PERIODICA POLYTECHNICA SER. EL. ENG. VOL. 41, NO. 3, PP. 167-183 (1997)

ON TWO-POINT RESOLUTION OF IMAGING SYSTEMS

A. J. DEN DEKKER

Department of Applied Physics
Delft University of Technology
P.O. Box 5046, 2600 GA Delft
The Netherlands
Tel: +31 0 15-278-1823
Fax: +31 0 15-278-4263
E-mail: dekker@duttncb.tn.tudelft.nl

Received: Nov. 28, 1995

Abstract

In this paper a new criterion is proposed for optical two-point resolution, applicable to coherent, incoherent, and partially coherent imaging. Unlike classical resolution criteria, such as Rayleigh's, the new criterion takes account of the presence of errors in the observed intensity distributions. Based on a parameter estimation approach, it shows how the resolvability of the imaged point sources depends on these errors. Additionally, a test for the resolvability of the point sources from a given set of observations is presented. Moreover, a procedure is proposed for the computation of the errors having minimum energy among all errors undermining the resolution. The results presented include, as a special case, earlier results on two-point resolution of strictly incoherent imaging systems.

Keywords: two-point resolution, imaging system, resolution limit, parameter estimation, partially coherent imaging.

1. Introduction

The two-point resolution of an imaging system, that is, its ability to resolve two point sources of equal intensity, is widely in use as a measure for the system's resolving capabilities. In the past, many resolution criteria have been proposed based on the assumption that the resolving power of an imaging system is limited by the diffraction at the aperture of the imaging lens of the system. Due to this diffraction, a point source is not imaged as a point image, but as the Fraunhofer diffraction pattern of the aperture. Hence, this diffraction pattern can be regarded as the point spread function of the imaging system. Of all the proposed criteria referred to above, the classical Rayleigh resolution criterion [1] is surely the most famous. According to the Rayleigh criterion, originally derived for incoherent imaging, two point sources are to be considered as just resolved if the central maximum of the intensity diffraction pattern produced by one point source coincides with the first zero of the intensity diffraction pattern produced by the other point source. This means that Rayleigh's resolution limit is

given by the distance between the central maximum and the first zero of the intensity point spread function of the imaging system concerned. In Rayleigh's considerations the human visual system has been employed in the role of a sensor used to detect differences in intensity at various points of the composite intensity distribution produced by the two point sources together. So, obviously, Rayleigh based his criterion not only on the properties of the diffraction-limited imaging system, but also on presumed resolving capabilities of the human visual system. Consequently, Rayleigh's resolution limit cannot be regarded as a law of physics, but rather as a rule of thumb. Other notable examples of resolution criteria, involving both the properties of the imaging system and the human visual system, are those of BUXTON, HOUSTON, SCHUSTER and SPARROW [2–3]. All these so called classical criteria have in common that they provide resolution limits that are completely set by the functional form of the point spread function of the diffraction-limited imaging system. No mention is made of the presence of errors in the observed intensity distributions.

Nowadays, it has been recognized for some time that diffraction does not impose an absolute limit to the resolving power of an imaging system. When visual inspection is replaced by intensity measurement, knowledge of the point spread function of the imaging system makes it possible to attain unlimited resolution, provided that the intensity measurements are noise free. For, on this condition, numerically fitting a mathematical model of point spread functions to the observed composite intensity distribution would make it possible to resolve the two point sources exactly.

When we turn to spatial frequency domain, viewing the imaging system as a linear spatial filter transforming the light field at the object plane into the light field at the image plane, the work on superresolution [4–6] has shown that in spite of the bandlimiting filter characteristics of the aperture of the diffraction-limited imaging system, knowledge of the point spread function, together with some reasonable a priori information about the object (e.g., knowledge that the object is of finite size which implies that its spatial Fourier transform is analytical) makes it possible to reconstruct the object exactly by applying mathematical operations on these spatial frequencies that do pass the aperture.

To sum up, it may be stated that, in the absence of measurement errors, an imaging system with a known point spread function can *theoretically* attain as high a resolving power as desired. *In practice*, however, the intensity measurements are always disturbed by noise (non-systematic errors). Furthermore, the point spread function will rarely be known exactly, which means the introduction of systematic errors. It may be obvious that it will be these errors, both systematic and non-systematic, that prevent an infinite degree of resolution. This insight necessitates a reevaluation of

two-point resolution in order to establish a resolution criterion that, unlike classical resolution criteria, takes into account the presence of errors in the observed intensity distribution. Such a resolution criterion is proposed in this paper.

The theory developed in this paper is related to earlier work by VAN DEN BOS on resolution in model-fitting [8–10]. In references [8] and [9] the two-point resolution of fitting a nonlinearly parametric sum model of intensity point spread functions to intensity observations was considered for one-dimensional and two-dimensional imaging systems respectively. These results are only applicable to incoherent imaging. The present study generalizes the results and extends them to include models used to describe one-dimensional partially and fully coherent imaging, which are *no* sum models.

