SELECTION OF OPTIMAL SPECTRAL CHANNELS FOR SOLVING THE PROBLEMS OF ABSORPTION GAS ANALYSIS AND REMOTE SENSING. PART I. GENERALIZED ALGORITM AND SEARCH CRITERION

A.A. Mitsel'

Institute of Atmospheric Optics,
Siberian Branch of the Russian Academy of Sciences , Tomsk

Received March 31, 1992

A generalized algorithm for selecting optimal spectral channels for solving the sounding and absorption gas analysis problems is proposed. The algorithm is based on calculating information distance in space of gaseous object states. The conditions under which the problem of calculating the information distance is solved analytically and four algorithms that could be of practical interest are proposed.

INTRODUCTION

Knowledge of informative spectral channels is of great importance when optical devices are designed for operation in the atmosphere. The effective mathematical algorithms are needed to analyze absorption spectra of atmospheric and foreign gases in a wide wavelength range and to choose optimal spectral channels. In the light propagation problem the channels are selected based on the requirement for radiation energy transport over the atmospheric paths with minimum losses. These spectral intervals have to be chosen within the so–called atmospheric transmission windows. In the problems of gas analysis based on absorption spectroscopy the intervals are chosen in absorption bands of the gases under study for providing maximum sensitivity and selectivity with respect to gases.

In Ref. 1 we have proposed an algorithm for selecting spectral channels informative for solving the problems of gas analysis using a laser–excited optoacoustic detectors (OAD). This algorithm is based on the analysis of the average risk owing to adoption of one of the alternative statistical hypotheses H_1 or H_2 (e.g., H_1 means the absence of the gas under study in the mixture, H_2 – the presence of the gas).

In this paper a new algorithm for selecting optimal spectral channels informative for solving the problems of the absorption gas analysis and light propagation is presented.

BAYES CRITERION OF STATE DISTINGUISHABILITY

In the problems of absorption gas analysis the relation between the measured signals and the analyzed parameter in the general form can be written as

$$y = y_0\, \eta\, f(Kx + \beta)\ , \qquad (1)$$

where y and y_0 are the measured energy characteristics of radiation at the input and output of a medium under study, η is the sensitivity of the measuring system, $f(\cdot)$ is some functional (nonlinear, in the general case) of concentration x (analyzed parameter) of a gas and the absorption coefficient β of other gases and aerosol, and K is the coefficient of absorption per unit concentration of the gas.

In general, y and y_0 are assumed to be n–dimensional random vectors with the components $y_i = y(\lambda_i)$, $y_{0i} = y_0(\lambda_i)$, $i = 1, ..., n$; x is the m–dimensional vector of concentrations of the analyzed gases of the mixture with components x_j, where x_j is the jth gas concentration, $j = 1, ..., m$, $m \le n$; β is the n–dimensional absorption vector of foreign gases; K – $(n{\times}m)$ is the matrix of absorption coefficients.

Let us assume that there are two classes of states Y_1 and Y_2, and y can belong to one of them. The classes of states Y_1 and Y_2 are fixed by conventional probability densities $p(y/Y_1)$ and $p(y/Y_2)$. The fact that the signal y belongs to the class of results Y_r is, in turn, associated with the fact that the gas concentration vector x belongs to the class of states X_r, $r = 1, 2$. Let q and $p = 1 - q$ be the *a priori* probabilities of the fact that the unknown parameter y belongs to the classes Y_1 and Y_2. Moreover, the probability density $p(y_0)$ of the interfering parameter y_0 is assumed to be known.

To check the hypotheses that y falls either into the Y_1 or Y_2 class we make use of the Bayes solving rule which minimizes the solution error.[2] In what follows we shall concentrate on the solution error ε equal to

$$\varepsilon = q\varepsilon_1 + p\varepsilon_2\ ,$$

where

$$\varepsilon_2 = \int\limits_{-\infty}^{\mu} p(h/Y_2)\mathrm{d}h\ ,\ \varepsilon_1 = \int\limits_{\mu}^{\infty} p(h/Y_1)\mathrm{d}h\ ;$$

$$\mu = \ln(q/p)\ ,\ h = \ln[\tilde{p}(y/Y_2)/\tilde{p}(y/Y_1)]\ .$$

Here $p(h/Y_r)$ is the probability density of the random value h: $\tilde{p}(y/Y_r)$ is the density of the random vector y averaged over $p(y_0)$.

