state problems, provided that relevant conditions are met. Individual derivatives are the only ones that are important and all arguments concerning the method of separation of the derivatives in the Enskog's method should be taken as referring to the individual derivatives. This is obvious from the physical point of view, since the Enskog method leads to the hydrodynamic representation of the motion of gas.

Of course, instead of confining ourselves to the first Enskog approximation for f, we could attempt to obtain a more accurate asymptotic representation by taking into account the second derivatives of f^+ in the Taylor expansion. This would have yielded the Burnett approximation (it would not take us into the higher rarefaction region, since the effective domain of the functional dependence would not be altered). The question whether this method would yield better accuracy in the general case is open to doubt, since this would demand the smoothness of f^+ and consequently of f beyond that given by the Boltzmann equation.

BIBLIOGRAPHY

1. Chapman, S. and Cowling, T. G., Mathematical Theory of Nonuniform Cases. Cambridge University Press, 1952.
2. Kogan, M. N., Dynamics of a Rarefied Gas. M., "Nauka", 1967.
3. Struminskii, V. V., On a method of solution of the Boltzmann equation. Dokl. Akad. Nauk SSSR, Vol. 158, №2, 1964.
4. Evgrafov, M. A., Asymptotic Estimates and Entire Functions, 2nd ed., M., Fizmatgiz, 1962.

Translated by L. K.

APPLICATION OF THE LEAST SQUARES METHOD TO A MODEL WHICH IS NONLINEAR WITH RESPECT TO THE PARAMETERS

PMM Vol. 32, №6, 1968, pp. 1143-1146

A. Z. VOLYNETS, B. S. DARKHOVSKII and B. A. KADER
(Moscow)

(Received May 28, 1968)

We consider the application of the least squares method to models which are nonlinear with respect to the parameters and for which a linearizing transformation Ψ exists (such problems e. g. arise in the experimental determination of the parameters of the exponential criterial equations). We prove that the values of the parameters obtained after the transformation deviate from the required values, and we show the logarithmic transformation as an example illustrating the method used to obtain formulas yielding the estimates of these errors.

Iteration method which we propose, retains the advantages of the computations based on the linear model, but removes the error mentioned above, and we give the sufficient conditions for its convergence. To illustrate the method, we use empirical data on the mass transfer at the wall in a turbulent fluid flow at large Schmidt numbers.

Let us consider a random variable function of the form $Y(x) = f(x) + \xi(x)$ (without loss of generality, we can assume that this function is defined for $a \leqslant x \leqslant b$ and is equal to zero outside this interval), where $f(x) = \langle Y(x) \rangle$ and $\xi(x)$ is a stationary random function such, that $\langle \xi(x) \rangle = 0$. Usually, the problem of estimating the regression curve $f(x)$

on the obtained realization $y(x)$, $a \leqslant x \leqslant b$ can be solved using the least squares method, i.e. minimizing the functional

$$\int_a^b [y(x) - \varphi(x, \theta)]^2 dx \tag{1}$$

where $\varphi(x, \theta)$ denotes a certain class of functions approximating the regression line $f(x)$ and θ is a set of unknown parameters fully defining the function $\varphi(x, \theta)$. As a rule, we also assume that $\varphi(x, \theta)$ is a linear function of its parameters, i.e. $\varphi(x, \theta) = \theta_1\varphi_1(x) + \ldots + \theta_N\varphi_N(x)$. In this case the problem of minimizing the functional (1) reduces to solving a simple linear system of equations (see e.g. [1]).