The main results may be summarized as follows. A mathematical model of amplitude point spread functions, which is assumed to underlie the intensity observations, is fitted in least-squares sense to these error corrupted observations with respect to the intensities and the locations of the two point sources. The locations are non-linearly present in the model to be fitted. Now, the two point sources have been defined as being resolved if and only if the model-fitting solutions for their locations are distinct. It has been shown that the Euclidean space of the errors, or equivalently that of the observations, can be divided into two regions, separated from each other by a hypersurface called the bifurcation set. In the one region two distinct solutions for the locations are found, while in the other region the solutions exactly coincide. Thus defined, the bifurcation set represents the error limit to resolution achievable by model-fitting. Furthermore, it has been found to be relatively easy to decide on which side of the bifurcation set a given set of observations is located. This offers the experimenter the possibility to determine beforehand whether or not the two point sources can be resolved from the available observations. Additionally, an operational procedure has been developed to compute the errors that have minimum energy (in the sense of their Euclidean norm) among all errors causing coincidence of the solutions. This minimum energy offers the experimenter a scalar criterion to determine to what extent coincidence of the solutions is to be expected for assumed error energy. Besides, it can be used to compare the resolving capabilities of different imaging systems.

This paper is organized as follows. In section 2 the model to be fitted to the intensity observations is described. As a model-fitting criterion the least-squares criterion is chosen, since it is most frequently used in practice. In section 3 the possible structures of the model-fitting criterion under influence of the errors are discussed. The structure of the criterion is decisive for the kind of solution. In section 4 the resolution discriminant and the bifurcation set are derived. Section 5 introduces the concept of

critical errors. In section 6 an illustrative numerical example is presented. In section 7 conclusions are drawn.

2. Fitting a Model to the Observed Intensity Distribution

Consider an object consisting of two point sources of light. Let this two-point object be imaged by a one-dimensional diffraction-limited imaging system. Assume that the two point sources A and B have equal intensity and that they are located at the positions $x = \beta_1$ and $x = \beta_2$ with respect to the optical axis, where x is the coordinate in the object plane. Due to the diffraction at the aperture of the imaging lens, each point source is not imaged as an image point, as predicted by geometrical optics, but as the Fraunhofer diffraction pattern of the aperture. Let $h(x')$ be the Fraunhofer amplitude diffraction pattern, i.e., *the amplitude point spread function (apsf)* of the imaging system concerned, where x' is the coordinate in the image plane. Then, the composite intensity distribution in the image plane produced by the two point sources together is given by:

$$g(x';\alpha,\beta_1',\beta_2') = \alpha[|h(x'-\beta_1')|^2 + |h(x'-\beta_2')|^2 + \\ +2\Re\{\gamma_{12}\cdot h(x'-\beta_1')\cdot h^*(x'-\beta_2')\}] \,, \tag{1}$$

where β_1' and β_2' are the locations of the geometrical image points of A and B respectively. $\Re$ is the real operator, * denotes the complex conjugate, γ_{12} is the complex degree of coherence [11] and α is the amplitude, which is proportional to the intensity of the point sources. It may be shown [11] that $|\gamma_{12}| \leq 1$. From now on it will be assumed that the *apsf* is real. Let $\gamma = \Re\{\gamma_{12}\}$. The value $\gamma = 0$ corresponds to mutually fully incoherent point sources, $\gamma = +1$, -1 imply completely cophasal and antiphasal coherent point sources respectively, whereas $|\gamma| < 1$ implies partially coherent point sources.

Suppose that N observations $w_1, \ldots, w_N$ have been made on $g(x;\alpha,\beta_1',\beta_2')$, defined by:

$$w_n = g_n(\alpha,\beta_1',\beta_2') + v_n \,, \quad n = 1,\ldots,N \tag{2}$$

with $g_n = g(x_n';\alpha,\beta_1',\beta_2')$, where the x_n' are exactly known values of x' (the measurement points). The v_n are the errors in the observations. Systematic errors are defined as the expectation $E[v_n]$ of the v_n, whereas non-systematic errors are defined as $v_n - E[v_n]$. The errors are assumed to be

small in comparison with the errorless observations. Now, in this paper, two-point resolution will be defined as the ability of the imaging system, including detection and digital computing facilities, to obtain two distinct solutions for the location parameters b_1 and b_2 of the model $g(x'_n; a, b_1, b_2)$ when the latter is fitted to the observations, with respect to $\mathbf{t} = (a, b_1, b_2)^T$. If a least-squares procedure is used to estimate the parameters, the model-fitting solutions for the amplitude and the locations are equal to the solution of the following minimization problem:

$$\underset{a,b_1,b_2}{\text{minimize}}\, J_2(a, b_1, b_2) = \sum_n d_n^2(a, b_1, b_2) \ , \tag{3}$$

where $J_2(a, b_1, b_2)$ is the least-squares criterion and the deviations $d_n(a, b_1, b_2)$ are defined by:

$$d_n(a, b_1, b_2) = w_n - g_n(a, b_1, b_2) \ . \tag{4}$$

The least-squares solution is the absolute minimum of $J_2(a, b_1, b_2)$. A necessary condition for a minimum is that it must be a stationary point of the criterion. Stationary points are defined as points where the gradient of the criterion is equal to zero. So, from *Eqs.* (1–4) it follows that the stationary points $(\hat{a}, \hat{b}_1, \hat{b}_2)$ of $J_2(a, b_1, b_2)$ satisfy the so-called normal equations given by:

$$\sum_n d_n(\hat{a}, \hat{b}_1, \hat{b}_2)[h_n^2(\hat{b}_1) + h_n^2(\hat{b}_2) + 2\gamma h_n(\hat{b}_1) h_n(\hat{b}_2)] = 0 \ , \tag{5}$$

$$\sum_n d_n(\hat{a}, \hat{b}_1, \hat{b}_2)[h_n(\hat{b}_1) h_n^{(1)}(\hat{b}) + \gamma h_n^{(1)}(\hat{b}_1) h_n(\hat{b}_2)] = 0 \ , \tag{6}$$

$$\sum_n d_n(\hat{a}, \hat{b}_2, \hat{b}_2)[h_n(\hat{b}_2) h_n^{(1)}(\hat{b}_2) + \gamma h_n(\hat{b}_1) h_n^{(1)}(\hat{b}_2)] = 0 \ , \tag{7}$$

where $h_n(\hat{b}_k) = h(x'_n - \hat{b}_k)$ and $h^{(\ell)}(\hat{b}_k)$ is the ℓ-th order derivative of $h_n(b_k)$ with respect to b_k evaluated at $\hat{b}_k$.

