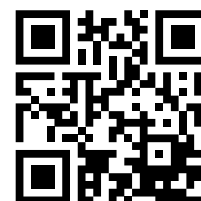


[16] Statistical estimation of the probability of the correct substance detection in FTIR spectroscopy

A.N. Morozov¹, I.V. Kochikov², A.V. Novgorodskaya¹, A.A. Sologub¹, I.L. Fufurin¹

¹Bauman Moscow State Technical University, Moscow, Russia

²Research Computer Center M.V. Lomonosov Moscow State University, Moscow, Russia



Abstract

In the present paper the problem of substance identification in FTIR spectroscopy is considered. The spectral library hitlist search is chosen as the main tactic. In the paper the Pearson correlation coefficient as the similarity criteria between two spectra is suggested. The situation when one of the measured spectra has an additive delta-correlated white noise component with Gaussian distribution is considered. In this case the probability density of correlation coefficient is found. The definition of correct detection probability is formulated and theoretical expression is found. In addition, we consider the threshold correlation coefficient search algorithm, which allows to find a threshold value providing the required correct detection. Computational experiments have shown the applicability of the method.

Keywords: SPECTROSCOPY, IDENTIFICATION, PROBABILITY OF CORRECT DETECTION, THE THRESHOLD VALUE OF DETECTION.

Citation: MOROZOV A.N. STATISTICAL ESTIMATION OF THE PROBABILITY OF THE CORRECT SUBSTANCE DETECTION IN FTIR SPECTROSCOPY / MOROZOV A.N., KOCHIKOV I.V., NOVGORODSKAYA A.V., SOLOGUB A.A., FUFURIN I.L. // COMPUTER OPTICS. – 2015. – VOL. 39(4). – P. 614-621

Introduction

The problem of remote control of air pollution, as well as air chemical control of industrial and other important sites is sufficiently relevant to date. The reason for this is the increasing growth of the number of pollutants, as well as growth of industries whose emissions cannot be defined in direct contact. As a result, for more than two centuries technologies and methods of contactless identification are being developed. One of the characteristic properties strictly individual for each substance is its optical spectrum; it can be used as a kind of ID. One of the common methods for obtaining spectra of substances is an infrared (IR) spectrometry. The choice of IR range is caused by the fact that the radiation in this range excites vibration motion of molecules or their fragments, resulting in the weakening of the intensity only at frequencies of molecular vibrations, so the spectrum of each substance is unique, and spectral lines are selective and pronounced.

IR spectrometers are divided into diffraction and IR Fourier spectrometers based on interferometers. In this work a second class of devices will be considered; however, the methods described can be used for diffraction method also. Fourier spectrometers special feature is their ability to work with lower intensities

[1] than in diffraction devices, which allows to detect their self-radiation spectra.

An important problem arising in the development of infrared Fourier spectrometer is the choice of method for identifying a substance from the reconstructed spectrum. Of course, chemical composition may be analyzed by the presence of some specific bands in the spectrum, but more reliable and accurate method is consistent comparison of the obtained spectrum with the spectra of already known substances stored in the database of reference spectra data. It is obvious that identifiable signal may contain considerable noise, and it even further complicates the situation. At present we know many different methods and tools for solving problems of recognition, such as: trained neural networks [2], the use of indistinct logic [3], the methods of PCA (principal component analysis) [4], as well as introduction of a variety of measures of similarity between the data [5-11]. Major work on search and comparison of spectra of substances in the database have been made by Clerc [12] and Luinge [13]. Also a fundamental work on calculation of similarity measures in the database has been done in [14]. One of the possible measures of similarity is Pearson statistical coefficient of correlation [15] (hereinafter, we shall assume that the spectra are presented as numerical vectors of finite length):

$$r = \frac{(\bar{x} - \bar{x})(\bar{y} - \bar{y})}{\|\bar{x} - \bar{x}\| \|\bar{y} - \bar{y}\|}, \quad (1)$$

where \bar{x}, \bar{y} are vectors of compared spectra, \bar{x} is the arithmetic average of the components of the vector \bar{x} ,

$\|\bar{x}\|$ is Euclidian vector norm.

Here and further the expression $\bar{x} - \bar{x}$ means subtraction of the same scalar from all the components.