In practice we often encounter a situation in which the form of the function $f(x) = f(x, \theta)$ is well known, but the parameters θ which are being estimated, appear in it nonlinearly (this, for example, happens in approximating the experimental data obtained in the process of heat, mass or momentum transfer, using the criterial equations of the type $f(x, \theta) = kx_1^{l_1} \ldots x_n^{l_n}$, $\theta = (k, l_1, \ldots, l_n)$. In these cases, there exists a transformation $\Psi(f)$ linearizing the model which is nonlinear in θ, i.e. such transformation, that

$$\Psi[f(x, \theta)] = \theta_1 f_1(x) + \ldots + \theta_n f_n(x) \tag{2}$$

This means that the problem of estimation of the parameters θ_i reduces to the standard problem of the least squares method. Let us denote the value of θ estimated by this method by $\theta^{(1)}$, and by $\theta^{(2)}$ the estimate based on the requirement of minimization of the functional (1) in which $\varphi(x, \theta)$ is replaced with $f(x, \theta)$.

We note that either estimate could be considered optimal. In the latter case, the "quality" of the estimate is judged by the value of the integral (1), and in the former case – by some other metric. It must however be borne in mind that use of an arbitrary criterion causes the resulting method to deviate from the classical least squares method in the sense of [1] and, as we shall show later, hinders the process of comparing the parameters $\theta^{(1)}$ computed over various intervals (a, b), since $\theta^{(1)}$ is found to be strongly dependent on the interval of measurement. For this reason we shall, in the following, adopt $\theta^{(2)}$ as the "best", or even the "correct" estimate.

It can easily be proved that if Ψ is not an identity transformation, then $\theta^{(2)} \neq \theta^{(1)}$. Indeed, let us assume that $\Psi(Y)$ has a one-sided Laplace transform, i.e.

$$\Psi(Y) = \frac{1}{2\pi i} \int_{\gamma - i\infty}^{\gamma + i\infty} \Phi(p)\, e^{pY}\, dp$$

Then

$$\langle \Psi(Y) \rangle = \frac{1}{2\pi i} \int_{\gamma - i\infty}^{\gamma + i\infty} \Phi(p)\, e^{pf(x)} \langle e^{-p[-\xi(x)]} \rangle\, dp = \frac{1}{2\pi i} \int_{\gamma - i\infty}^{\gamma + i\infty} \Phi(p)\, \Lambda(p)\, e^{pf(x)}\, dp$$

where $\langle e^{-p[-\xi(x)]} \rangle = \Lambda(p)$ is a two-sided Laplace transform for the probability density of $-\xi(x)$ (assuming that it exists).

Thus $\langle \Psi(Y) \rangle = G[f(x)]$ where $LG = \Phi(p)\,\Lambda(p)$, and by the convolution theorem

$$G(t) = \int_{-\infty}^{t} \Psi(t - \tau)\, \lambda(\tau)\, d\tau = \int_0^{\infty} \Psi(\tau)\, \lambda(t - \tau)\, d\tau \tag{3}$$

where L denotes the Laplace transform and λ is one-dimensional probability density of $-\xi(x)$ of the process.

For the equality $\theta^{(1)} = \theta^{(2)}$ to hold, function $\Psi(Y)$ should obviously satisfy

$$\Psi(t) = \int_0^{\infty} \Psi(\tau)\, \lambda(t - \tau)\, d\tau \tag{4}$$

When the least squares method is used in practice, the results showing large deviations from the regression curve $f(x)$ are usually assumed erroneous and subsequently neglected. This fact enables us to consider only the distributions $\lambda(\tau)$ concentrated on the finite segment $\tau \in (A, B)$. We can assume, without loss of generality, that $Y(x) > 0$ for any $a \leqslant x \leqslant b$ and write (4) as

$$\Psi(t) = \int_{\delta}^{\gamma} \Psi(\tau)\,\lambda(t-\tau)\,d\tau \tag{5}$$

where δ and γ denote, respectively, the minimum and maximum of

$$\{f(x) \pm A,\ f(x) \pm B\}, \qquad a \leqslant x \leqslant b$$

It can easily be confirmed that the identity transformation $\Psi(t) \equiv t$ will be a solution of the integral equation (5) and, that it will be unique (this follows from the derivation of (4)). This in turn implies that $\theta^{(1)} = \theta^{(2)}$ if and only if the transformation is an identity transformation.