Next, suppose that a model describing the intensity distribution as if it was produced by *one* point source is fitted to the same observations. The least-squares criterion for this so-called *first order* model $a^* h^2(b)$ is given by:

$$J_1(a^*, b) = \sum_n d_{n,1}^2(a^*, b) \ , \tag{8}$$

with

$$d_{n,1}(a^*, b) = w_n - a^* h_n^2(b) \ . \tag{9}$$

The stationary points $(\hat{a}^*, \hat{b})$ of $J_1(a^*, b)$ with respect to a^* and b satisfy

$$\sum_n d_{n,1}(\hat{a}^*, \hat{b}) h_n^2(\hat{b}) = 0 \ , \tag{10}$$

$$\sum_n d_{n,1}(\hat{a}^*, \hat{b}) h_n(\hat{b}) h_n^{(1)}(\hat{b}) = 0 \ . \tag{11}$$

From the *Eqs.* (5–7) and (10–11), it can be concluded that all points $(\hat{a}, \hat{b}_1, \hat{b}_2)$, with $\hat{b}_1 = \hat{b}_2 = \hat{b}$ and $\hat{a} = \hat{a}^*/(2 + 2\gamma)$ are stationary points of $J_2(a, b_1, b_2)$. In this way, a stationary point of J_1 generates a stationary point of J_2. So, the stationary points of J_2 can be divided into a group that contains the stationary points for which the location parameters b_1 and b_2 are distinct and a group that contains the stationary points characterized by *exactly* coinciding location parameters. Stationary points belonging to the first group and the last group will be called *second order* and *first order stationary points* respectively.

It may be obvious that the structure of the criterion J_2 is decisive for the kind of solution and by that, ultimately, for the resolvability of the model-fitting solutions for the location parameters. In the next section we will analyze this structure and the influence of the errors upon it.

3. The Structure of the Least-Squares Criterion

3.1. Stationary Points of the Criterion

For the purpose of this paper the location parameters β_1' and β_2' are considered to be very close, so that difficulties with resolution may be expected. For the moment, let's consider errorless observations. Then, for reasons of symmetry, it will be clear that the least-squares criterion J_2 has two closely spaced absolute minima $(\alpha, \beta_1', \beta_2')$ and $(\alpha, \beta_2', \beta_1')$, at which the criterion is equal to zero. Since the location parameters are close to each other, a first order model will also fit quite well to the observations in the sense of least-squares. So, intuitively, least-squares fitting of the first order model will result in a so-called *first order solution* $(\hat{a}^*, \hat{b})$, , where $\hat{a}^*$ and $\hat{b}$ will approximately be equal to $(2 + 2\gamma)\alpha$ and the average of β_1' and β_2' respectively. Now, as we saw in the preceding section, this first order solution $(\hat{a}^*, \hat{b})$ generates a stationary point $(\hat{a}, \hat{b}, \hat{b})$, with $\hat{a} = \hat{a}^*/(2 + 2\gamma)$, of J_2. So, in between the two (second order) minima, the criterion J_2 will have an extra (first order) stationary point. This first order stationary point is a one-saddle point. This can be seen as follows: since the criterion $J_1(a^*, b)$ is the intersection of the plane $b_1 = b_2$ and the (appropriately scaled) criterion $J_2(a, b_1, b_2)$ and $(\hat{a}^*, \hat{b})$ is the minimum of $J_1(a^*, b)$, the first order

stationary point must be a minimum in all directions but $b_1 - b_2$. Travelling on the criterion in parameter space in the latter direction from the absolute minimum $(\alpha, \beta_1', \beta_2')$ through the first order stationary point $(\hat{a}, \hat{b}, \hat{b})$ to the absolute minimum $(\alpha, \beta_2', \beta_1')$ on the other side of the plane $b_1 = b_2$, means going uphill from $(\alpha, \beta_1', \beta_2')$ to $(\hat{a}, \hat{b}, \hat{b})$ and downhill from there. In conclusion, the first order stationary point is a maximum in the direction $b_1 - b_2$, but a minimum in all other directions. The minima are close to the saddle point, since the location parameters were assumed close.

A remarkable phenomenon appearing from simulation experiments is the fact that errors in the observations may change the structure of the criterion in such a way that the two minima and the saddle point merge into one minimum at the first order stationary point $(\hat{a}, \hat{b}, \hat{b})$. Then resolution of the two point sources is no longer possible, since the least-squares solutions for the location parameters exactly coincide. In the next subsection this coincidence phenomenon will be studied by investigating the influence of the errors on both the nature of the first order stationary point and the behavior of the criterion around it.