It is known that $r \in [-1, 1]$, while equality to unit can only be achieved when signals are linearly dependent, that in the case of spectra allows to speak about their identity. The less "similar" are spectra, the closer is the value of (1) to zero. Note that when r (1) approaches unity, spectra difference norm approaches the global minimum. In [16-22] a method is described for identifying based on correlation coefficient: first, the value of (1) is calculated for the tested spectrum with all the spectra in the database, then the substance is selected with the highest correlation coefficient, and if (1) exceeds a predetermined empirical threshold, the substance is considered identified.

Probabilistic characteristics of the correlation coefficient

Fisher and Kenney [23, 24] obtained the exact theoretical expression for the function of probability density of Pearson correlation coefficient when two random variables are compared with two-dimensional Gaussian distribution with a certain correlation. A significant limitation is that both vectors must be random. In turn, when identifying on the database in IR Fourier spectrometry reference spectra can be considered exactly known because they were obtained in laboratory studies by multiple averaging. It is also impossible to know in advance a probabilistic correlation coefficient without a set of multiple statistics.

Suppose that the analyzed vector contains white noise with a normal distribution, then the resulting spectrum will have the same noise characteristics due to the properties of the Fourier transformation. That is, the analyzed spectrum can be represented as:

$$\bar{x} = \bar{\tau}^* + \bar{\xi}, \quad \bar{y} = \bar{\tau}^*, \quad (2)$$

where $\bar{\tau}^*$ is the vector of the reference spectrum, $\bar{\xi}$ is a noise vector, each component of which has a normal distribution with known characteristics $\mathcal{N}(0, \sigma_{\xi}^2)$.

Let us denote by r^j the correlation coefficient of spectrum \bar{x} with a substance from the database with number $j = 1, \dots, M$, where M is the number of substances in the database. Let us consider that the database includes a substance with $\bar{\tau}^*$ spectrum, and its number is j^* .

We will get an explicit expression for correlation coefficient in the case where \bar{x} spectrum is compared with $\bar{\tau}^*$ spectrum. Having substituted (2) into (3), we will get:

$$r^{j^*} = \frac{\hat{\sigma}_{\tau^*} + \sigma_{\xi} r_{\xi}^{j^*}}{\sqrt{\hat{\sigma}_{\tau^*}^2 + \delta^2 + 2\hat{\sigma}_{\tau^*} \sigma_{\xi} r_{\xi}^{j^*}}}, \quad (3)$$

where $\hat{\sigma}_{\tau^*}^2 = \frac{\|\bar{\tau}^* - \bar{\tau}^*\|^2}{N}$ is a mean-square deviation of the $\bar{\tau}^*$ spectrum, characterizing its intensity.

$$r_{\xi}^{j^*} = \frac{\bar{\xi}^T (\bar{\tau}^* - \bar{\tau}^*)}{N \sigma_{\xi} \hat{\sigma}_{\tau^*}} \quad (4)$$

is a coefficient of pure noise correlation with the reference spectrum $\bar{\tau}^*$,

$$\delta^2 = \frac{\|\bar{\xi} - \bar{\xi}\|^2}{N} \quad (5)$$

is mean square deviation of noise,

N is the number of points in the experimental spectrum.

It is seen that r^{j^*} is a function of two random values (4) and (5), having a normal and chi-squared distribution respectively. Accurate analysis of function (3) in this case is presented in [25]. However, we can use the fact that in the considered spectra the number N is sufficiently large (usually from 200 to 800), in order to examine the expression (5) as a point estimate of dispersion σ_{ξ}^2 .

Note that in the denominator (3) δ^2 and the quantity $2\hat{\sigma}_{\tau^*} \sigma_{\xi} r_{\xi}^{j^*}$ are added. If we assume $\delta^2 = \sigma_{\xi}^2$, than an error of such an assessment would be

$$\frac{(2N-2)\sigma_{\xi}^4}{N^2} [26],$$

and dispersion of the value $2\hat{\sigma}_{\tau^*} \sigma_{\xi} r_{\xi}^{j^*}$ can be obtained by the theorem on the sum of random variables:

$$\frac{4\sigma_{\xi}^2 \hat{\sigma}_{\tau^*}^2}{N} \cdot \text{If the condition} \quad (6)$$

$$\frac{(2N-2)\sigma_{\xi}^4}{N^2} \ll \frac{4\sigma_{\xi}^2 \hat{\sigma}_{\tau^*}^2}{N} \Leftrightarrow$$

$$\Leftrightarrow \frac{\hat{\sigma}_{\tau^*}^2}{\sigma_{\xi}^2} \gg 0,5 - \frac{1}{2N},$$

is satisfied, we can take the value δ^2 as a determined variable and equal to σ_{ξ}^2 . In the left part of expression (6) there is the signal/noise ratio (SNR). So, the condition (6) may be interpreted as a requirement of a significant excess of the signal level above the noise level.

