ELSEVIER

Analytica Chimica Acta 429 (2001) 171–183

ANALYTICA CHIMICA ACTA

www.elsevier.com/locate/aca

Data analysis in the chemical laboratory II. The end-point estimation in instrumental titrations by nonlinear regression

Karel Kupka [a], Milan Meloun [b,*]

[a] *TriloByte, Statistical Software Ltd., U sokolovny 21, 530 02 Pardubice, Czech Republic*
[b] *Department of Analytical Chemistry, University of Pardubice, 532 10 Pardubice, Czech Republic*

Received 22 December 1999; received in revised form 26 October 2000; accepted 26 October 2000

Abstract

The regression model for a two-segments titration curve with a break-point at the end-point is analyzed. Both linear and nonlinear shapes of the titration curve segments are treated. An effective and simple method discriminates which of two segments is linear or curved. The point and interval estimates of the end-point are calculated by the nonlinear least squares of curve fitting technique. The nonlinear regression method is applied to any, linear or nonlinear, type of a two-segments titration curve *without excluding* any titration points to reach the most probable point estimate of the end-point together with its $100(1-\alpha)$% confidence interval. An accuracy and precision of the proposed end-point estimation is examined on several instrumental titrations. A sample program in S-Plus^{TM} is provided.

Keywords: Two-segments titration curve; Break-point; Two lines intersection; End-point; Nonlinear least squares; Titration curve

1. Introduction

Titrimetric procedures based on a determination of the end-point, i.e. the point at which volumetric titration is completed, have been successfully employed over a wide range of concentration and have always been popular because of their simplicity, speed, accuracy, and good reproducibility. The importance of titrimetric analysis has increased with the advance of instrumental method of the end-point detection which are generally sensitive. The accuracy and precision of the results of a titrimetric determination depend on a nature of the titration reaction, but they are also influenced by the technique of the end-point location, [1–2,4–17].

Among numerous methods of an instrumental end-point detection the following techniques have found wide application: the *potentiometry* with a change in the potential of an indicating electrode, the *amperometry* with a change in the diffusion current at a polarizable electrode, the *conductimetry* with a change in conductivity of the solution being titrated and the *photometry* with a change in absorbance of the titrated solution. The relation between the signal response of titrand y (here the *dependent* or *response* variable, i.e. pH, emf, current, conductivity, absorbance, etc.) and the volume of titrant added x (here the *independent* or *predictor* variable) is called the *response function* of the instrumental titration system. Under real conditions, i.e. when perturbations in

* Corresponding author. Tel.: +42-40-603-7026; fax: +42-40-603-7068.
E-mail address: milan.meloun@upce.cz (M. Meloun).

0003-2670/01/$ – see front matter
PII: S0003-2670(00)01270-8

signal are present, the response function has a random character.

There are two types of a mathematical model for the response function. One is linear, for example, in most titrimetric, photometric, conductimetric or polarographic methods while the other is exponential, for example, in the potentiometric method. Linear response functions are generally preferred, and when the response function is nonlinear, a linearization procedure has been commonly used, with a suitable change of variables. In practice, the term linear dependence is used when the ratio of increments Δy of dependent variable to the increment Δx of independent variable assumes values which are randomly spread about the mean. If the ratio $\Delta y/\Delta x$ exhibits a certain trend, the dependence is nonlinear.

Ortiz-Fernández and Herrero-Gutiérrez [12] applied the robust regression by the least median squares LMS to data from various titrations. This robust approach tales the curvature between straight lines of a titration curve as outliers and/or leverages on each straight line segment. Omitting these point (e.g. Mach et al. [4]) or a robust approach lead to unwanted changes in the estimated values of the slopes and intercept terms. All curvature points are taken as erroneous experimental data, and therefore, excluded from a next computation. However, excluding some curvature points of titration curve is a results of a false regression model fitting (usually two straight line segments) and could lead to a lost of some experimental information.

In this paper the mathematical method of the end-point estimation is considered as one of the most important factors in a titrimetric procedure, since the volume of titrant used is directly affected by its accuracy and precision. The nonlinear regression method will be applied to any, linear or nonlinear, type of a two-segments titration curve *without excluding* any titration points to reach the most probable point estimate of the end-point together with its $100(1-\alpha)\%$ confidence interval. An accuracy and precision of nonlinear regression estimation also will be examined.