3.2. Taylor Expansion of the Criterion

The least-squares criterion J_2 is Taylor expanded around the first order stationary point $\hat{\mathbf{t}} = (\hat{a}, \hat{b}, \hat{b})^T$. The constant term is equal to $J_2(\hat{a}, \hat{b}, \hat{b})$, or equivalently $J_1(\hat{a}^*, \hat{b})$. The linear terms are absent, since the origin of the Taylor expansion is a stationary point. The quadratic terms are given by:

$$\frac{1}{2}\, {}^1\hat{\mathbf{t}}\mathbf{H}_2\, {}^1\hat{\mathbf{t}}^T \,, \tag{12}$$

where ${}^1\mathbf{t} = \mathbf{t} - \hat{\mathbf{t}}$. $\mathbf{H}_2$ is the 3×3 Hessian matrix of J_2 with respect to $\mathbf{t} = (a, b_1, b_2)^T$ at $\hat{\mathbf{t}} = (\hat{a}, \hat{b}, \hat{b})^T$. The elements of H_2 are defined as:

$$(H_2)_{ij} = \frac{\partial^2 J_2(\hat{\mathbf{t}})}{\partial t_i \partial t_j} \,, \quad i, j = 1, \ldots, 3 \,. \tag{13}$$

The eigenvalues of the Hessian matrix determine the nature of the stationary point. In order to get more insight in the eigenstructure of $\mathbf{H}_2$, the parameters are subsequently linearly transformed into

$${}^2a = {}^1a(2 + 2\gamma) \,, \quad {}^2b_1 = \frac{1}{2}({}^1b_1 + {}^1\, b_2) \,, \quad {}^2b_2 = \frac{1}{2}({}^1b_1 - {}^1\, b_2) \,. \tag{14}$$

It may be shown that in these coordinates the Taylor expansion becomes

$$J_2({}^2\mathbf{t}) = J_2(\hat{\mathbf{t}}) + \frac{1}{2}\, {}^2\mathbf{t}^T \mathrm{diag}\, (\mathbf{H}_1\ \rho_2)\, {}^2\mathbf{t} + O({}^2\mathbf{t}^3) \,, \tag{15}$$

where $\mathbf{H}_1$ is the 2×2 Hessian matrix of $J_1(a^*, b)$ at the first order solution $(\hat{a}^*, \hat{b})$. The symbol $O(\mathbf{t}^n)$ is the order symbol of Landau. It represents all terms of degree n and higher. The element ρ_2 is given by:

$$\rho_2 = -4\hat{a}^* \left[\frac{(1-\gamma)}{(1+\gamma)} \sum_n d_{n,1}(\hat{a}^*, \hat{b}) (h_n^{(1)}(\hat{b}))^2 + \sum_n d_{n,1}(\hat{a}^*, \hat{b}) h_n(\hat{b}) h_n^{(2)}(\hat{b}) \right] . \tag{16}$$

By the assumption that the first order solution $(\hat{a}^*, \hat{b})$ always exists, i.e., that the first order model fits well to the observations, independent of the particular realization of the errors, the Hessian matrix $\mathbf{H}_1$ is always positive definite. Then all eigenvalues of $\mathbf{H}_1$ are positive. Hence, the nature of the stationary point $\hat{\mathbf{t}}$ is completely set by the sign of ρ_2, since this sign corresponds to that of the remaining eigenvalue of $\mathbf{H}_2$. The point $\hat{\mathbf{t}}$ is a one-saddle, a non-degenerate minimum or a degenerate minimum as ρ_2 is negative, positive or zero respectively.

The analysis of the higher order terms of the Taylor expansion can be simplified drastically by using the catastrophe theory. Catastrophe theory [12] is concerned with the structural change of a parametric function under influence of its parameters. It tells us that a structural change of the function is always preceded by a degeneracy of one of its stationary points. The theory also shows that the independent variables of the function can be split into *essential* and *inessential* variables. The essential variables correspond to the directions in which degeneracy occurs. The number of essential variables is equal to the number of possible vanishing eigenvalues of the Hessian matrix of the function at the stationary point that may become degenerate. In order to analyze structural change, the parametric function can be replaced by a Taylor expansion in the essential variables, around the latter stationary point. Terms in the inessential variables do not play a role at all in the structural change. According to catastrophe theory the global structure of a parametric function with only one essential variable is completely set by its Taylor expansion up to the degree of which coefficient cannot vanish under the influence of its parameters. In most practical applications, including the one considered in this paper, the required degree of the Taylor polynomial is very low.

The parametric function studied in this paper is the least-squares criterion as a function of the model parameters. Its parameters are the errors in the observations. *Eq.* (15) shows that only the eigenvalue corresponding to the coordinate ${}^2b_2$ may vanish. So, ${}^2b_2$ is the only essential variable. Since $\mathbf{H}_1$ is symmetric, it can be diagonalized by a not specified non-singular linear transformation of the coordinates ${}^2a$ and ${}^2b_1$ into ${}^3a$ and