Taking into account (6), the expression (3) can be represented as:

$$r^{j*} = \frac{\sqrt{SNR} + r_{\xi}^{j*}}{\sqrt{SNR + 1 + 2\sqrt{SNR}r_{\xi}^{j*}}}. \quad (7)$$

A similar result was obtained in [27], under assumption that $r_{\xi}^{j*} \equiv 0$, that is possible only when $N \rightarrow \infty$. Correlation coefficient (7) is a function of only one random value with a known distribution:

$$w(r_{\xi}^{j*}) = \sqrt{\frac{N}{2\pi}} \exp\left(-\frac{(r_{\xi}^{j*})^2 N}{2}\right). \quad (8)$$

Using the function (8), moments of all orders can be obtained.

For the case $j \neq j^*$ correlation coefficient will be written as follows:

$$r^j = \frac{r^{jj*} \sqrt{SNR} + r_{\xi}^j}{\sqrt{SNR + 1 + 2\sqrt{SNR}r_{\xi}^j}}, \quad (9)$$

where it is introduced:

$$r^{jj*} = \frac{(\bar{\tau}^j - \bar{\tau}^j)^T (\bar{\tau}^* - \bar{\tau}^*)}{\|\bar{\tau}^j - \bar{\tau}^j\| \|\bar{\tau}^* - \bar{\tau}^*\|},$$

r_{ξ}^j is determined analogously to (4), but for the $\bar{\tau}^j$ spectrum, $\bar{\tau}^j$ is a reference spectrum of j substance in the database.

Function (9) depends on two random values, and we must bear in mind that those values are not independent. In order to find their joint distribution we must know the correlation function of noise. We assume that the noise is uncorrelated $\langle \xi_i \xi_j \rangle = \sigma_{\xi}^2 \delta_{ij}$, where δ_{ij} is the Kronecker symbol. Then, taking into consideration (8), covariance of correlation coefficients can be written

$$\begin{aligned} \text{cov}(r_{\xi}^j, r_{\xi}^j) &= \langle r_{\xi}^j, r_{\xi}^j \rangle - \langle r_{\xi}^j \rangle \langle r_{\xi}^j \rangle = \\ &= \frac{(\bar{\tau}^* - \bar{\tau}^*)^T (\bar{\tau}^j - \bar{\tau}^j)}{N^2 \hat{\sigma}_{\tau^*} \hat{\sigma}_{\tau^j}} = \frac{r^{jj*}}{N}, \end{aligned} \quad (10)$$

and the probabilistic correlation coefficient will equal r^{jj*} .

Thus, the simplified formulas for correlation coefficients (7) and (9) are derived, and also properties of random values included in them are found.

Correct detection of a substance

Since the analyzed spectrum is noisy, it is not always possible to identify the substance correctly. There may be cases where, for example, another substance will be identified, or when the signal is regarded as pure noise and omitted. In practice it is necessary to know with what probability the detection was made. Let us formulate a definition of correct detection. Suppose the spectrum \bar{x} of substance j^* from the database is investigated. Then

detection is considered correct if correlation coefficient r^{j*} has exceeded correlation coefficient with all other substances, and moreover, has exceeded some previously predetermined threshold $r_{j^*}^*$.

Mathematically, the probability of such an event can be written as follows:

$$P_{\text{correct}} = P\left[\left(r^{j*} > r^j, \forall j \neq j^*\right) \wedge \left(r^{j*} > r_{j^*}^*\right)\right] \quad (11)$$

Let us consider separately each of the expressions in parentheses. Substituting in (11) previously obtained expressions (7) and (9), then

$$r_{\xi}^{j*} - r_{\xi}^j > \sqrt{SNR} (r^{jj*} - 1). \quad (12)$$

Let us introduce the value Φ_j :

$$\Phi_j = r_{\xi}^{j*} - r_{\xi}^j = \frac{\bar{\xi}^T (\bar{\tau}^* - \bar{\tau}^*)}{N \sigma_{\xi} \hat{\sigma}_{\tau^*}} - \frac{\bar{\xi}^T (\bar{\tau}^j - \bar{\tau}^j)}{N \sigma_{\xi} \hat{\sigma}_{\tau^j}}.$$

Formula (11) shows that probability of correct detection can be represented as a function

of only one random vector $\vec{\Omega} = \{\Phi_1, \Phi_2, \dots, \Phi_{j^*}, \dots, \Phi_M\}$

having the length M (the number of substances in the database), and on the position of j^* there is the value r_{ξ}^{j*} . It is evident that all the components of the vector $\vec{\Omega}$ are statistically dependent and all of them are subject to the multidimensional Gaussian distribution.