2. Theoretical

Supposing a two-segments titration curve of n pairs $\{x, y\}$ of titration curve points being described by a response function with the measured dependent, *response* variable y (i.e. potential, polarographic current, conductivity or absorbance) and independent, *predictor* variable x (i.e. the volume of titrant added). Usually, the two segments are supposed to be linear and can be modeled by two regression lines

$$f(x, \boldsymbol{p}) = \left\{ \begin{array}{ll} a_1 + b_1 x & \text{for } x \le x_{\text{ep}} \\ a_2 + b_2 x & \text{for } x > x_{\text{ep}} \end{array} \right\} \tag{1}$$

where $\boldsymbol{p} = (x_{\text{ep}}, a_1, a_2, b_1, b_2)^{\text{T}}$ being unknown represents the x value at the end-point (x_{ep}). We may construct a conditioned regression model with four unknown parameters x_{ep}, a_1, a_2, b_1, and $b_2 = b_1 + (a_1 - a_2)/p$ being expressed

$$y = \left\{ \begin{array}{ll} a_1 + b_1 x + \epsilon & \text{for } x \le x_{\text{ep}} \\ a_2 + b_2 x + \epsilon & \text{for } x > x_{\text{ep}} \end{array} \right\} \tag{2}$$

This regression model is nonlinear in parameter x_{ep}. Assuming that random errors $\epsilon_i, i = 1, \ldots, n$, has normal distribution, $\epsilon \sim N(0, s^2)$ with constant variance, we may use the nonlinear least squares for an estimation of four unknown parameters $\boldsymbol{p} = (x_{\text{ep}}, a_1, a_2, b_1)^{\text{T}}$, $r = 4$. The least squares regression minimizes the residual square sum function

$$U(\boldsymbol{p}) = \sum_{i=1}^{n} (y_i - f(x_i, \boldsymbol{p}))^2 \tag{3}$$

to find the optimal (least squares) estimates $\boldsymbol{p}^*$. Let us denote the number of unknown parameters r. Given the covariance matrix $\boldsymbol{C}(r \times r)$ of the parameters $\boldsymbol{p}$ and the estimated residual variance s^2, we may estimate variances of the respective parameters

$$\text{var}(\boldsymbol{p}) = s^2 \times \text{diag}(\boldsymbol{C}) \tag{4}$$

where $\text{diag}(\boldsymbol{C})$ is a vector consisting of diagonal elements of matrix $\boldsymbol{C}$. The residual variance is estimated from residuals as

$$s^2 = \frac{1}{n-r} \sum_{i=1}^{n} (y_i - f(x_i, \boldsymbol{p}))^2 \tag{5}$$

The covariance matrix $\boldsymbol{C}$ is computed as an inverse of Hessian i.e. an inverse of a second order partial derivatives of the objective function $U(\boldsymbol{p})$

$$\boldsymbol{C} = \boldsymbol{H}^{-1}, \qquad H_{ij} = \frac{\delta^2}{\delta p_i \delta p_j} U(p) \tag{6}$$

The Hessian matrix $\boldsymbol{H}$ is usually computed from the Jacobi matrix $\boldsymbol{J}(n \times r)$, cf. [3]

$$J_{ij} = \frac{\delta}{\delta p_j} U(\boldsymbol{p}), \qquad \boldsymbol{H} \approx \boldsymbol{J}^{\mathrm{T}}\boldsymbol{J} \tag{7}$$

All the above mentioned quantities, $\boldsymbol{C}$, s^2, var($\boldsymbol{p}$) are usually part of output of a regression program and do not need to be computed manually. In our algorithm, all derivatives are calculated numerically according to the simple symmetric rule $\mathrm{d}f(x)/\mathrm{d}x \approx (f(x + d) - f(x - d))/2d$, where $d = 10^{-12}$ $(x_{\max} - x_{\min})$. Since x_{ep} will never lie at an experimental point x_i, there is no need to calculate derivatives of $f(x)$ at x_{ep} (the probability that $x_{\mathrm{ep}} = x_i$ is of the order 10^{-e}, where e is number of decimal points of the computer arithmetic's, typically e is from 15 to 18). The variance of the end-point x_{ep}, denoted s_p^2 (here it is represented by the first element of var($\boldsymbol{p}$)), can be used to estimate the $100(1 - \alpha)$% confidence interval of parameter $\boldsymbol{p}$ for $r = 4$

$$x_{\mathrm{ep}}^* - s_p t_{1-\alpha/2}(n - r) < x_{\mathrm{ep}} < x_{\mathrm{ep}}^* + s_p t_{1-\alpha/2}(n - r), \quad L_{\mathrm{L}} < x_{\mathrm{ep}} < L_{\mathrm{U}} \tag{8}$$

here x_{ep}^* is the best estimate of x_{ep}, $t_{1-\alpha/2}(n-r)$ is the α-quantile of the t-distribution with $(n - r)$ degrees of freedom, L_{L} is the lower limit and L_{U} is the upper limit of an interval estimate of the end-point.