${}^3b_1$. Although this transformation does not effect the essential coordinate ${}^2b_2$, the superscript ^{3} is also used for ${}^3b_2 = {}^2b_2$. Now, the quadratic terms of the Taylor expansion can be described as:

$$\frac{1}{2}\lambda_1' \, {}^3a^2 + \frac{1}{2}\lambda_2' \, {}^3b_1^2 + \frac{1}{2}\rho_2 \, {}^3b_2^2 \, , \tag{17}$$

where λ_1' and λ_2' are the eigenvalues of $\mathbf{H}_1$. Next, the cubic and quartic terms of the Taylor expansion are considered. The coefficient of ${}^3b_2^3$ happens to be equal to zero, because of the symmetry of the model. The cubic cross terms containing the inessential coordinates $({}^3a, {}^3b_1)$ are removed by applying a procedure described by [12]. First, all cubic terms in which ${}^3a$ appears are collected. Let the sum of these terms be $\lambda_1' \, {}^3aQ_1 \updownarrow Q_1$ is then homogeneously quadratic. The sum of $\frac{1}{2}\lambda_1' \, {}^3a^2$ and these terms may be written as:

$$\frac{1}{2}\lambda_1'({}^3a + Q_1)^2 - \frac{1}{2}\lambda_1' Q_1^2 \, , \tag{18}$$

where the last term is homogeneously quartic. This procedure is next applied to all remaining cubic terms containing ${}^3b_1$, which can be described as $\lambda_2' \, {}^3b_1 Q_2$. Subsequently, the coordinates ${}^3a$ and ${}^3b_1$ are nonlinearly transformed into the following curvilinear coordinates:

$${}^4a = {}^3a + Q_1 \, , \quad {}^4b_1 = {}^3b_1 + Q_2. \tag{19}$$

The coordinate ${}^3b_2$ is again not affected by this transformation: ${}^4b_2 = {}^3b_2$. Substituting *Eq.* (19) in the Taylor expansion removes all cubic terms. Notice that the quartic terms resulting from this procedure are described by:

$$-\frac{1}{2}\sum_m \lambda_m' Q_m^2 \, , \quad m = 1, 2. \tag{20}$$

Also notice that the quartic terms (20) and those already present in the original expansion are still in the old coordinates. To change this, *Eq.* (19) is used to express the old coordinates as a power series in the new ones and the result is substituted in the quartic and higher order terms. Now, the coefficient τ of ${}^4b_2^4$ described by:

$$\tau = \nu - \frac{1}{2}(\eta_1^2/\lambda_1') - \frac{1}{2}(\eta_2^2/\lambda_2') \, , \tag{21}$$

where ν, η_1 and η_2 are the coefficients of ${}^3b_2^4$, ${}^3a \, {}^3b_2^2$ and ${}^3b_1 \, {}^3b_2^2$ respectively in the original expansion. Without derivation, the expression for ν is given

by:

$$\nu = \hat{a}^{*2}\left[\frac{(1-\gamma)^2}{(1+\gamma)^2}H_{040} + H_{202} + 2\frac{(1-\gamma)}{(1+\gamma)}H_{121}\right] - \frac{1}{6}\hat{a}^*\left[3D_{00200} + 4\frac{(1-\gamma)}{(1+\gamma)}D_{01010} + D_{10001}\right] , \tag{22}$$

where

$$H_{klm} = \sum_n h_n^k (h_n^{(1)})^l (h_n^{(2)})^m \tag{23}$$

and

$$D_{ijklm} = \sum_n d_n h_n^i (h_n^{(1)})^j (h_n^{(2)})^k (h_n^{(3)})^l (h_n^{(4)})^m , \tag{24}$$

with $h_n = h_n(\hat{b})$, $h_n^{(l)} = h_n^{(l)}(\hat{b})$ and $d_n = d_{n,1}(\hat{a}^*, \hat{b})$. Also without derivation, the expressions for η_1 and η_2 are given by:

$$\eta_1 = 2\hat{a}^*\left[H_{301} + \frac{(1-\gamma)}{(1+\gamma)}H_{220}\right] + 2\left[\frac{1}{(1+\gamma)}D_{02000} - D_{10100}\right] , \tag{25}$$

and

$$\eta_2 = \Phi - \left(2\hat{a}^*\frac{H_{310}}{H_{400}}\right)\eta_1 , \tag{26}$$

where

$$\Phi = 4\hat{a}^{*2}\left[H_{211} + \frac{(1-\gamma)}{(1+\gamma)}H_{130}\right] - 2\hat{a}^*\left[\frac{1}{2}D_{00200} + D_{10010} + \frac{(3-\gamma)}{(1+\gamma)}D_{01100}\right] . \tag{27}$$

Next, a similar procedure could be used to remove all quartic terms in which ${}^4a$ and ${}^4b_1$ appear and so on. Notice that such a procedure would not further affect the coefficient τ. So, the resulting Taylor expansion would be of the form

$$\lambda_1(a')^2 + \lambda_2(b_1')^2 + \rho(b_2')^2 + \tau(b_2')^4 + O(\mathbf{t}'^5) , \tag{28}$$

where $\lambda_1 = \frac{1}{2}\lambda_1'$, $\lambda_2 = \frac{1}{2}\lambda_2'$ and $\rho = \frac{1}{2}\rho_2$. The new coordinates a' and b_1' have been obtained by applying the procedure described above to the quartic terms and $b_2' = {}^4b_2 = b_1 - b_2$.