To find their joint distribution we construct the covariance matrix for the values Φ_i и Φ_j . Let us use the terms from (10) for noise correlation, then

$$\text{cov}(\Phi_i, \Phi_j) = \frac{1 + r^{ij} - (r^{ij*} + r^{jj*})}{N}, \quad (13)$$

where $i, j \neq j^*$.

Covariance of values r_{ξ}^{j*} and Φ_j will equal

$$\text{cov}(r_{\xi}^{j*}, \Phi_j) = \frac{1 - r^{jj*}}{N}. \quad (14)$$

Let us consider the expression in the second bracket in (11). Substituting (7), we obtain:

$$\frac{\sqrt{SNR} + r_{\xi}^{j*}}{\sqrt{SNR + 1 + 2\sqrt{SNR}r_{\xi}^{j*}}} > r_{j^*}^*.$$

Let us denote the upper and lower bounds of the solution to this inequation with respect to r_{ξ}^{j*} as Γ_{up} and Γ_{down} respectively.

Combining (13) and (14), we obtain the final covariance matrix:

$$\Sigma = \frac{1}{N} \begin{pmatrix} 2(1 - r^{1j^*}) & 1 + r^{1,2} - (r^{1j^*} + r^{2j^*}) & \dots & 1 - r^{1j^*} & \dots \\ 1 + r^{2,1} - (r^{2j^*} + r^{1j^*}) & 2(1 - r^{2j^*}) & \dots & 1 - r^{2j^*} & \dots \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 1 - r^{1j^*} & 1 - r^{2j^*} & \dots & 1 & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots \end{pmatrix}$$

The joint probability density for vector $\bar{\Omega}$ will have the form

$$w(\Omega) = \frac{1}{\sqrt{(2\pi)^M |\Sigma|}} \exp\left(-\frac{1}{2} \Omega^T \Sigma^{-1} \Omega\right). \quad (15)$$

As a result, the probability of (11) can be found as an integral of (15) over the domain of admissible values:

$$P_{correct} = \int w(\Omega) d\Omega, \quad (16)$$

where Γ is the area bounded by $\Gamma_{up}, \Gamma_{down}$ and an inequation (12).

Let us briefly describe the resulting algorithm for calculating the probability of correct detection:

1. The tested spectrum is compared sequentially with all substances from the database,
2. A substance is selected with which the correlation coefficient was the highest,
3. If the maximum correlation coefficient exceeds empirical threshold, then the substance is considered identified.

4. The identified substance index is remembered.

5. By this index a covariant matrix Σ is built and region boundaries of values Γ are found.

6. According to the formula (16) the probability of correct detection is found.

Experimental test

To verify the obtained results, numerical experiments were carried out with the actual base of the spectral transmittance coefficients consisting of more than 50 spectra. Numerical simulation was carried out in MATLAB by generating random variables, followed by calculation of different statistical characteristics. Typically, the experiment was repeated 1000-10000 times with various substances, and non-biased statistical estimators were found.