It is often doubtful, whether the segments of the titration curve are linear or nonlinear, curved. The curved shape of one or both segments may result from the theoretical model, which is in fact a more or less complicated logarithmic or exponential expression. It may appear to be linear only due to, for example, big differences in dissociation constants, etc. Therefore, the nonlinearity may result from non-ideal behavior in the solution at higher concentrations or because of unstable compounds. It shows up to be redundant to use the theoretical models in order to determine the end-point. On the other hand, the use of the linear approximation of the apparently curved segments may lead to wrong and unreliable results, as it does when omitting the curved part of data, and thus, losing information and precision.

(1) An application of the second-order polynomial seems to be quite effective to fit both segments of titration curve, $r = 6$

$$f(x, \boldsymbol{p}) = \left\{ \begin{array}{ll} a_1 + b_1 x + c_1 x^2 & \text{for } x \le x_{\mathrm{ep}} \\ a_2 + b_2 x + c_2 x^2 & \text{for } x > x_{\mathrm{ep}} \end{array} \right\} \tag{9}$$

with $p = (x_{\mathrm{ep}}, a_1, a_2, b_1, c_1, c_2)$ and $b_2 = b_1 + ((a_1 - a_2)/x_{\mathrm{ep}}) + x_{\mathrm{ep}}(c_1 - c_2)$. However, using the model (9) to fit linear segments leads to excessively high variances of $\boldsymbol{p}$ due to lower degrees of freedom ($n - 6$ instead of $n - 4$) and also due to a higher multicollinearity. The significance of quadratic terms was used to justify their presence in (9). The regression parameter β_i is significant if the relation (10)

$$|\beta_i^*| > s(\beta)_{p,i} \cdot t_{1-\alpha/2}(n - r) \tag{10}$$

is valid.

(2) For titrations with curvature around end-point, we suggest models of the type

$$f(x, \boldsymbol{p}) = \left\{ \begin{array}{l} a_1 + b_1 x + g_1(x, p_1) \text{ for } x < x_0 \\ a_2 + b_2 x + g_2(x, p_2) \text{ for } x \ge x_0 \end{array} \right\} \tag{11}$$

where x_0 is the model break-point not necessarily identical with the end-point of titration, x_{ep}, $g_1(x)$ converges to 0 for $x \to -\infty$ and $g_2(x)$ converges to 0 for $x \to +\infty$. Thus, $f(x, \boldsymbol{p})$ will converge to a line for $|x - x_0| \gg 0$. The task of finding the end-point is then reduced to estimation of g_1 and g_2 and finding the parameters of the two lines and their interception, x_{ep}. This methodology takes in account nonlinear behavior of the curve and satisfies the assumption of homoscedastic errors with zero mean. Since mostly the theoretical titration curve models are exponential, we used model

$$f(x, \boldsymbol{p}) = \left\{ \begin{array}{l} a_1 + b_1 x + c_1 \exp(d_1 x) \text{ for } x < x_0 \\ a_2 + b_2 x + c_2 \exp(d_2 x) \text{ for } x \ge x_0 \end{array} \right\} \tag{12}$$

with the condition of continuity and smoothness at the break-point x_0

$$y_1(x_0) = y_2(x_0) \text{ and } y_1'(x_0) = y_2'(x_0) \tag{13}$$

Applying relation (13) into Eq. (12), we reduce the number of parameters from nine to seven and obtain Eqs. (14) and (15)

$$y = a_2 + b_2 x + c_2 \exp(d_2 x_0) - b_1 x_0 - \frac{b_2 + c_2 d_2 \exp(d_2 x_0) - b_1}{d_1 \exp(d_1 x_0)} \exp(d_1 x_0) + b_1 x + \frac{b_2 + c_2 d_2 \exp(d_2 x_0) - b_1}{d_1 \exp(d_1 x_0)} \exp(d_2 x) \quad \text{for } x < x_0 \tag{14}$$

$$y = a_2 + b_2 x + c_2 \exp(d_2 x) \quad \text{for } x \geq x_0 \tag{15}$$

where b_1 is the slope of the left branch, d_1 the exponential term for the left branch, a_2 the absolute term for the linear part of the right branch, b_2 the slope for the linear part of the right branch, c_2 the multiplicator of the right exponential term, d_2 the exponential term for the left branch, x_0 the break-point of the two models. The parameters b_1, d_1, a_2, b_2, c_2, d_2 and x_0 may be easily estimated with a nonlinear regression software. Note that for very small values of c_1, d_1 or c_2, d_2 relation (12) can be simplified to (1).

