estimate very well the actual statistical errors if the number of fringes are greater than 12.

The general conclusion is that the fringes must be harmonic with maximum contrast to obtain maximum precision in measuring its parameters. However, even for strongly non-harmonic fringes, the estimation of the parameters (N_f, φ) is very good if the fringe profile is periodic and if there are at least 10-12 fringes available. Special care must be taken for processing of quasiperiodic fringes (fringes with non-uniform contrast, or modulated fringes). In this case especially the parameters (N_f, φ) can have large displacements and fluctuations if there are not at least 20 fringes in the fringe pattern to be processed.

The results of this paper are applicable, for example, to predict the precision that one can achieve when measuring wavelengths, lengths and displacements using high precision interferometers with stabilized lasers. They give the answer to the following questions: What resolution (number of pixels) must have the CCD used to acquire the fringe pattern to measure wavelengths, lengths or displacements with a given accuracy? How do these errors depend on the number of fringes? It has been shown that these errors do not depend on the number of fringes, if this number is large enough. They depend only on the number of samples N and the fringe contrast.

4. Appendix

4.1. Least squares fitting for linear combinations of functions

Let x, y be two quantities which have a functional relationship that we approximate by a linear combination of r independent functions:

$$y(x)=a_1\cdot \mathrm{f}_1(x)+a_2\cdot \mathrm{f}_2(x)+\ldots+a_r\cdot \mathrm{f}_r(x)=\mathbf{a}^T\cdot\mathbf{f}(x)$$
$$\mathbf{a}=\begin{bmatrix}a_1\\a_2\\\vdots\\a_r\end{bmatrix},\quad \mathbf{f}(x)=\begin{bmatrix}\mathrm{f}_1(x)\\\mathrm{f}_2(x)\\\vdots\\\mathrm{f}_r(x)\end{bmatrix} \tag{A1}$$

The symbol ^{T} represents the operation of matrix transposition. Let us consider that $y(x)$ is a scalar function and x may be a single variable or a set of independent variables.

Let $\mathbf{y}$ be a set of experimental results (statistical data), measured with variance σ_y^2 to which a set of data $\mathbf{x}$ is associated. We suppose the data $\mathbf{x}$, belonging to the variables x, have zero variance.

$$\mathbf{y} = \begin{bmatrix} y^{(1)} \\ y^{(2)} \\ \vdots \\ y^{(N)} \end{bmatrix}, \quad \mathbf{x} = \begin{bmatrix} x^{(1)} \\ x^{(2)} \\ \vdots \\ x^{(N)} \end{bmatrix} \tag{A2}$$

In the most general case σ_y^2 is the covariance matrix of the values $\mathbf{y}$, has the dimension $N \times N$, and contains the variances on the main diagonal while the other elements are the covariations between the various data $y^{(i)}$, i=1,2,...,N. In many practical situations all the values $\mathbf{y}$ are measured with the same accuracy and are statistically independent, so that the quantity σ_y^2 becomes a scalar.

Alternatively we use as well another equivalent quantity, called statistical weight, defined by $\mathbf{P}_y = \left[\boldsymbol{\sigma}_\mathbf{y}^2\right]^{-1}$.

We are attempting to compute the coefficients $\mathbf{a}$ from relation (A1) so that this relation best approximate the experimental data. For this purpose we use the least squares fitting method from statistical mathematics, widely used for data processing. Firstly we define a column vector whose elements are the difference between the data $\mathbf{y}$ and the assumed function $y(x)$:

$$\begin{aligned} &\boldsymbol{\delta}(\mathbf{a}) = y(\mathbf{x}) - \mathbf{y} = \mathbf{f}^T(\mathbf{x}) \cdot \mathbf{a} - \mathbf{y} = \mathbf{D} \cdot \mathbf{a} - \mathbf{y}, \\ &\mathbf{D} = \mathbf{f}^T(\mathbf{x}) = \begin{bmatrix} \mathrm{f}_1\left(x^{(1)}\right) & \mathrm{f}_2\left(x^{(1)}\right) & \cdots & \mathrm{f}_r\left(x^{(1)}\right) \\ \mathrm{f}_1\left(x^{(2)}\right) & \mathrm{f}_2\left(x^{(2)}\right) & \cdots & \mathrm{f}_r\left(x^{(2)}\right) \\ \vdots & \vdots & \cdots & \vdots \\ \mathrm{f}_1\left(x^{(N)}\right) & \mathrm{f}_2\left(x^{(N)}\right) & \cdots & \mathrm{f}_r\left(x^{(N)}\right) \end{bmatrix}, \quad \mathbf{D}^T = \mathbf{f}\left(\mathbf{x}^T\right) \end{aligned} \tag{A3}$$

As a global assessment of all these deviations one builds up a scalar function $\chi^2(\mathbf{a})$ which represents the sum of the squared normalized deviations of the experimental data $\mathbf{y}$ from the assumed function $y(x)$:

$$\chi^2(\mathbf{a}) = \sum_{i,j=1}^{N} P_y^{(ij)} \cdot \delta^{(i)}(\mathbf{a}) \cdot \delta^{(j)}(\mathbf{a}) = \boldsymbol{\delta}^T(\mathbf{a}) \cdot \mathbf{P}_y \cdot \boldsymbol{\delta}(\mathbf{a}) \tag{A4}$$

Least squares fitting consists in finding the values of the parameters $\mathbf{a}$ for which the function $\chi 2(\mathbf{a})$ reaches its minimum. The values of the parameters $\mathbf{a}$ for which the function $\chi^2(\mathbf{a})$ is minimum are obtained by solving an equation system consisting of r linear equations established by equaling to zero the partial derivatives of the function $\chi^2(\mathbf{a})$ relative to each of the parameters $\mathbf{a}$. This is called the "normal equation system":

$$\frac{\partial \chi^2(\mathbf{a})}{\partial \mathbf{a}} = 2\mathbf{D}^T \cdot \mathbf{P}_y \cdot [\mathbf{D} \cdot \mathbf{a} - \mathbf{y}] = 0 \Rightarrow \left[\mathbf{D}^T \cdot \mathbf{P}_y \cdot \mathbf{D}\right] \cdot \mathbf{a} = \left[\mathbf{D}^T \cdot \mathbf{P}_y\right] \cdot \mathbf{y} \tag{A5}$$

The solution of this system is:

$$\bar{\mathbf{a}} = \mathbf{B} \cdot \mathbf{y}, \quad \mathbf{B} = \left[\mathbf{D}^T \cdot \mathbf{P}_y \cdot \mathbf{D}\right]^{-1} \cdot \left[\mathbf{D}^T \cdot \mathbf{P}_y\right] \tag{A6}$$

It is interesting to notice that there is a linear relationship between the data $\mathbf{y}$ and the computed parameters $\bar{\mathbf{a}}$. The parameters $\mathbf{a}$ are statistical quantities. Relation (A6) gives their

statistical mean $\bar{\mathbf{a}}$. Using the error propagation theorem we obtain the covariance matrix of these parameters, that contains on the main diagonal their variances:

$$\boldsymbol{\sigma}_{\bar{\mathbf{a}}}^2 = \mathbf{B} \cdot \boldsymbol{\sigma}_{\mathbf{y}}^2 \cdot \mathbf{B}^T = \left[\mathbf{D}^T \cdot \mathbf{P}_y \cdot \mathbf{D}\right]^{-1} = \left[\mathbf{P}_{\bar{\mathbf{a}}}\right]^{-1} \tag{A7}$$