Let the initial distributions $p(y/Y_r)$ and $p(y_0)$ be normal with the parameters (M_{yr}, V_y) and (M_0, V_0), where M_{yr} is the vector of the average value in the class Y_r, $r = 1, 2$; $V_y = \mathrm{diag}\{\sigma_y^2(1), ..., \sigma_y^2(n)\}$ is the diagonal matrix of covariation of the vector y, the same for both classes Y_1 and Y_2; M_0 is the vector of the average value y_0; $V_0 = \mathrm{diag}\{\sigma_0^2(1), ..., \sigma_0^2(n)\}$ is the matrix of covariation of the vector y_0.

0235-6880/92/09 636-05 $02.00

Taking into account the fact that the measurement errors ξ_y and ξ_0 are independent and σ_0 does not depend on y_0 we obtain normal distribution $\tilde{p}(y/Y_r)$ with the parameters (M_r, V_r), where

$$M_r(i) = M_0(i)\,\eta_i\, f_i\,(Kx_r + \beta)\ ,\ r = 1,\ 2, \tag{2}$$

$$V_r = \mathrm{diag}\{\sigma_r^2(1)\ \dots\ \sigma_r^2(n)\}\ ,\ r = 1,\ 2\ . \tag{3}$$

Here x_r is the vector of gas concentration in the class X_r, $r = 1, 2$; $M_0(i)$, η_i, and f_i are the ith components of the vectors M_0, η, and f.

The values of variances $\sigma_r^2(i)$ are calculated by formula

$$\sigma_r^2(i) = \sigma_y^2(i) + \sigma_0^2(i)\ \eta_i^2\ f_i\,(Kx_r + \beta)\ . \tag{4}$$

To determine the classification error ε it is necessary to know the probability distribution of the logarithm of the likelihood ratio $p(h/Y_r)$. The following assumption is introduced: in both hypotheses the variances $\sigma_r^2(i)$, $r = 1, 2$ are taken equal and independent of the value h, i.e.,

$$\sigma_1^2(i) = \sigma_2^2(i) = \sigma^2(i)\ ,\quad i = 1,\ \dots,\ n.$$

Then we have $V_1 = V_2 = V$ and the distribution $p(h/Y_r)$ is normal with the parameters (η_r, Σ), $r = 1, 2$, $\eta_1 = -\eta_2 = -E$, $\Sigma = 2E$, and the expression for ε takes the form with $q = p = 0.5$

$$\varepsilon = 0.5\left[1 - \Phi\left(\frac{\sqrt{E}}{2}\right)\right], \tag{5}$$

where

$$E = \frac{1}{2}\,(M_2 - M_1)^{\mathrm{T}}\ V^{-1}(M_2 - M_1); \tag{6}$$

$$\Phi(x) = \frac{2}{\sqrt{\pi}}\int\limits_0^x \mathrm{e}^{-t^2}\mathrm{d}t\ .$$

The vectors M_1 and M_2 entering into expression (6) are given by formula (2); the symbol T denotes the transposed matrix; elements of the diagonal matrix V are determined by formula (4) where it is reasonable to take for x_r that vector, of x_1 or x_2, for which the elements $\sigma^2(i)$ are maximum

$$\sigma^2(i) = \max\,\{\sigma^2(a_i,\ x_1),\ \sigma^2(a_i,\ x_2)\} \tag{7}$$

or

$$V(x) = \max\,\{V(x_1),\ V(x_2)\}\ , \tag{7a}$$

where a_i is the ith row of the absorption coefficient matrix K. By taking the maximum value of the two values $\sigma^2(i)$ we shall overestimate ε, this makes it possible to improve the accuracy of distinguishing between the signals $y(x_1)$ and $y(x_2)$.

THE RELATIONSHIP OF CLASSIFICATION ERROR ε AND THE INFORMATION DISTANCE IN THE SPACE X.