With known parameters estimates of b_1, d_1, a_2, b_2, c_2, d_2 and x_0 the intercept x_{ep} of the two lines can be easily calculated as

$$x_{ep} = \frac{1}{b_1 - b_2}\left((b_1 - b_2)x_0 - c_2 \exp(d_2 x_0) + \frac{b_2 + c_2 d_2 \exp(d_2 x_0) - b_1}{d_1 \exp(d_1 x_0)} \exp(d_1 x_0)\right) \tag{16}$$

From the regression analysis we know asymptotic variances and also standard deviations for all the parameters and we can determine the standard deviation and uncertainty of x_{ep} using the law of error propagation [18]. Equations of the two lines will be $p_0 + b_1 x$ and $a_2 + b_2 x$, respectively, where

$$p_0 = a_2 + (b_2 - b_1)x_0 + c_2 \exp(d_2 x_0) - \frac{b_2 + c_2 d_2 \exp(d_2 x_0) - b_1}{d_1 \exp(d_1 x_0)} \exp(d_1 x_0) \tag{17}$$

3. Experimental

3.1. Procedure

The procedure proposed may be formulated for two cases

3.1.1. Case of Eq. (9)

1. The vector of parameter $\boldsymbol{p}$ in Eq. (9) and the vector of variances var($\boldsymbol{p}$) is estimated with the use of the nonlinear regression procedure.
2. With the criterion (10) a test of the parameters c_1, c_2 in Eq. (9) is applied to investigate their significance at the significance level α.
3. The insignificant quadratic terms is excluded from Eq. (9) and the regression analysis is repeated.
4. The confidence interval of x_{ep} is calculated using Eq. (8).

3.1.2. Case of Eq. (12)

1. Parameters a_1 through d_2 and x_0 are computed using nonlinear regression.
2. The end-point x_{ep} is computed using (16).
3. The law of error propagation is used to estimate the confidence interval of the end-point x_{ep} (Fig. 4b).

3.2. Software

The method of the nonlinear regression was used with the nonlinear regression module of the QC-Expert 2.0^{TM} [19] or ADSTAT 2.0^{TM} [20] statistical packages and checked also with the use of S-Plus^{TM} package [3]. The full, ready-to-run source for S-Plus^{TM} is included in Appendix A. Generally, a good convergence of parameters estimates in nonlinear regression procedures depends on the distance of parameter estimates from optimal values and the shape of an elliptic hyperparaboloid $U(\boldsymbol{p})$ in $(r + 1)$-dimensional space. Since both models (1) and (9) are nearly linear, estimation of $\boldsymbol{p}$ is relatively easy. In S-Plus^{TM} a modified Gauss–Newton algorithm [3] was used to find $\boldsymbol{p}$ and var($\boldsymbol{p}$).

4. Results

4.1. Study case 1: reliability of the end-point determination in case of two straight line segments

Two straight lines of a titration curve were calculated with an intersection 5.90 ml and resulting points were loaded with a noise generated normal random error $N(0, 0.01)$ with the standard deviation $s = 0.1$. Reliability i.e. accuracy and precision of the estimated end-point are examined when two strategies, algorithmic and heuristic, of nonlinear regression model building are applied to both straight line segments of titration curve.

1. An algorithmic search of the regression model made automatically (Fig. 1a) found that both segments of titration curve are linear and the estimated end-point being 5.90 ml with the 95% confidence interval [$L_L = 5.82$ ml and $L_U = 5.99$ ml]. As a true value 5.90 ml lays in the calculated confidence interval, the estimated end-point is *accurate* and unbiased. Narrow confidence interval L_L and L_U means *precise* estimation.