We note that $\lambda(t-\tau) \to \delta(t-\tau)$ as $\sigma \to 0$ and Eq. (4) becomes an identity $\Psi(t) \equiv \equiv \Psi(t)$, i.e. the error in determination of the correlation equation parameters diminishes under an arbitrary transformation Ψ, with decrease in the standard deviation σ of the experimental data from the regression line.

Formula (3) can be used to estimate the discrepancy between $\theta^{(1)}$ and $\theta^{(2)}$ and to find the proper correction $(\theta^{(2)} - \theta^{(1)})$ to the value of $\theta^{(1)}$.

Let us, for example, consider the regression line of the form $f(x) = k x^l$ assuming, for definiteness, that $l > 0$, so that $f(x)$ increases monotonically with x changing from a to b. In addition we shall assume that the distribution of the experimental points relative to the regression line obeys the "truncated normal law" with the density

$$\lambda(\tau) = \frac{1}{(1-\alpha)\sqrt{2\pi}\sigma} \exp\left(-\frac{\tau^2}{2\sigma^2}\right) \qquad \tau \in (A, B)$$

$$\lambda(\tau) = 0, \qquad \tau \notin (A, B)$$

where

$$1-\alpha = \frac{1}{\sqrt{2\pi}\sigma}\int_{A}^{B} \exp\left(-\frac{\tau^2}{2\sigma^2}\right)d\tau, \quad A = -q\sigma, \quad B = q\sigma, \quad 0 < q < \infty$$

Then by the previous argument, we have

$$G(t) = \frac{1}{(1-\alpha)\sqrt{2\pi}\sigma}\int_{-q\sigma}^{q\sigma} \ln(t-\tau)\exp\left(-\frac{\tau^2}{2\sigma^2}\right)d\tau$$

or, taking into account the fact that $t > \tau$ by definition, (see e.g. [2])

$$G(t) = \ln t - \frac{1}{(1-\alpha)\sqrt{2\pi}}\sum_{k=1}^{\infty}\frac{\sigma^{2k}}{2k}\int_{-q}^{q} u^{2k}e^{-1/2u^2}du = \ln t - \Delta(\alpha, \sigma)$$

$$\Delta(\alpha, \sigma) \approx \frac{1}{\sqrt{2\pi}}\sum_{k=1}^{\infty}\frac{\sigma^{2k}}{2k}\frac{2^{1/2(2k+1)}}{t^{2k}}\Gamma\left(\frac{2k+1}{2}\right) \tag{6}$$

where the divergent series should be interpreted in the asymptotic sense.

Restricting ourselves to the first term of (6) and estimating $k^{(2)}$ and $l^{(2)}$ from the condition of minimum of the functional

$$\int_{a}^{b}[\ln(kx^l)^{(1)} - \Delta - \ln(kx^l)^{(2)}]^2\,dx$$

we find, that when $l^{(1)} \neq 1/2$

$$\ln k^{(2)} - \ln k^{(1)} = \pm 2\sqrt{\frac{2}{\pi}}\left\{\frac{\sigma^2}{k^2}\frac{1}{1-2l}\frac{1}{C^{2l}}\left[\frac{1}{2}+\frac{l(\ln C-1)}{1-2l}\right]\right\}^{(1)}$$

$$l^{(2)} - l^{(1)} = \mp 2\sqrt{\frac{2}{\pi}}\left[\frac{\sigma^2}{k^2}\frac{l}{(1-2l)^2}\frac{1}{C^{2l}}\right]^{(1)} \qquad (7)$$

where $\ln C = {}^1/_2(\ln a + \ln b)$ is the logarithmic mean interval of measurement.