In order to investigate whether or not the Taylor expansion up to the fourth degree in the essential variable is sufficient, let's consider the possible signs of the coefficients ρ and τ. As we saw before, for errorless observations the first order stationary point will be a saddle point, so ρ will be negative. However, considering *Eq.* (16), it can easily be understood that already relatively small errors in the observations may make ρ positive and thus the stationary point a minimum. This can be seen as follows. The *apsf h* and its first and second order derivatives are relatively independent of the errors, since the first order solution $(\hat{a}^*, \hat{b})$ will hardly change under influence of the errors. The derivations d_n on the other hand, depend strongly on the particular realization of the errors; relatively small errors are required to change their sign and by that the sign of ρ. From *Eqs.* (21-27), it follows that τ consists of a sum of terms that do not contain the deviations d_n and a sum of terms that do contain these deviations. The first sum of terms only depends on the first order solution $(\hat{a}^*, \hat{b})$ and is therefore relatively independent of the particular realization of the errors. It may be shown that this sum of terms is positive. The second sum consists of terms of order $|\mathbf{d}|_1$ and $|\mathbf{d}|_2$, where $|\mathbf{d}|_p$ denotes the ℓ_p norm of the vector $\mathbf{d} = (d_1, d_2, \ldots d_N)^T$. These terms will be negligible compared to the first sum of terms, since, by assumption, the first order model fits well to the observations. So, τ consists of a nearly constant positive term and terms of order $|\mathbf{d}|_1$ and $|\mathbf{d}|_2$ and will therefore always be positive. Consequently, the expansion (28) is sufficient to describe the possible structures of the least-squares criterion. This means that the study of the least-squares criterion under influence of the errors can be replaced by a study of the following quartic Taylor polynomial in the essential variable b_2':

$$\rho(b_2')^2 + \tau(b_2')^4 \ , \tag{29}$$

where the constant term has been omitted, since it does not influence the structure.

4. Resolution Discriminant and Bifurcation Set

The structure of the quartic Taylor polynomial (29) is fully established by the number and the nature of its stationary points. From *Eq.* (29) it follows, that the polynomial is always stationary at the point $b_2' = b_1 - b_2 = 0$. This stationary point is a maximum, a non-degenerate minimum or a degenerate minimum, if ρ is negative, positive or equal to zero respectively. These correspond to a one-saddle point, a non-degenerate minimum and a degenerate minimum of the least-squares criterion J_2 at the origin of expansion (29), i.e., at the first order stationary point $(\hat{a}, \hat{b}, \hat{b})$. Further analysis of *Eq.* (29) shows that the criterion will have two additional stationary

points if $\frac{\rho}{\tau}$ is negative. Given the positiveness of τ, it may be shown that these two additional stationary points are minima. To sum up, it can be concluded that if ρ is negative the criterion has locally two (second order) minima and one (first order) saddle point. This structure corresponds to the one for errorless observations, found on intuitive grounds in Section 3. In this case one will find two distinct least-squares solutions for the location parameters, so that resolution is guaranteed. If, however, under influence of the errors in the observations, ρ becomes positive, the first order stationary point changes from a saddle point into a minimum, while the two original minima vanish through coincidence with this first order stationary point. In this case the first order stationary point, at which $b_1 = b_2$, has become the absolute minimum of the criterion. Then the least-squares solutions for the location parameters will exactly coincide so that resolution is no longer possible. Simulation experiments have confirmed the existence of the two structures described above and up till now no local structures have been found different from these. Since the sign of ρ is decisive for the resolvability of the two point sources, ρ is called the *resolution discriminant*. To test a given set of observations with respect to resolvability, only ρ has to be computed. From *Eq.* (16) it follows that, for this purpose, fitting of a first order model suffices. Notice, that the resolution test is applicable to any value of γ, except for $\gamma = -1$.

In the Euclidean N space of the errors the subset of all points for which $\rho = 0$ separates both structures described above. The errors belonging to this subset and the corresponding first order solution $(\hat{a}^*, \hat{b})$ must satisfy the normal equations for $J_1(a^*, b)$ with respect to (a^*, b), and, in addition, the equation $\rho = 0$. This is a set of three equations in $N+2$ variables, or, equivalently one equation in the N errors. This equation constitutes a subspace of the Euclidean N space of errors of codimension 1, which is called a *bifurcation set* in catastrophe theory. It divides the error space into a region, including the origin, where the point sources can be resolved, and a region where resolution is impossible. Thus defined, the bifurcation set represents the error limit to resolution achievable by model-fitting. Since the first order solution will hardly change under influence of the errors and the condition $\rho = 0$ is linear in the errors, the bifurcation set is approximately a hyperplane. Simulation experiments have affirmed the high accuracy of this approximation. Notice that the error space can easily be transformed into that of the observations by a simple translation.

When the functional form of the *apsf* is not identical to the one underlying the observations, modelling errors are introduced. These systematic errors may, possibly in combination with the additive errors v_n, also cause coincidence of the model-fitting solutions. This means that exact observations, i.e. all $v_n = 0$, do not guarantee resolution. It may be shown

that fitting a wrong model causes a shift of the bifurcation set in the error space. Nevertheless, the resolution test derived in this section applies to fitting a wrong model, too.

5. Critical Errors

Among all errors belonging to the bifurcation set, those that have minimum energy, in the sense of their Euclidean norm, are called the *critical errors*. Errors having smaller energy than the critical energy cannot cause coincidence of the solutions, errors having higher energy may do so. Knowledge of the critical energy offers the experimenter the possibility to find out to what extent resolution is to be expected for assumed error energy. Furthermore, it can be used to test beforehand whether or not the experimental set-up is good enough to meet the demands. Additionally, it can be used to compare the resolving capabilities of different imaging systems used for the same purpose. By computing the critical energy as a function of the distance between the two closely located point sources, a relation can be established between this distance and the maximum allowable energy of the errors.