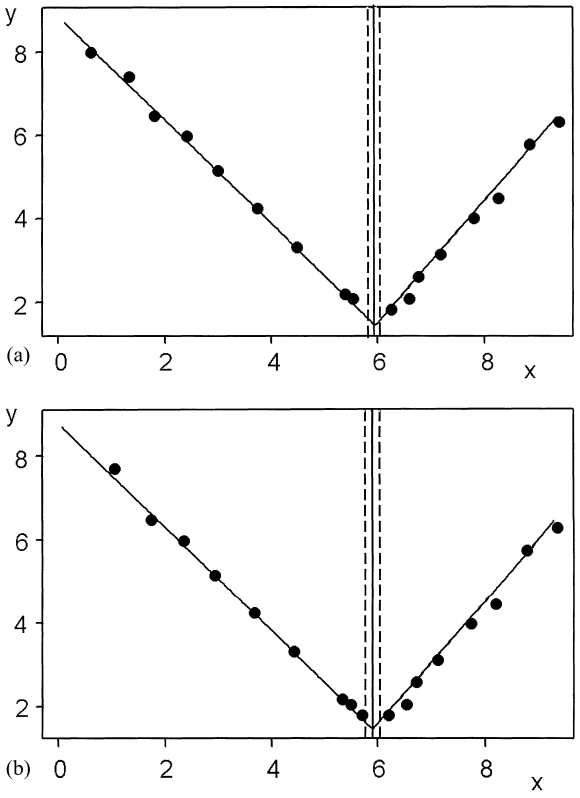

Fig. 1. (a) Algorithmic operation: the regression model with both linear segments is found; data: generated linear data with a true value of the end-point 5.90 ml; estimated end-point: 5.90 ml with the 95% confidence interval $L_L = 5.82$ ml and $L_U = 5.98$ ml; conclusion: accurate and precise estimation. (b) Heuristic operation: the regression model with both quadratic segments is forced; data: generated linear data with a true value of the end-point 5.90 ml; estimated end-point: 5.90 ml with the 95% confidence interval $L_L = 5.76$ ml and $L_U = 6.03$ ml; conclusion: accurate and imprecise estimation.

2. When heuristically operated the regression model with both quadratic segments made manually (Fig. 1b) is forced, the estimated end-point was 5.90 ml with the 95% confidence interval ($L_L = 5.76$ ml and $L_U = 6.03$ ml). As a true value 5.90 ml lays in the calculated confidence interval, the estimated end-point is *accurate* and unbiased. Wider confidence interval L_L and L_U means rather *imprecise* estimation.

It may be concluded that both models tested led to unbiased value of the end-point but in case of quadratic segments the 95% confidence interval is rather wider and more uncertain than in case of model with linear segments.

4.2. Study case 2: reliability of the end-point determination in case of two curved segments

Two quadratic segments of a titration curve were generated with an intersection 3.60 ml and resulting points were loaded with a noise generated normal random error $N(0, 0.04)$ with the standard deviation $s = 0.2$. Reliability, i.e. an accuracy and precision, of the estimated end-point are examined when two strategies, algorithmic and heuristic, of nonlinear regression model building are applied to both curved segments of titration curve.

1. An algorithmic search of the regression model made manually (Fig. 2a) found that both segments of titration curve are quadratic and the estimated end-point being 3.63 ml with the 95% confidence interval ($L_L = 3.41$ ml and $L_U = 3.85$ ml). As a true value 3.60 ml lays in the calculated confidence interval, the estimated end-point is *accurate* and unbiased. Narrow confidence interval L_L and L_U means *precise* estimation.
2. When heuristically operated the regression model made manually (Fig. 3b) with both linear segments is forced and the estimated end-point being 2.90 ml with the 95% confidence interval (2.60 and 3.19 ml). As a true value 3.60 does not lay in the calculated confidence interval, the estimated end-point is *inaccurate* and biased. Wide confidence interval L_L and L_U means *imprecise* estimation.

4.3. Study case 3: analysis of experimental data in conductimetry

Data from a conductometric titration of 0.1 M HCl with 0.1 NaOH, where x is the volume of 0.1 M NaOH in ml, $y = (1000 - a)/a$, in which a is the bridge reading in mm [5].

x (ml)	2	4	6	8	10	12	14	16	17	18	20	22	24
y (mm)	1.265	1.141	1.028	0.906	0.777	0.641	0.510	0.372	0.388	0.441	0.544	0.644	0.752