We find out that the function $\chi^2(\mathbf{a})$ is an *r*-dimensional parabola in the components of the vector **a**:

$$\begin{aligned} \chi^2(\mathbf{a}) &= \left[\mathbf{D}\cdot\mathbf{a}-\mathbf{y}\right]^T \cdot \mathbf{P}_y \cdot \left[\mathbf{D}\cdot\mathbf{a}-\mathbf{y}\right] = \\ &= \left[\mathbf{D}\cdot(\mathbf{a}-\bar{\mathbf{a}})+\mathbf{D}\cdot\bar{\mathbf{a}}-\mathbf{y}\right]^T \cdot \mathbf{P}_y \cdot \left[\mathbf{D}\cdot(\mathbf{a}-\bar{\mathbf{a}})+\mathbf{D}\cdot\bar{\mathbf{a}}-\mathbf{y}\right] = \\ &= \left[\mathbf{a}-\bar{\mathbf{a}}\right]^T \cdot \mathbf{P}_{\bar{\mathbf{a}}} \cdot \left[\mathbf{a}-\bar{\mathbf{a}}\right] + \chi_0^2, \quad \chi_0^2 = \left[\mathbf{D}\cdot\bar{\mathbf{a}}-\mathbf{y}\right]^T \cdot \mathbf{P}_y \cdot \left[\mathbf{D}\cdot\bar{\mathbf{a}}-\mathbf{y}\right] = const \end{aligned} \tag{A8}$$

A section of this parabola to one component, say a_1, is a simple 1D parabola, shown in Fig. 3:

$$\chi^2(a_1) = \frac{(a_1-\bar{a}_1)^2}{\sigma_{\bar{a}_1}^2} + \chi_0^2 \tag{A9}$$

Thus, the parabola $\chi^2(\mathbf{a})$ reaches its minimum in **a** and has an extent corresponding to the variances $\sigma_{\bar{a}}^2$.

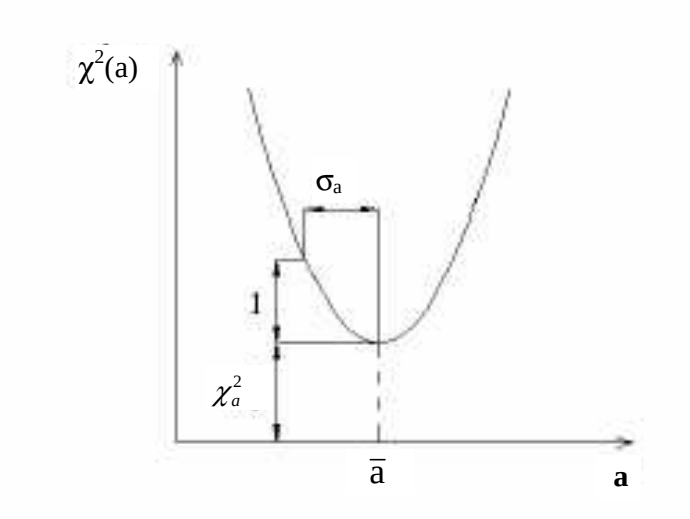


Fig. 6. For linear LSF the function χ^2 is a parabola.

Using the criterion of minimizing the function $\chi^2(\mathbf{a})$ for computing the parameters **a** that give the best fit of the experimental data **y** by the formula (A1) is based on the maximum plausibility principle from statistical mathematics. It is right to use this criterion only with the assumption that the experimental data **y** deviates from the formula (A1) only due to normal distributed fluctuations with 0 mean and σ_y^2 variances. Otherwise, a different criterion is necessary.

If the deviations $\delta(\bar{\mathbf{a}})$ are normally distributed, then the minimum of the function $\chi^2(\mathbf{a})$, namely the value $\chi_0^2 = \chi^2(\bar{\mathbf{a}})$, which is also a statistical quantity, has a χ^2 distribution with *N* degrees of freedom. The value χ_0^2 is useful to check if the relation (A1) fits well the experimental data **y**. We distinguish two main cases:

- If $\chi_0^2 < N \cdot r$ it is highly probable that the relation (A1) approximates very well the experimental data **y**, which deviate from the relation (A1) only due to normal distributed statistical fluctuation with 0 mean and 2 y σ variances.

- If $\chi_0^2 < N \cdot r$ the relation (A1) is not suitable for fitting the experimental data **y** or the assumption that the deviations $\delta(\bar{\mathbf{a}})$ have normal distribution with 0 mean and σ_y^2 variances is not valid.