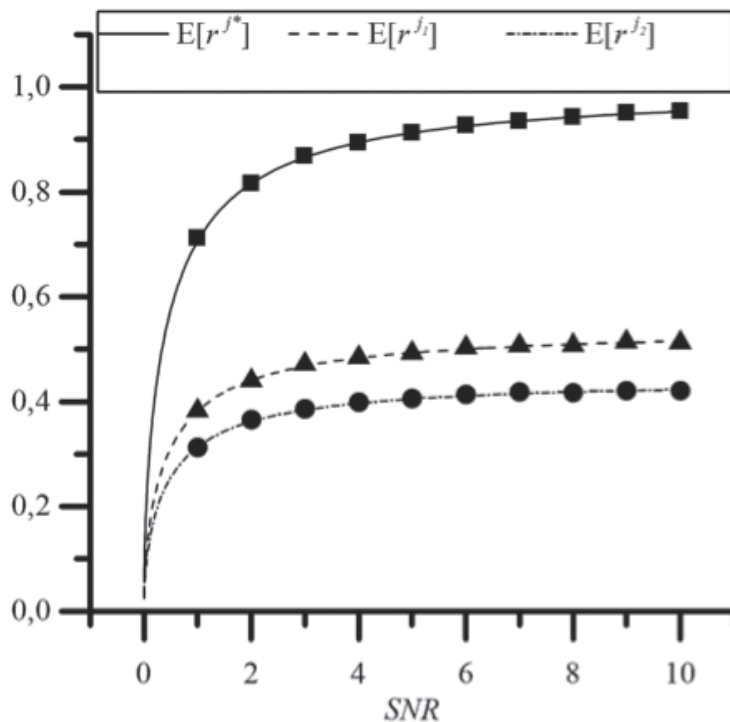


Fig. 1. Dependence of average value $E[r^j]$ of correlation coefficient on the "signal/noise" ("S / N") ratio

Figure 1 shows a comparison of the theoretical average values obtained from the formulas (7) and (9), with mathematical expectations of the correlation coefficient (1) (in the figure they are shown with points). We see good correlation of results, indicating the applicability of the assessment (6) for calculation of the mo-

ments of the first order. Further in Fig.2 experiments are shown on calculation of the mean square deviation in comparison to the dispersions of correlation coefficients for two pairs of test substances. However, in this case discrepancies were already observed at low values of the signal to noise ratio.

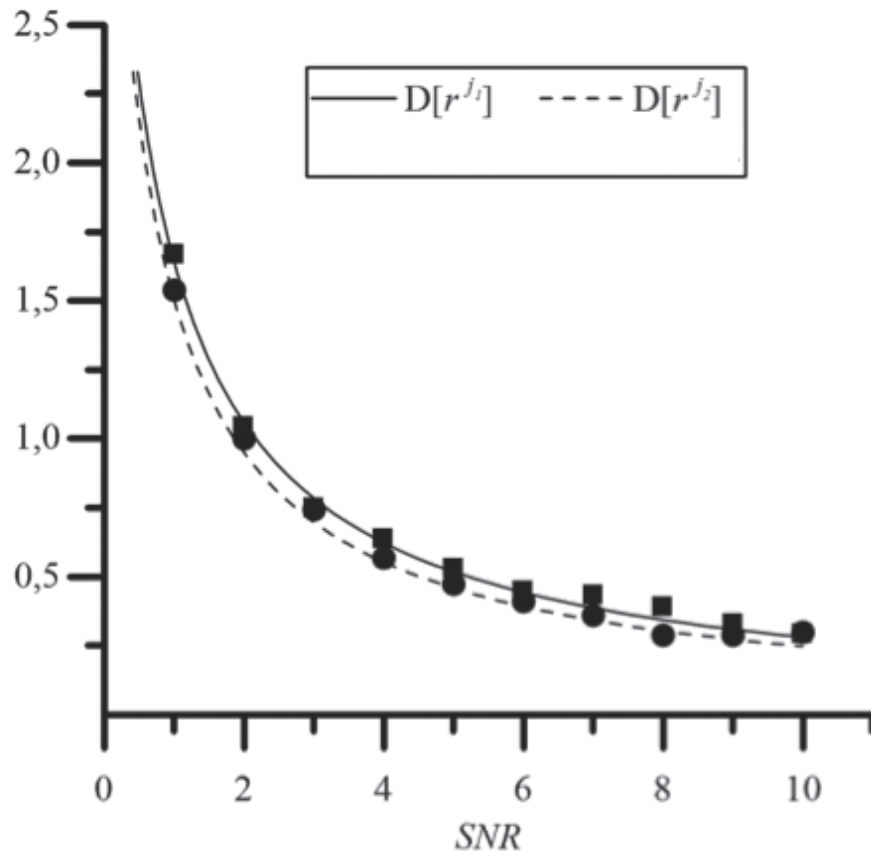


Fig.2. Dependence of the dispersion $D[r^j]$ of correlation coefficient on the "signal to noise" ratio (SNR)