Consider the value $d(Y_i,\ Y_j) = 2E(M_i,\ M_j)$ which is a distinguishability measure for the pair of states Y_i and Y_j in the measurement space (see Eq. (6)):

$$d(Y_i,\ Y_j) = d(i,\ j) = (M_i - M_j)^{\mathrm{T}}\ V^{-1}\ (M_i - M_j)\ , \tag{8}$$

and satisfies the conditions
a) $d(i,\ j) \ge 0$, $d(i,\ j) = 0$ when $i = j$; b) $d(i,\ j) = d(j,\ i)$; and, c) $\{d(i,\ j)\}^{0.5} \le \{d(i,\ k)\}^{0.5} + \{d(k,\ j)\}^{0.5}$.

The first two conditions are obvious, the third condition (inequality of triangle for pair states (i, j), (i, k), and (k, j)) can readily be proved using the Cauchy inequality.[3] The value d satisfying the conditions "a", "b", and "c" is the squared distance in the space of results between the classes Y_i and Y_j.

Let the class X_1 be a gaseous object characterized by the state x, and X_2 be the class characterized by the state $x + \mathrm{d}x$, where $\mathrm{d}x$ is the differential of vector x. Then in the space of results the vector $(M_2 - M_1)$ can be represented in the form

$$M_2 - M_1 = F(K(x + \mathrm{d}x) + \beta) - F(Kx + \beta) = R\mathrm{d}x\ , \tag{9}$$

where F is the n–dimensional vector with the components

$$F(i) = M_0(i)\ \eta_i\ f\left(\sum_{j=1}^{m} K_{ij}\,x(j) + \beta(i)\right);$$

R is the $(n{\times}m)$–dimensional matrix with the components

$$R_{ij} = \frac{\mathrm{d}F(i)}{\mathrm{d}x(j)}\ ,\quad i = 1,\ \dots,\ n\ ;\quad j = 1,\ \dots,\ m\ .$$

By substituting Eq. (9) into Eq. (8) one comes to the following relation:

$$(\mathrm{d}L)^2 = \mathrm{d}x^{\mathrm{T}}\ R^{\mathrm{T}}\ V^{-1}\ R\ \mathrm{d}x = \mathrm{d}x^{\mathrm{T}}\ G\ \mathrm{d}x\ , \tag{10}$$

where $G = R^{\mathrm{T}}\,V^{-1}\,R$.

Relation (10) is a differential quadratic form of the variables $x_1, \dots, x_m$ being the projections of the vector x and determines the squared distance between the states x and $x + \mathrm{d}x$ in the space X. Elements of the matrix G_{ij} determine a matrix tensor in the Riemann space.[4] A finite distance between the points given by the radius–vectors x_1 and x_2 are determined as a minimum length of the parametric curve $x(t)$, $t_1 \le t \le t_2$.

$$L = \min_{x\in X}\int \sqrt{z^{\mathrm{T}}Gz}\ \mathrm{d}t\ . \tag{11}$$

Minimum is taken over all the unbroken differential curves $x(t)$ the end points of which are fixed, i.e., $x(t_1) = x_1$, $x(t_2) = x_2$ (see Ref. 4). The components of vector z are given in the form $z(j) = \mathrm{d}x(j)/\mathrm{d}t$. The elements of the matrix G depend on t with respect to the vector projection x, i.e.,

$$G_{ij} = G_{ij}\,[x_1(t),\ \dots,\ x_m(t)]\ .$$

It should be noted that V.P. Kozlov (see Refs. 5 and 6) was the first who proposed to employ the Riemann information metrices for estimating the limiting capabilities of spectrophotometric measurement methods.

From formulas (6)–(8) and (10) follows the relation for the classification error ε and the information distance L in the state space X

$$\varepsilon(i,\,j) \geq \frac{1}{2}\left[1 - \Phi\left(\frac{L(i,\,j)}{2\sqrt{2}}\right)\right]. \tag{12}$$

or

$$\varepsilon \geq \frac{1}{2}\left[1 - \Phi\left(\frac{L}{2\sqrt{2}}\right)\right]. \tag{12a}$$

CALCULATION OF INFORMATION DISTANCE *L*

Consider the method of calculating L. Minimization of Eq. (11) is equivalent to the solution of the system of differential equations (the Euler equations)[7]

$$\frac{\partial^2\Theta}{\partial z_j^2}\, z'_j + \frac{\partial^2\Theta}{\partial x_j\,\partial z_j}\, z_j + \frac{\partial^2\Theta}{\partial t\,\partial z_j} - \frac{\partial\Theta}{\partial x_j} = 0\ , \tag{13}$$

where

$$\Theta = \sqrt{z^{\mathrm{T}} G\, z} = \Theta[x_1(t),\ \ldots,\ x_m(t);\ \ z_1(t),\ \ldots,\ z_m(t)].$$