The critical errors can exactly be found by minimizing the quadratic sum of all errors with respect to $(v_1, \ldots, v_N, a^*, b)$ subject to equality constraints (10), (11) and $\rho = 0$. This nonlinear minimization problem under nonlinear equality constraints can be solved by nonlinear programming methods and software. Since, however, this is an enormous, time-consuming task, an alternative procedure will be proposed in this section, providing an approximate expression for the critical errors that has proved to be very accurate. Compared to numerically finding a solution, this approximation means a drastic simplification.

Notice, that the minimization problem in consideration is a Lagrange problem whose solution must be under the stationary points $(\hat{v}_1, \ldots, \hat{v}_N, \hat{\mu}, \hat{\kappa}, \hat{\xi}, \hat{a}^*, \hat{b})$ of the Lagrange function

$$L = \sum_N v_n^2 + \mu \sum_n d_n h_n^2 + \kappa \sum_n d_n h_n h_n^{(1)} + \xi \left[\frac{(1-\gamma)}{(1+\gamma)} \sum_n d_n (h_n^{(1)})^2 + \sum_n d_n h_n h_n^{(2)} \right], \tag{30}$$

where μ, κ and ξ are the Lagrange multipliers. Here and in what follows h_n, $h_n^{(\ell)}$ and d_n are defined as in *Eqs.* (22-24). Differentiating L with respect

to the v_n and equating the result to zero yields

$$\hat{v}_n = -\frac{1}{2}\mu h_n^2 - \frac{1}{2}\kappa h_n h_n^{(1)} - \frac{1}{2}\xi\left[\frac{(1-\gamma)}{(1+\gamma)}(h_n^{1)})^2 + h_n h_n^{(2)}\right] . \tag{31}$$

Subsequent differentiation of L with respect to the Lagrange multipliers and equating the result to zero yields, of course, the *Eqs.* (10), (11) and $\rho = 0$. Substituting (31) for the errors in these equations, produces equations identical to the normal equations for minimizing the deviations

$$\hat{d}_n = g_n(\alpha, \beta_1', \beta_2') + \hat{v}_n - \hat{a}^* h_n^2(\hat{b}) \tag{32}$$

in least-squares sense with respect to the Lagrange multipliers, which are, by *Eq.* (31), linearly present in the v_n. If $\{g_n(\alpha, \beta_1', \beta_2') - a^* h^2(b)\}$ is Taylor expanded about $(\hat{a}^*, \hat{b})$ in powers of $((2+2\gamma)\alpha - \hat{a}^*)$, $(\beta_1' - \hat{b})$ and $(\beta_2' - \hat{b})$, it can easily be shown that this least-squares minimization makes the deviations quadratic in the elements of $((2+2\gamma)\alpha - \hat{a}^*)$, $(\beta_1' - \hat{b})$ and $(\beta_2' - \hat{b})$. So, it can be concluded that for critical errors $\hat{d}_n \approx 0$. Hence, it follows from *Eq.* (32) that the critical errors are approximately equal to the additive inverse of the deviations of the exact observations and the first order model best fitting to the observations corrupted by the critical errors. Finally, differentiating L with respect to (a^*, b) and equating the result to zero shows that $(\hat{a}^*, \hat{b})$ must be approximately, but very accurately, equal to the solution that would have been obtained in the absence of errors. This means that the critical errors are approximately equal to the additive inverse of the deviations of the best fitting first order model to exact observations. The proof will be left out here, since it requires more or less the same consecutive steps as described in [10]. In this reference an approximate solution is derived for the critical errors associated with coincidence of model-fitting solutions for closely located nonlinear parameters of weighted-sum models, which is a topic closely related to the one described in this section.

Since the bifurcation set is approximately a hyperplane, it can be shown that an arbitrary error signal $\mathbf{v} = (v_1, \ldots, v_N)^T$ is not on the same side of the bifurcation set as the origin if [13]

$$\mathbf{v} \cdot \mathbf{e} \geq \mathbf{e} \cdot \mathbf{e} , \tag{33}$$

where $\mathbf{v} \cdot \mathbf{e}$ is the inner product of $\mathbf{v}$ and the critical error vector $\mathbf{e} = (e_1, \ldots, e_N)^T$. The inner product $\mathbf{e} \cdot \mathbf{e}$ is equal to the critical energy. Rule (33) can be used to find out whether or not for arbitrary, given errors coincidence of the solutions for the locations occurs, or for statistical errors, how probable this coincidence is [13].

6. Numerical Example

The purpose of this numerical example is to give an impression of the magnitude of the critical energy as a function of γ. For this purpose the critical energy is computed for the model

$$\alpha \left[h^2(x'_n - \frac{1}{2}\Delta\beta') + h^2(x'_n + \frac{1}{2}\Delta\beta') + 2\gamma h(x'_n - \frac{1}{2}\Delta\beta')h(x'_n + \frac{1}{2}\Delta\beta') \right] , \tag{34}$$

where $\alpha = 1$ and $h(x')$ is a gaussian shaped *apsf*: $h(x') = \exp\{-(1/2)x'^2\}$. This is done for $-1 < \gamma \leq 1$. The measurement points are described by $x'_n = -0.01 + (n - 11) \times 0.4$, $n = 1, \ldots, 21$. *Fig. 1* shows the square root of the critical energy as a fraction of $\hat{a}^*$ as a function of γ, respectively for $\Delta\beta' = 0.02$, $\Delta\beta' = 0.04$ and $\Delta\beta' = 0.08$. The results show that coincidence may already occur for relatively low error energy. They also show that resolution decreases with decreasing $\Delta\beta'$, as might be expected. Moreover, *Fig. 1* demonstrates that resolution decreases with increasing γ; undermining the resolution requires for fully cophasal coherent point sources errors that are, roughly, three times smaller than for fully incoherent point sources.