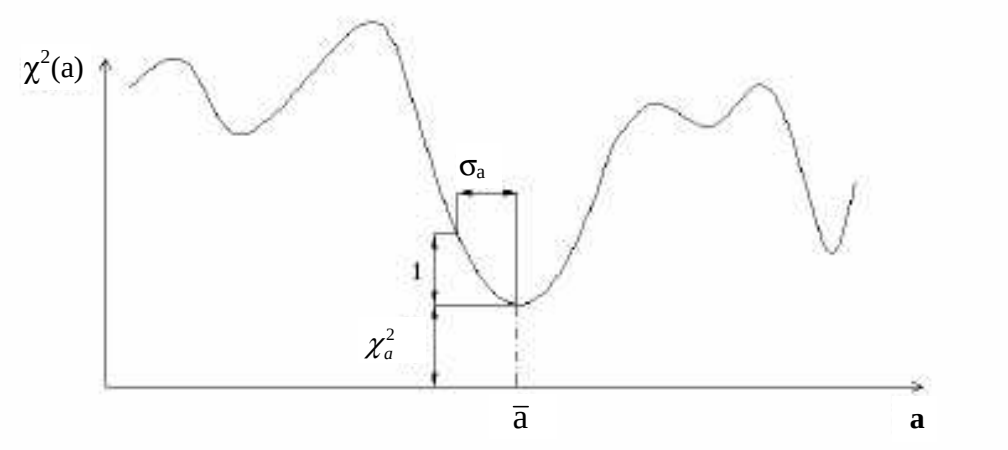


Fig. 7. For nonlinear LSF the function χ^2 has a quadratic form around it's global minimum.

4.2. Least squares fitting for nonlinear functions

Let's assume a certain nonlinear relationship between two quantities y and x, which cannot be expressed as a linear combination of independent functions:

$$y = \mathrm{f}(x, \mathbf{a}) \tag{A10}$$

The function f(x, **a**) has a known behavior, but depends on the set of parameters **a**, which are to be determined in order to find the best fit of an experimental data set (**x**, **y**) by this function, in the sense of minimizing the function χ^2(**a**). We use the assumption that the data **y** is affected by normal distributed fluctuations with 0 mean and σ_y^2 variances.

The normal equations for determining the set of parameters **a** are generally not linear. They cannot be always solved analytically, and the numerical solution can often prove to be difficult. The problem can be tackled iteratively, as shown further. We start from an arbitrary initial approximation $\mathbf{a}^{(0)}$ of the parameter set **a**, obtained by some method or other, with an arbitrary accuracy. The non-linear function f(**x**, **a**) is developed as a Taylor series around the values $\mathbf{a}^{(0)}$ and the linear term is saved:

$$\mathrm{f}(\mathbf{x}, \mathbf{a}) \cong \mathrm{f}(\mathbf{x}, \mathbf{a}^{(0)}) + \mathbf{D} \cdot \left(\mathbf{a} - \mathbf{a}^{(0)}\right)$$

$$\mathbf{D} = \left.\frac{\partial \mathrm{f}(\mathbf{x}, \mathbf{a})}{\partial \mathbf{a}}\right|_{\mathbf{a}=\mathbf{a}^{(0)}} = \left[\begin{array}{cccc} \frac{\partial \mathrm{f}(x^{(1)}, \mathbf{a})}{\partial a_1} & \frac{\partial \mathrm{f}(x^{(1)}, \mathbf{a})}{\partial a_2} & \cdots & \frac{\partial \mathrm{f}(x^{(1)}, \mathbf{a})}{\partial a_r} \\ \frac{\partial \mathrm{f}(x^{(2)}, \mathbf{a})}{\partial a_1} & \frac{\partial \mathrm{f}(x^{(2)}, \mathbf{a})}{\partial a_2} & \cdots & \frac{\partial \mathrm{f}(x^{(2)}, \mathbf{a})}{\partial a_r} \\ \vdots & \vdots & \cdots & \vdots \\ \frac{\partial \mathrm{f}(x^{(N)}, \mathbf{a})}{\partial a_1} & \frac{\partial \mathrm{f}(x^{(N)}, \mathbf{a})}{\partial a_2} & \cdots & \frac{\partial \mathrm{f}(x^{(N)}, \mathbf{a})}{\partial a_r} \end{array}\right]_{\mathbf{a}=\mathbf{a}^{(0)}} \tag{A11}$$