We can easily see from (7) (which are estimates), that the corrections depend not only on the extent of scatter of the experimental data, but also on the interval of measurements. This makes the comparison of the experimentally obtained parameters estimated according to the nonlinear linearized model and referring to different intervals of measurement either very difficult, or impossible.

We can, however, propose a simple iterative method of obtaining $\theta^{(2)}$ retaining the advantages of the linear model computations. At the n th step we seek θ_n which minimizes the functional

$$\int_a^b \beta(x, \theta_{n-1})\{\Psi[y(x)] - \Psi[f(x, \theta_n)]\}^2\,dx \qquad (8)$$

where

$$\beta(x, \theta_{n-1}) = \left\{\frac{y(x) - f(x, \theta_{n-1})}{\Psi[y(x)] - \Psi[f(x, \theta_{n-1})]}\right\}^2$$

It can be shown that this iterative process converges, if the function $\Psi'[f(x, \theta)]$ is bounded and different from zero for $a \leqslant x \leqslant b$ and, if such value θ_0 is chosen as the initial approximation that $f(x, \theta_0)$ approximates the observed values (e.g. $\theta^{(1)}$ can be used as θ_0).

Convergence of the process improves considerably if θ_0 in the first approximation is determined graphically, or if we set $\beta(x, \theta) = \{\Psi'[y(x)]\}^{-2}$ (9)

We shall illustrate the results obtained by applying them to the experimental data obtained in [3]. Wide range of variation of the arguments (Schmidt number N_{Sc} varies between 430 and 100000) and narrow spread of the experimental data (obtained at the fixed value of the Reynolds number $N_{Re} = 10^4$) relative to the approximating function $N_{Nu} = kN_{Sc}^l$ (N_{Nu} is the Nusselt number) ensures objective test of the methods.

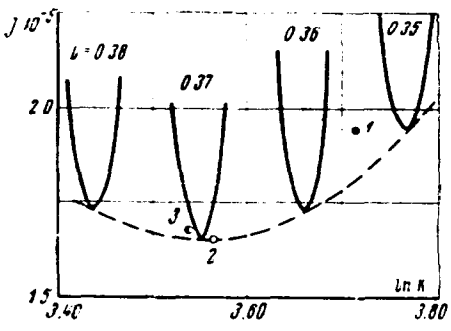

Fig. 1

Let us find $k^{(2)}$ and $l^{(2)}$ corresponding to the minimum value of the functional (1) which in this case has the form

$$J = \sum_{i=1}^{28}(N_{Nu_i} - kN_{Sc_i}^l)^2$$

We construct an envelope to the parabolas and find its minimum. Each parabola is computed for a fixed l and varying k (see Fig. 1). Point 1 of the graph corresponds to the values $k = 40.3$ and $l = 0.355$ recommended in [3] and obtained using (8) with $\beta \equiv 1$, i.e. by the usual least squares method applied to $\Psi(f)$. Using these values as initial and proceeding to the second approximation, we obtain $k = 35.4$ and $l = 0.369$ (point 2) which is seen to be very accurate.

We note an interesting fact, that the first approximation with $c\beta_i = [N_{Nu_i}]^2$ (compared with (9)) gives already an excellent result of $k = 34.4$ and $l = 0.372$ (point 3 on the

graph). Results of similar accuracy are obtained using Formulas (7) ($k = 35.8$ and $l = 0.376$) and this illustrates the usefulness of such formulas in estimating probable errors in the values of parameters arising from the nonlinearity present in the linearizing transformation.

BIBLIOGRAPHY

1. Cramer, H., Mathematical Methods of Statistics. Princetown Univ. Press, 1946.
2. Ryzhik, I. M. and Gradshtein, I. S., Tables of Integrals, Sums, Series and Products. M.-L., Gostekhizdat, 1951
3. Harriott, P. and Hamilton, R. M., Solid–liquid mass transfer in turbulent pipe flow. Chem. Eng. Sci., Vol. 20, №12, 1965.

Translated by L. K.