Analytical solution of system (13) can be obtained only in the case when the vector y is linearly related to the vector x, i.e., $y \simeq Kx + \beta$. Here the elements of the matrix $G_{ij} = \{R^{\mathrm{T}} V^{-1} R\}_{ij}$ are independent of $x_j(t)$. Therefore the second and fourth terms in Eq. (13) vanish and the following system of differential equations appears

$$\left[G_{ij}\Theta^2 - \left(\sum_{l=l}^{m} G_{il}\, z_l\right)\right] z'_j + \sum_{l=1}^{m} G_{il}\, z'_l\Theta^2 -$$

$$- \sum_{l=1}^{m} G_{il}\, z_l \sum_{j\ k}^{m} z'_j G_{ik}\, z_k = 0\ .$$

A solution of this system is

$$z_j = \mathrm{const} = a_j\ ,\quad x_j = a_j\, t + b_j\ ,\quad j = 1,\ \ldots,\ m,$$

and the integration constants a_j and b_j are determined from the conditions

$$x_j(t_1) = a_j\, t_1 + b_j = \rho_1^j;\quad x_j(t_2) = a_j\, t_2 + b_j = \rho_2^j;\ t_1 = 0,\quad t_2 = 1.$$

The resulting expression is

$$x_j(t) = (\rho_2^j - \rho_1^j)t + \rho_1^j\ ;\quad z_j(t) = (\rho_2^j - \rho_1^j)\ ;$$

$$L = \int_0^1 \sqrt{z^{\mathrm{T}} G\, z}\ dt = \sqrt{z^{\mathrm{T}} G\, z}\ , \tag{14}$$

where ρ_1^j, ρ_2^j are the boundaries of the jth gas concentration range.

By comparing Eqs. (14) and (6) one can notice that the information distance $L = \sqrt{2E}$ for the linear relationship between the vector y and gas concentration x ($y \simeq Kx + \beta$).

The analytical solution of system (13) cannot be obtained for nonlinear relation between y and x vectors. However, it is possible to separate out three important examples where the variational problem (11) possesses a solution even for a nonlinear model.

The first example is related to a single–parameter state ($m = 1$). The expression for L when $m = 1$ takes the form

$$L = \int_{\rho_1}^{\rho_2} \sqrt{R^{\mathrm{T}}\, V^{-1} R}\ \mathrm{d}x\ , \tag{15}$$

i.e., the value L can be calculated using a conventional integration of a function of one variable.

The second example is related to a single wave method of y ($n = 1$) measurement. In this case the elements of matrix G_{ij} are reduced to the products of components of the vector Q, i.e., $G_{ij} = Q_i Q_j$ and relation (11) takes the form

$$L = \sum_{j=1}^{m} \int_{\rho_1^j}^{\rho_2^j} Q_j(x_1,\ \ldots,\ x_m)\ \mathrm{d}x_j\ . \tag{16}$$

And, finally, the analytical solution of variation of problem (11) is related to the situation when $n = 1$ and $m = 1$. From Eq. (11) immidiately follows the result:

$$L = \frac{M_0\eta}{\sigma} \int_{\rho_1}^{\rho_2} \left|\frac{\partial f}{\partial x}\right|\ \mathrm{d}x\ , \tag{17}$$

where the function f is assigned by formula (1).

CRITERION OF SEARCHING FOR OPTIMAL SPECTRAL CHANNELS

The information length L depends on characteristics of the measuring system (σ_y, σ_0, M_0, η) and the matrix of absorption coefficients K which, in turn, is wavelength dependent. For simplicity, let us consider the case of $n = 1$ and $m = 1$. The value L is determined by formula (17) and is a function of the wavelength λ. The value L changes with λ. The larger the value L the smaller is the classification error ε of a gaseous object and, hence, more reliable is the detection of a gas with a given device. On the contrary, the smaller the value L the larger is ε and, hence, less promising is a given spectral interval for measuring the gas content. If for all atmospheric gases with the absorption bands in a given spectral region the value L is small (it is desirable that $L \ll 1$), then it is apparent that this spectral region holds promise for light propagation problems.