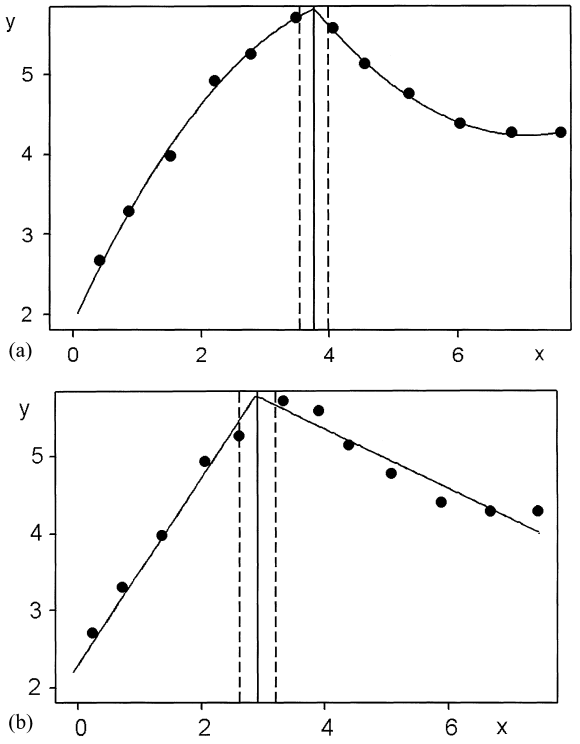

Fig. 2. (a) Algorithmic operation: the regression model with both quadratic segments is found; data: generated nonlinear data with a true value of the end-point 3.60 ml; estimated end-point: 3.63 ml with the 95% confidence interval $L_L = 3.41$ ml and $L_U = 3.85$ ml; conclusion: accurate and precise estimation. (b) Heuristic operation: the regression model with both linear segments is forced; data: generated data with a true value of the end-point 3.60 ml; estimated end-point: 2.90 ml with the 95% confidence interval $L_L = 2.60$ ml and $L_U = 3.19$ ml; conclusion: inaccurate and imprecise estimation.

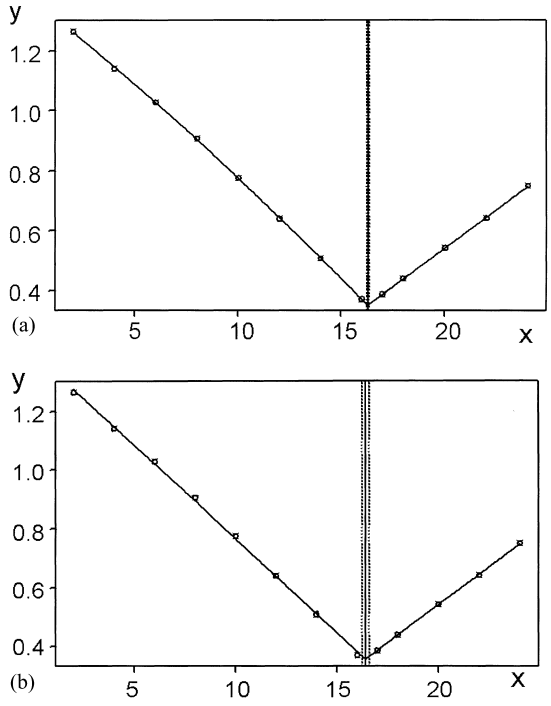

Fig. 3. (a) Algorithmic operation: the regression model with quadratic-linear segments is found; data: experimental data [5]; estimated end-point: 16.28 ml with the 95% confidence interval $L_L = 16.22$ ml and $L_U = 16.35$ ml; conclusion: accurate and precise estimation, the residual sum of squares for the left segment $U(\boldsymbol{p}) = 0.0001047$. (b) Heuristic operation: the regression model with both linear segments is forced; data: experimental data [5]; estimated end-point: 16.41 ml with the 95% confidence interval $L_L = 16.24$ ml and $L_U = 16.58$ ml; conclusion: inaccurate and imprecise estimation; the residual squares-sum for the left segment $U(\boldsymbol{p}) = 0.001057$.

The algorithm found that the data have the left branch of nonlinear nature.

1. An algorithmic search of the regression model made automatically (Fig. 3a) found that the left branch of a titration curve is quadratic and the estimated end-point being 16.28 ml with the quite narrow 95% confidence interval ($L_L = 16.22$ ml and $L_U = 16.35$ ml).
2. When heuristically operated the regression model with both linear segments (Fig. 3b) is forced and the estimated end-point being 16.41 ml with the broader 95% confidence interval ($L_L = 16.24$ ml and $L_U = 16.58$ ml).

It may be concluded that when fitting both segments by two straight lines a biased end-point estimate is obtained with three times broader confidence interval (*imprecise* estimation) and 10 times higher value of the residual sum of squares comparing to a model with the quadratic left branch. The systematic error of a forced linear–linear model is +0.13 ml (i.e. 0.8%) and the estimation is, therefore, *inaccurate*.

Used titration curve was analyzed previously by other authors: Liteanu and Hopirtean [5] supposed both segments linear and found the point estimate for the end-point 16.41 ml with the 95% interval estimate ($L_L = 16.27$ ml and $L_U = 16.57$ ml). Jandera

and co-workers [2] approximate by two straight lines the hyperbolae which delimit the confidence interval of the intersection point being formed by two equal segments and found the end-point 16.41 ml with the 95% interval estimate ($L_L = 16.23$ ml and $L_U = 16.59$ ml).