From (15) and (16) we can see that in order to calculate probability of correct detection it is necessary to perform integration over the region of very large dimension (in this case $\dim(\Gamma) = 58$). This fact significantly slows down the speed of real systems and completely eliminates the possibility of working in real time. But during the numerical experiments it was found out that the estimate of the value (16) there may be replaced by the mathematical expectation of the probability of exceeding the correlation coefficient r^{j^*} over all others r^j :

$$P_{\text{correct}} = \frac{1}{M} \left[\sum_j P(r^{j^*} > r^j) \right] P(r^{j^*} > r_{j^*}^*), \quad (17)$$

wherein each probability in the sum (17) is expressed in terms of the error function:

$$P(r^{j^*} > r^j) = \frac{1}{2} \left(\operatorname{erf} \left(\sqrt{\frac{N}{1-r^{jj^*}}} \right) + \operatorname{erf} \left(\sqrt{\frac{N \cdot \text{SNR}(1-r^{jj^*})}{4}} \right) \right)$$

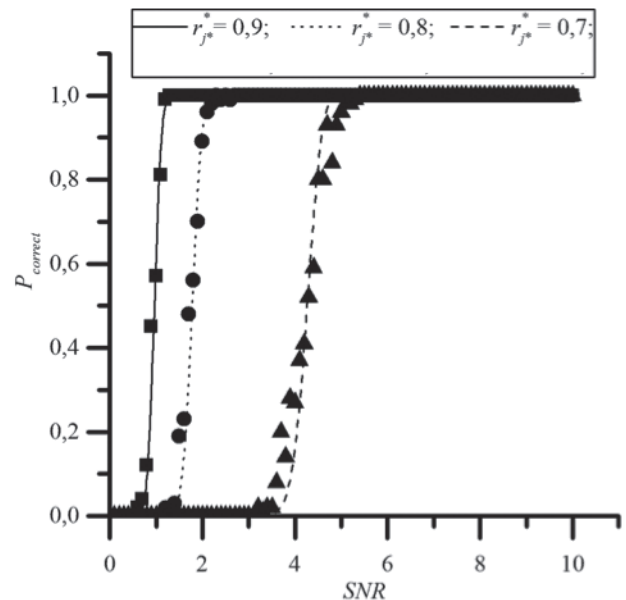


Fig.3. Probability of correct detection P_{correct} from the "s/n" ratio (SNR)

The results of comparison (17) with the experiment are shown in Fig.3. Each experimental point was calculated 1000 times and then averaged. The results are identical even for low SNR value. For the case $SNR > 1$ for all the substances it is shown that the probability of the event $r^{j*} < r^j$ is extremely small. Therefore, a crucial role is played by the right multiplier in (17):

$$P(r^{j*} > r_{j^*}^*) = \int_{\Gamma_{down}(r_{j^*}^*, SNR)}^{\Gamma_{up}(r_{j^*}^*, SNR)} w(r_{\xi}^{j*}) dr_{\xi}^{j*}. \quad (19)$$

It is possible to express (19) analytically using special functions, but the resulting expression appears too bulky to bring it to the article. Note that with the help of (19) we can find thresholds for a given probability of correct detection. Although

the relationship $r_{j^*}^*(P_{correct})$ is not expressed explicitly, determining the detection threshold for a given probability can be carried out using an interpolation search algorithm, as $P(r_{j^*}^*)$ is a monotonically decreasing function. Computational complexity of such an algorithm [28] may be estimated as

$$O\left(\log_2 \left[\log_2 \left(\frac{1}{\varepsilon} \right) \right] \right),$$

where ε is a permissible error. This result makes it possible to find substances detection thresholds for different SNR values, ensuring a predetermined correct detection probability (Figure 4).