The N-element vector that contains the deviations of the function $f(\mathbf{x}, \mathbf{a})$ from the experimental values $\mathbf{y}$ is:

$$\boldsymbol{\delta}(\mathbf{a}) = y(\mathbf{x}) - \mathbf{y} = \mathrm{f}(\mathbf{x}, \mathbf{a}) - \mathbf{y} = \mathbf{D} \cdot \left(\mathbf{a} - \mathbf{a}^{(0)}\right) - \boldsymbol{\delta}_0, \quad \boldsymbol{\delta}_0 = \mathbf{y} - \mathrm{f}\left(\mathbf{x}, \mathbf{a}^{(0)}\right) \tag{A12}$$

and the sum of the squared and normalized deviations is:

$$\chi^2(\mathbf{a}) = \boldsymbol{\delta}^T(\mathbf{a}) \cdot \mathbf{P}_y \cdot \boldsymbol{\delta}(\mathbf{a}) = \left[\mathbf{D} \cdot \left(\mathbf{a} - \mathbf{a}^{(0)}\right) - \boldsymbol{\delta}_0\right]^T \cdot \mathbf{P}_y \cdot \left[\mathbf{D} \cdot \left(\mathbf{a} - \mathbf{a}^{(0)}\right) - \boldsymbol{\delta}_0\right] \tag{A13}$$

The normal equation system has the same form as (A5), but instead of $\mathbf{a}$ we have $\mathbf{a}$-$\mathbf{a}^{(0)}$, and $\mathbf{y}$ is replaced with $\boldsymbol{\delta}_0 = \mathbf{y} - \mathrm{f}(\mathbf{x}, \mathbf{a}^{(0)})$:

$$\frac{\partial \chi^2(\mathbf{a})}{\partial \mathbf{a}} = 2\mathbf{D}^T \cdot \mathbf{P}_y \cdot \left[\mathbf{D} \cdot \left(\mathbf{a} - \mathbf{a}^{(0)}\right) - \boldsymbol{\delta}_0\right] = 0 \Rightarrow \left[\mathbf{D}^T \cdot \mathbf{P}_y \cdot \mathbf{D}\right] \cdot \left(\mathbf{a} - \mathbf{a}^{(0)}\right) = \left[\mathbf{D}^T \cdot \mathbf{P}_y\right] \cdot \boldsymbol{\delta}_0 \tag{A14}$$

Obviously, its solution has the same form as (A6):

$$\bar{\mathbf{a}} - \mathbf{a}^{(0)} = \mathbf{B} \cdot \boldsymbol{\delta}_0, \quad \mathbf{B} = \left[\mathbf{D}^T \cdot \mathbf{P}_y \cdot \mathbf{D}\right]^{-1} \cdot \left[\mathbf{D}^T \cdot \mathbf{P}_y\right] \tag{A15}$$

By now we have the result of the first iteration of a series that must be continued. The result $\mathbf{a}$ thus obtained is input for a new iteration instead of the zero-order approximation $\mathbf{a}^{(0)}$ and the procedure is repeated until the difference between the results of two successive iterations becomes negligible. Generally the algorithm has a rapid convergence, but not always. After the final result was obtained with sufficient accuracy, its variances can be calculated using the same formula as in the linear case, (A7). Let's analyze the form of the function $\chi^2(\mathbf{a})$. If we develop the nonlinear function $f(\mathbf{x}, \mathbf{a})$ in first order Taylor series around the values $\bar{\mathbf{a}}$:

$$\mathrm{f}(\mathbf{x}, \mathbf{a}) = \mathrm{f}(\mathbf{x}, \bar{\mathbf{a}}) + \mathbf{D} \cdot (\mathbf{a} - \bar{\mathbf{a}}), \quad \mathbf{D} = \left.\frac{\partial \mathrm{f}(\mathbf{x}, \mathbf{a})}{\partial \mathbf{a}}\right|_{\mathbf{a}=\bar{\mathbf{a}}} \tag{A16}$$