4.4. Study case 4: analysis of experimental data in amperometry

Data from an amperometric titration of Pb^{2+} with 0.1 M CrO_4^{2-}, where x is the volume of CrO_4^{2-} (μl) and y is the corrected current (μA) [12]

x (μl)	0	100	200	300	400	500	600	700	800	900
y (μA)	311.75	290.37	255	223.51	188.96	157.92	126.29	98.73	78.48	57.11
x (μl)	1000	1100	1200	1300	1400	1500	1600	1700	1800	
y (μA)	49.25	71.37	113.59	158.9	207.48	261.91	310.59	355.97	417.72	

Using the model (14) and (15), we received the best least squares estimates of the parameters with their standard deviations, from which x_{ep} was calculated as the interception of the asymptotic lines according to Eq. (16) $x_{ep} = 977.3$ μl. Using simple error propagation law with covariances the standard deviation of x_{ep} was calculated, $s_{ep} = 5.7$ μl and the 95% confidence interval of x_{ep} ($L_L = 966.0$ μl and $L_U = 988.6$ μl), which is four times more precise than estimated in [12] (Fig. 4a and b).

5. Discussion

In practice, one aspect of titration methodology in analytical chemistry has troubled a chemist: the difficulty of assigning an uncertainty estimate to an end-point determined from a titration curve. Nonlinear regression methods are a well-known means of estimating parameter uncertainties and may be applied to an end-point determination. Quite useful seems to be the $100(1-\alpha)$% confidence interval, which may be interpreted as an interval in which (in case of $\alpha = 0.05$) in average 19 out of 20 results would fall if the analysis is repeated many times under the same conditions. So, it is necessary to expect that the real value (e.g. a concentration) to be *anywhere* within this interval. Described technique was used to evaluate the confidence interval of the end-point in photometric titration. Thanks to possible quadratic shape of a branch, also nonlinear shapes may be analyzed with some restrictions.

Examples bring a fact that the quadratic fit on linear data when a model is over-determined, causes wider confidence interval of the end-point. On the other hand, linear fit on the nonlinear data when a model is under-determined causes biased end-point estimate and a wider confidence interval. For most of the data tested, this method with an automatic discrimination of parameters gave good results.

It is not advisable to use this method for exponential data ranging across more than one order. In such cases the errors are usually heteroscedastic (i.e. non-constant variance) and the least squares may not give correct results and the weighted regression or multiplicative-error models should be used instead. It is recommended that the graph be viewed every time to ensure absence of big errors in y or x. For automated analysis, a robust modification of the regression analysis such as the L_1 method, the method of trimmed squares, the method of M-estimates, LMS (cf. [3]) being enabled in the software ADSTAT, QC-Expert or S-Plus™ may be employed to reduce the problem with outliers.

The described method of fitting V-shape titration curves with conditional models may also be applied to the titration lines with curvature around the end-point. Curvature is usually caused by deviations from ideal behavior of analyte in solution or interference of other ions at very low concentrations of the measured ion near the end-point. Efforts have been made to describe titration curves analytically, but much simpler linearization techniques are still being widely used with

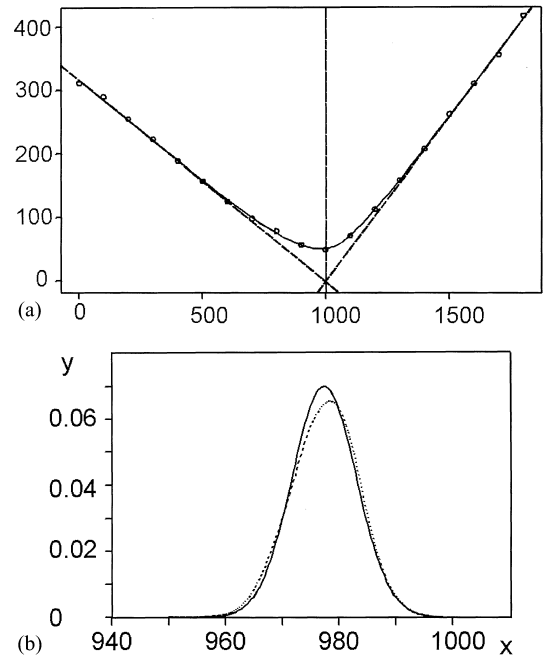

Fig. 4. (a) Fitted titration curve (14) and (15) with analytical asymptotic lines; data: experimental data [12]; estimated end-point: 977.3 μl with the 95% confidence interval $L_L = 966.0$ μl and $L_U = 988.6$ μl. (b) Monte Carlo — generated probability density of the estimate (dashed) with the normal curve was used for estimation of end-point confidence interval using an error propagation law [18]; data: experimental data [12].

good results [5,6,9]. This shape is usual, for example, for the widely used Gran's linearization method [6–8,10]. End-point is at the intersection of the linear or linearized branches of the titration curve. Methods used to find the two lines are usually based on treating the nonlinear part of the measured dependence signal — volume as outliers and deleting them using robust or other methods [12]. With these methods, however, we suppose that part of the data were measured with a great one-sided error with non-constant variance (heteroscedatic errors with non-zero mean), which is not true and may lead to a loss of correctly measured data and to biased estimates of the lines, especially when using weighted least squares linear regression. For the end-point volume computed in this way it is also difficult to calculate its confidence interval or uncertainty, when only a small portion of data obey linearity.