7. Discussion and Conclusions

An earlier presented theory with respect to the concept of optical two-point resolution, which was applicable to strictly incoherent imaging, has been extended to include coherent and partially coherent imaging as well. A new resolution criterion has been proposed that, unlike classical resolution criteria, takes account of the errors in the observed intensity distributions and what's more, it's even based on these errors. The criterion states that which errors make resolution impossible and which errors do not, depends on the distance between the two point sources, the intensity of the point sources, the choice of measurement points, the degree of coherence and the point spread function of the imaging system. A resolution test has been derived that offers the experimenter the possibility to determine beforehand whether or not the two point sources can be resolved from error-corrupted observations. Also a procedure has been presented to compute the errors having minimum energy among all errors undermining resolution. This minimum energy provides the experimenter with a scalar criterion to determine to what extent resolution is to be expected for the point spread

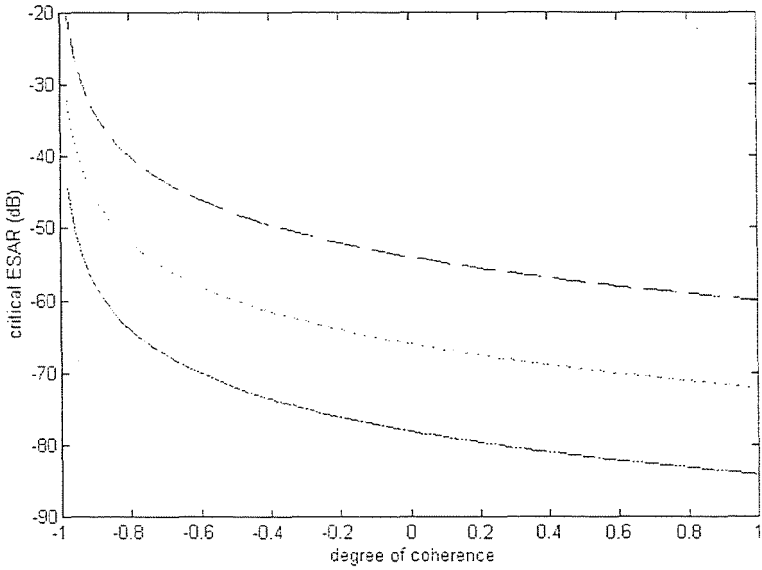

Fig. 1. Critical Error-to-Signal Amplitude Ratio (ESAR) for model (34) as a function of γ, for $\Delta\beta' = 0.02$ (solid), $\Delta\beta' = 0.04$ (dotted) and $\Delta\beta' = 0.08$ (dashed)

function assumed to underlie the observations and, perhaps a guess of, the error energy. Additionally, it can be used to compare the resolving capabilities of different instrumental set-ups used for the same purpose.

Finally, the author has found that the theory presented here is not confined to strictly one-dimensional imaging systems. For two-dimensional imaging systems a so called *resolution matrix* can be derived of which the smallest eigenvalue plays the role of resolution discriminant. These results will be reported later.

References

1. Lord RAYLEIGH, title of: STRUTT, J. W. (1899): *Scientific Papers*, Vol. 1, p. 415, Cambridge University Press, Cambridge,UK.
2. RAMSAY, B. P. – CLEVELAND, E. L. – KOPPIUS, O. T. (1941): *Journal of the Optical Society of America*, Vol. 31 (8), p. 26.
3. SPARROW, G. (1916): On Spectroscopic Resolving Power, *Astrophys. J.*, Vol. 44, p. 76.
4. HARRIS, J. L. (1964): Diffraction and Resolving Power, *Journal of the Optical Society of America*, Vol. 54 (7), pp. 931–936.
5. RUSHFORTH, C. K. – HARRIS, R. W. (1968): Restoration, Resolution and Noise, *Journal of the Optical Society of America*, Vol. 58 (4), pp. 539–545.
6. FRIEDEN, B. R. (1967): Band-Unlimited Reconstruction of Optical Objects and Spectra, *Journal of the Optical Society of America*, Vol. 57 (8), pp. 1013–1019.

7. HELMSTROM, C. W. (1970): Resolvability of Objects from the Standpoint of Statistical Parameter Estimation, *Journal of the Optical Society of America*, Vol. 60 (5), pp. 659–666.
8. VAN DEN BOS, A. (1987): Optical Resolution: an Analysis Based on Catastrophe Theory, *Journal of the Optical Society of America*, Vol. 4, pp. 1402–1406.
9. VAN DEN BOS, A. (1992): Ultimate Resolution: a Mathematical Framework, *Ultramicroscopy*, Vol. 47, pp. 298–306.
10. VAN DEN BOS, A. – SWARTE, J. H. (1993a): Resolvability of the Parameters of Multiexponentials and Other Sum Models, *IEEE Transactions on Signal Processing*, Vol. 41 (1), pp. 313–322.
11. BORN, M. – WOLF, E. (1980): Principles of Optics, 6th edn., Pergamon, New York.
12. POSTON, T. – STEWART, I. N. (1978): Catastrophe Theory and Its Applications, Pitman, London.
13. VAN DEN BOS, A. (1993b): Critical Errors Associated with Parameter Resolvability, *Computer Physics Communications*, Vol. 76, pp. 184–190.