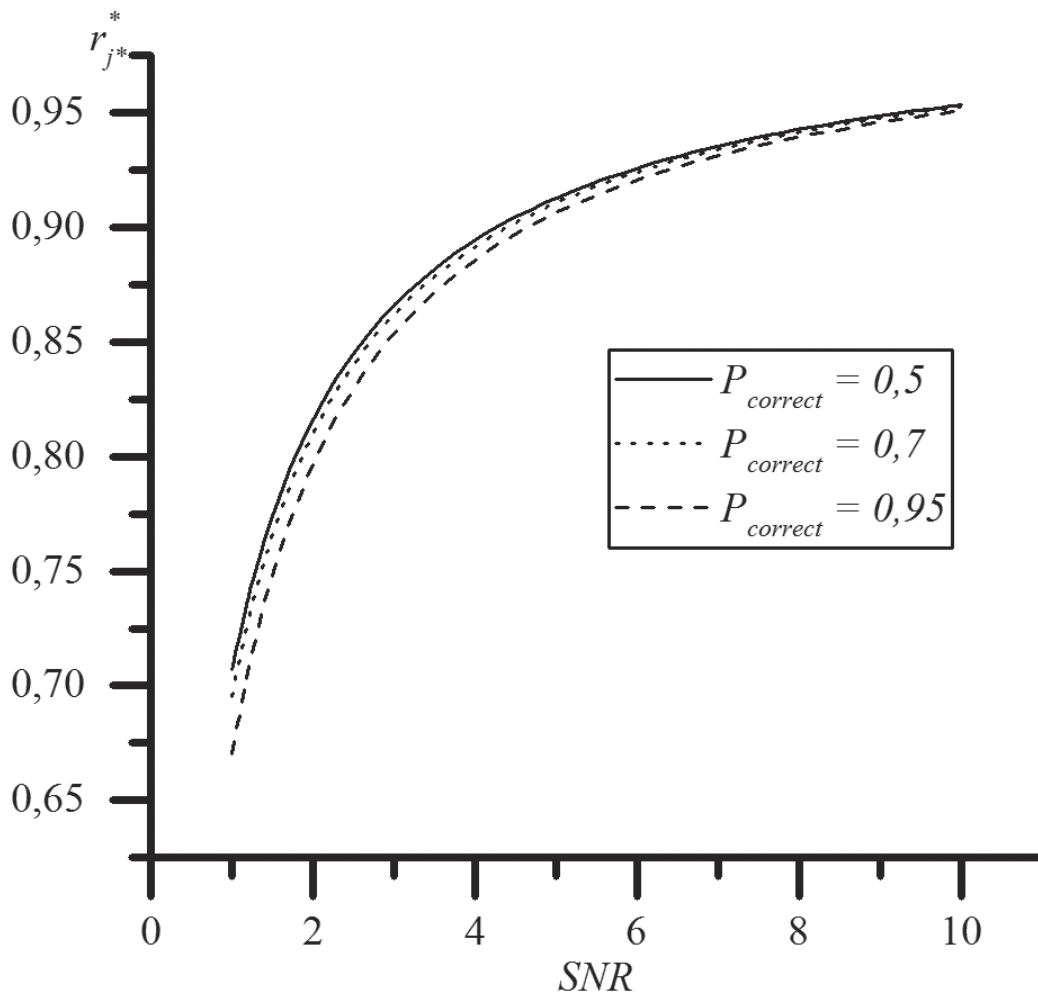


Figure 4. The threshold correlation coefficient depending on the signal to noise ratio for a given probability of correct detection.

Conclusion

The paper presents a method of determining the probability characteristics of the Pearson correlation coefficient in the case where one of the signals has an addition of Gaussian δ -correlated noise, and the other has no noise. This method is proposed to be used to identify substances by their spectra in the FTIR spectrometry. The basic assumption of the method is based on Replacement of the random value with chi-squared distribution by its statistical estimator. It is shown that this approach is applicable at high SNR.

Basing on the received probability characteristics the definition of correct detection probability was formulated and explicit theoretical expression for this probability was found. Also simplified expressions, that allow to find the probability a lot faster in numerical calculation, were found.

It is shown that the probability of correct detection depends on the threshold correlation coefficient, which previously was determined empirically. The technique is offered, allowing to find the detection threshold by a given probability of correct detection and signal to noise ratio.

The developed methods were tested on the actual database of substances spectra, consisting of 58 substances, and the applicability of proposed methods when performing introduced approximations was shown.