then we obtain an r-dimensional parabola form for the function $\chi^2(\mathbf{a})$, having the minimum in $\mathbf{a}$ and an extent that depends on the variances σ_y^2:

$$\begin{aligned} \chi^2(\mathbf{a}) &= \left[\mathrm{f}(\mathbf{x}, \mathbf{a}) - \mathbf{y}\right]^T \cdot \mathbf{P}_y \cdot \left[\mathrm{f}(\mathbf{x}, \mathbf{a}) - \mathbf{y}\right] = \\ &= \left[\mathbf{D} \cdot (\mathbf{a} - \bar{\mathbf{a}}) + \mathrm{f}(\mathbf{x}, \bar{\mathbf{a}}) - \mathbf{y}\right]^T \cdot \mathbf{P}_y \cdot \left[\mathbf{D} \cdot (\mathbf{a} - \bar{\mathbf{a}}) + \mathrm{f}(\mathbf{x}, \bar{\mathbf{a}}) - \mathbf{y}\right] = \\ &= \left[\mathbf{a} - \bar{\mathbf{a}}\right]^T \cdot \mathbf{P}_{\bar{\mathbf{a}}} \cdot \left[\mathbf{a} - \bar{\mathbf{a}}\right] + \chi_0^2, \quad \chi_0^2 = \left[\mathrm{f}(\mathbf{x}, \bar{\mathbf{a}}) - \mathbf{y}\right]^T \cdot \mathbf{P}_y \cdot \left[\mathrm{f}(\mathbf{x}, \bar{\mathbf{a}}) - \mathbf{y}\right] = const \end{aligned} \tag{A17}$$

Outside the vicinity of $\mathbf{a}$, the function $\chi^2(\mathbf{a})$ is no longer parabolic, as shown in figure 4 that presents a one-dimensional section of this r-variable function. The basic idea of this algorithm consists in the fact that the function $\chi^2(\mathbf{a})$ which depends on the set $\mathbf{a}$ of r parameters has an approximately r-dimensional quadratic shape around $\mathbf{a}$ (no matter if the function $\mathrm{f}(x, \mathbf{a})$ is linear or not) and its minimum can be determined as a solution of a linear equation system. In the non-linear case, the result obtained after each iteration progressively draws near the solution, except in some inappropriate singularity cases, when the convergence is not attained. As a general rule, to ensure convergent iterations, the first approximation $\mathbf{a}^{(0)}$ must be close enough to the exact values $\mathbf{a}$, otherwise the iterations may diverge, or they may lead to a local minimum. It is remarkable that all the operations described for this algorithm are carried out using only linear algebra.

References

[1] K. R. V. Murty, Applied Optics **3**, 531 (1964).
[2] D. Apostol, I. Apostol, I. N. Mihăilescu, Infrared Physics **16**, 269 (1976).
[3] D. Apostol, V. Damian, St. Cerc. Fiz. **1**, 97 (1989).
[4] J. Schwider, R. Burow, K. E. Elssner, J. Grzana, R. Spolaczyk, K. Merkel, App. Opt. **22**, 3421 (1983).
[5] C. Brophy, J. Opt. Soc. Am. A**7**, 537 (1990).
[6] G. Lai, T. Yatagai, J. Opt. Soc. Am. A**8**, 822 (1991).
[7] C. Rathjen, J. Opt. Soc. Am. A **12**, 1997 (1995).
[8] B. Zhao, App. Opt. **36**, 2070 (1997).
[9] D. W. Robinson, Applied Optics **14**, 2169 (1983).
[10] D. W. Robinson, G. T. Reid, Interferogram Analysis, Institute of Physics Publishing, Bristol and Philadelphia, 1993.
[11] G. Lai, T. Yatagai, App. Opt. **33**, 5935 (1994).
[12] www.riskglossary.com/articles/joint_normal_distribution.htm, about multivariate normal distribution.