Thus the problem on searching informative spectral intervals is reduced to calculating the length L. The search–for criterion for the problems of gas analysis can be presented in the form

$$L(\lambda) \to \max_{\lambda}\ ,\ \text{or}\ L(\lambda) \geq L_0\ ; \tag{18}$$

and for the problems of light propagation

$$L(\lambda) \to \max_{\lambda}\ ,\ \text{or}\ L(\lambda) \ll 1\ ; \tag{19}$$

In expression (18) L_0 is the threshold value of information distance which is uniquely related to the threshold value ε_0 of the classification error. For practical purposes, most common is the value $\varepsilon_0 = 0.05$ which corresponds to $L_0 = 3.3$.

In Ref. 6 Gal'tsev et al. proposed the value L to be interpreted as a number of pair distinguishable states of the gaseous object within the interval $[x_1, x_2]$ and $L_0 = 1$ being taken as the threshold value. When $L < L_0$ the states x_1 and x_2 cannot be distinguished using this device, when $L > L_0$ a number of the pair distinguishable states equals $L + 1$. Recognizing the importance of interpretation of L proposed in Ref. 6 it should be noted, however, that there is a subjective factor associated with the choice of the value $L_0 = 1$. When $L_0 = 1$ the error $\varepsilon = 0.3$, which is rather a large value. In practice, as was noted above, the value ε is conventionally taken at a level of 0.05 and lower which corresponds to $L = 3.3$ and larger.

CONCLUSIONS

The proposed criterion for selecting optimal spectral channels is a simplification of the approach we have earlier discussed in Ref. 1. In contrast to Ref. 1, in this paper we use a simple hypotheses in both of the state classes X_1 and X_2 of the information parameter x. Moreover, in these hypotheses the variances were taken equal and signal independent that enabled one to relate the solution error ε to a single parameter E (in Ref. 1 the average risk was calculated which was expressed through two parameters g_1 and g_2). Such an approach makes it possible to construct an effective search for algorithms for more complicated physical models describing the signals, in particular, for multidimensional and nonlinear models with any spectral resolution.

The introduction of a differential form of the information distance $(dL)^2$ (see Eq. (10)) has made it possible to pass from discrete states x_r to the continuous ones. As a result, the parameter L appeared which has more profound sense of a measure of the state distinguishability in the problem of object classifications. The value L can be treated as a number of distinguishable pair states in the given range of concentrations $[(\rho_1, \rho_2]$. The information distance in the classification problem was first introduced in Refs. 5 and 6 to determine the limiting capabilities of spectrophotometric methods for measuring gas contents. For practical purposes, Gal'tsev et al.[6] proposed an algorithm of the form of Eq. (17) ($n = 1$, $m = 1$). In this paper the conditions are described under which variational problem (11) is solved analytically and four general algorithms (14)–(17) have been derived. In the second part of this work from these algorithms the formulas have been derived for specific spectral devices operating based on the principle of absorption.

In conclusion the author thanks Drs. G.N. Glazov and S.D. Tvorogov for their useful discussions and critical comments at the starting point of this work.

APPENDIX

Let us introduce some definitions.

Gas object is a gas mixture of certain gases with absorption spectra in the optical region.

State of a gas object (in space of the information parameter X) is a gas mixture composed of different gases with some concentrations given by the vector x.

Measurement vector (observational vector) y is a set of measured energy characteristics of a gas object (e.g., optoacoustic signals at different wavelengths generated by an optoacoustic cell with a gas mixture when optical radiation passes through it).

State in the measurement space Y (in the space of results) is given by the measurement vector y uniquely associated with the vector x. The vector y represents some (in general, nonlinear) transformation of the vector of a gas object states x.

Class of states in space X is a set of random states of a gas object given by the vector x.

Class of states in space Y is a set of random states given by the vector y. It is determined by both states of a gas object and states of a measuring device.

Let us now consider briefly the problem of classification and its relation to the problem of choosing the informative wavelengths.

Let the measurement vector y be a random vector with the conventional probability density $p(y/Y_r)$, $r = 1, ..., l$. The probability density $p(y/Y_r)$ depends on the class Y_r which the vector y is being from. If the conventional probability density is known for each class Y_r, then the problem of classifying the gas objects is reduced to statistical varification of hypotheses.