References

1. Morozov AN, Svetlichniy SI. Basics of Fourier Spectroradiometry. M: Science; 2006.
2. Hemmer MC, Gasteiger J. Prediction of Three-Dimensional Molecular Structures Using Information from Infrared Spectra. *Analitica Chimica Acta* 2000; 420(2): 145-154.
3. Ehrentreich F. Joined knowledge-and signal processing for infrared spectrum interpretation. *Analitica Chimica Acta* 1999; 393: 193-200.
4. Schoonjans V, Questier F. Assessing molecular similarity/diversity of chemical structures by FT-IR spectroscopy. *Journal of Pharmaceutical and Biomedical Analysis* 2001; 24: 613-627.
5. Lebedev KS. Data base usage in IR and mass spectroscopy for the structural detection of organic substances. *Journal of Analytical Chemistry* 1993; 48: 851-863.
6. Varmuza K, Penchev PN, Scsibrany H. Maximum Common Substructures of Organic Compounds Exhibiting Similar Infrared Spectra. *Journal of Chemical Information and Computer Sciences Impact* 1998; 38: 420-427.
7. Varmuza K, Penchev PN, Scsibrany H. Large and frequently occurring substructures in organic compounds obtained by library search of infrared spectra. *Vibrational Spectroscopy* 1999; 19: 407-412.
8. Penchev PN, Varmuza K. Characteristic substructures in sets of organic compounds with similar infrared spectra. *Computers&Chemistry* 2001; 25: 231-237.
9. Varmuza K, Kochev NT, Penchev PN. Evaluation of Hitlists from IR Library Searches by the Concept of Maximum Common Substructures *Analytical Sciences* 2001; 17: 659-662.
10. Derendyaev BG. Taxonomy of structures selected from the IR spectroscopy database. *Journal of Structural Chemistry* 2001; 42(2): 271-280.
11. Ehrentreich F. Three-step procedure for infrared spectrum interpretation. *Analitica Chimica Acta*; 427(2): 233-244.
12. Clerc JT, Pretsch E, Zurcher M. Performance Analysis of Infrared Library Search Systems. *Microchimica Acta [Wien]* 1986; 2: 217-242.
13. Luinge HJ. Automated interpretation of vibrational spectra. *Vibrational Spectroscopy* 1990; 1: 3-18.
14. Zurcher M, Clerc JT, Farkas M, Pretsch E. General theory of similarity measures for library search systems. *Analitica Chimica Acta*; 206(0): 161-172.
15. Pearson K. Notes on regression and inheritance in the case of two parents. *Proceedings of the Royal Society of London* 1895; 58: 40-242.
16. Kochikov IV, Morozov AN, Svetlichniy SI, Fufurin IL. Substance detection in the free atmosphere by single interferogram in FTIR. *Optics and Spectroscopy* 2009; 106, № 5: 743-749.
17. Harig R, Matz G. Toxic cloud imaging by infrared spectrometry: A scanning FTIR system for identification and visualization. *Field Analytical Chemistry&Technology* 2001; 5: 75-90. DOI: 10.1002/fact.1008
18. Beil A, Daum R, Matz G, Harig R. Remote sensing of atmospheric pollution by passive FTIR spectrometry in Spectroscopic Atmospheric Environmental Monitoring Techniques. *Proceedings of SPIE* 1998; 3493: 32-43.
19. Clerbaux C, Hadji-Lazaro J, Turquety S, Magie G, Coheur P-F. Trace gas measurements from infrared satellite for chemistry and climate applications. *Atmospheric Chemistry and Physics* 2003; 3: 1495-1508. DOI:10.5194/acp-3-1495-2003
20. Kochikov I, Morozov A, Fufurin I. Numerical procedures for substances identification and concentration calculation in the open atmosphere by processing a single ftir measurement. *Computer Optics* 2012; 36, №4: 554-561.
21. Zaitshev KI, Gavdush AA, Karasik VE, Jurchenko SO. Highly accurate reconstruction of spectral optical characteristics of a medium using terahertz pulsed spectroscopy. *Herald of the Bauman Moscow State Technical University. Ser. Natural Sciences* 2014; 3: 69-92.
22. Morozov AN, Svetlichniy SI, Tabalin SE, Fufurin IL. Physical basics of interferometer with rotating plate estimation. *Journal of Optical Technology* 2013; 80, № 8: 37-41.
23. Fisher RA. On the probable error of a coefficient of correlation deduced from a small sample. *Metron* 1921; 1(4): 3-32.
24. Kenney JF, Keeping ES. *Mathematics of Statistics. Pt. 2 -D*: Van Nostrand Company, inc.; 1951.

- **25.** Vasiljev NS, Morozov AN. Subsnce identification by error deformed spectra. *Computer Optics* 2014; 30, № 4: 856-864.
- **26.** Pitjev JP, Shishmarev IA. The theory of probability for physics. Moscow State University Publishing; 1983.
- **27.** Benesty J, Chen Jingdong, Yiteng H. On the Importance of the Pearson Correlation Coefficient in Noise Reduction. *Audio, Speech and Language Processing, IEEE Transaction on* 2008; 16(4): 757-765.
- **28.** Perl Y, Itai A, Avni H. Interpolation search — a log logN search. *Communications of the ACM* 1978; 21 (7): 550-553.