Consider the case of two classes ($r = 1, 2$). It is necessary to check the hypothesis H_2 that the vector y belongs to the class Y_2 in contrast to the alternative H_1: $y \in Y_1$. The Bayes solving rule minimizing the solution error has the form[2]

$$p(y/Y_2)p(Y_2) \gtrless p(y/Y_1)p(Y_1) \to y \in \begin{cases} Y_2 \\ Y_1 \end{cases} \tag{A1}$$

or

$$l(y)=\frac{p(y/Y_2)}{p(y/Y_1)} \gtrless \frac{p(Y_1)}{p(Y_2)} \to y \in \begin{cases} Y_2 \\ Y_1 \end{cases}, \tag{A2}$$

where $p(Y_1)$ and $p(Y_2)$ are the *a priori* probabilities of the fact that the vector y is from the classes Y_1 and Y_2, respectively. The sign of inequality determines whether the object characterized by the vector y belongs to the class Y_1 or Y_2.

The value $l(y)$ is called the likelihood ratio, and $p(Y_1)/p(Y_2)$ is the threshold for a given solution rule. In many cases it is more convenient to use the logarithm of the likelihood ratio $h(y)$ instead of $l(y)$:

$$h(y) = \ln(l(y)) \gtrless \mu = \ln(p(Y_1)/p(Y_2)) \to y \in \begin{cases} Y_2 \\ Y_1 \end{cases} \tag{A3}$$

The solution rule (A3) does not provide a correct classification. The probability of the event that we misassign the measured vector y to a given class Y_r is called the probability of the solution error or merely the solution error. The distribution $p(h/Y_r)$ is shown schematically in the figure. The shaded sections under the distribution curves represent the errors of the first and second kinds ε_1 and ε_2 calculated by formulas

$$\varepsilon_2 = \int\limits_{-\infty}^{\mu} p(h/Y_2)\,\mathrm{d}h,\ \varepsilon_1 = \int\limits_{\mu}^{\infty} p(h/Y_1)\,\mathrm{d}h\ . \tag{A4}$$

The total error is found as a weighted mean sum of these errors

$$\varepsilon = p(Y_1)\,\varepsilon_1 + p(Y_2)\,\varepsilon_2\ . \tag{A5}$$

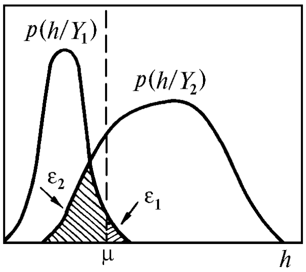

FIG. 1. Distribution of the logarithm of likelihood ratio.

It can readily be seen from relations (A4) and (A5) that the value ε does not depend on the measured vector y and therefore can be calculated before the experiment is carried out. Since $p(h/Y_r)$ depends on specifications of the measurement system, absorption coefficients, and gas concentrations, ε can be treated as a generalized parameter of the potential of a device for analyzing the gas mixture content. An essential feature of the problem under study is the dependence of ε on the absorption coefficients which are, in turn, wavelength dependent. This fact is used in our further discussions of the construction of the algorithm for selecting optimal spectral channels.

REFERENCES

1. M.Yu. Kataev, A.A. Mitsel, and S.R. Tarasova, Atm. Opt. **3**, No. 8, 763–772 (1990).
2. K. Fukunaga, *Introduction to the Statistical Theory of Image Recognition* (Nauka, Moscow, 1979), 367 pp.
3. R. Bellman, *Introduction to the Matrix Theory* (Nauka, Moscow, 1969), 367 pp.
4. P.K. Rashevskii, *Riemann Geometry and Tensor Analysis* (Nauka, Moscow, 1967), 568 pp.
5. V.P. Kozlov, Dokl. Akad. Nauk SSSR **166**, No. 4, 779–783 (1966).
6. A.P. Gal'tsev, V.P. Kozlov, and M.M. Rivin, in: *Problems of the Atmospheric Physics*, State University, Leningrad (1967), No. 5, pp. 129–145.
7. G.A. Corn and T.M. Corn, *Mathematical Handbook for Scientists and Engineers* (McGraw Hill, New York, 1961).