6. Conclusion

The described method has been proven on various examples of titration curves in photometric, conductometric and amperometric titrimetry. The automatic model selection based on *t*-test of the quadratic term ensures narrowest confidence intervals of the end-point and avoids misleading interpretations of near-linear curves as linear as it often happens using manual or visual line-based methods. Success of the end-point calculation depends heavily on the quality of an optimization algorithm and on the starting estimates of parameters. In the proposed algorithm in the S-Plus™ the estimates are generated automatically.

Appendix A. Algorithm in S-Plus™

This algorithm is written in a simple way to allow even a reader not familiar with S-Plus™ to understand it. Nevertheless, it runs quite fast on PC-586 or Unix-based computer. The "#" starts comment. The procedure requires only the measured data in an ($N \times 2$) matrix. Typical sample (Example 3 cited above) with the data input and the results output is enclosed. The S-Plus™ function **nls** solved most of the problems. For some data sets, however, alternative estimates had to be entered manually to reach convergence.

A.1. Main part of program

fit.two.branches_function(*data*, alpha = 0.05, linr = NULL, graf = T)

A.2. Symbols

data	means an ($n \times 2$) matrix containing vector *x* and *y*
alpha	means confidence level for interval of *p* (0, 1)
linr	means which branches are linear? (both, left, right, none)
NULL	means that program finds the best model automatically
graph	means Draw graph? (T, F)
two.branches	is the model function used by fit.two.branches

A.3. Execution

fit.two.branches(*titr*), data are denoted **titr**.

Example: data and execution of the Example 3. *Analysis of an experimental data of the conductometric titration curve*

```
xx_c(2,4,6,8,10,12,14,16,17,18,20,22,24)
yy_c(1.265,1.141,1.028,0.906,0.777,0.641,0.510,0.
  372,0.388,0.441,0.544, 0.644,0.752)
a_cbind(xx,yy)
```

A.4. Program listing

```
two.branches <- function(x, r, linear = "both")
{
######## linear - linear ##########################
# r[1:4] ... parameters (1=p, 2=a1, 3=a2, 4=b1)
# p ... intersection of the branches
# b1,b2 ... slopes of both lines
# a1,a2 ... absolute terms of both lines
    if(linear == "both") {
        p <- r[1]; a1 <- r[2]; a2 <- r[3]; b1 <- r[4]; b2 <- (a1 - a2)/p + b1
        z <- 1:length(x)
        for(i in 1:length(x)) {
            if(x[i] < p)
                z[i] <- a1 + b1 * x[i]
            else z[i] <- a2 + b2 * x[i]
        }
# END FOR
    }
# END IF

######## curved - curved ##########################
# r[1:6] ... parameters (1=p, 2=a1, 3=a2, 4=b1, 5=c1, 6=c2)
# p ... intersection of the branches
# a1,a2 ... absolute terms of the branches
# b1,b2 ... linear terms of the branches
# c1,c2 ... quadratic terms of the branches
    if(linear == "none") {
        p <- r[1]; a1 <- r[2]; a2 <- r[3]; b1 <- r[4]; c1 <- r[5]; c2 <- r[6]
        b2 <- b1 + (a1 - a2)/p + (c1 - c2) * p
        z <- 1:length(x)
        for(i in 1:length(x)) {
            if(x[i] < p)
                z[i] <- a1 + b1 * x[i] + c1 * x[i]^2
            else z[i] <- a2 + b2 * x[i] + c2 * x[i]^2
        } # END FOR
    } # END IF

######## curved - linear ##########################
# r[1:5] ... parameters (1=p, 2=a1, 3=a2, 4=b1, 5=c1)
# p ... intersection of the branches
# a1,a2 ... absolute terms of the branches
# b1,b2 ... linear terms of the branches
# c1  ... quadratic terms of the branches
    if(linear == "right") {
        p <- r[1]; a1 <- r[2]; a2 <- r[3]; b1 <- r[4]; c1 <- r[5]
        b2 <- b1 + (a1 - a2)/p + c1 * p
        z <- 1:length(x)
        for(i in 1:length(x)) {
            if(x[i] < p)
                z[i] <- a1 + b1 * x[i] + c1 * x[i]^2
            else z[i] <- a2 + b2 * x[i]
        } # END FOR
    } # END IF

######## linear - curved ##########################
# r[1:5] ... parameters (1=p, 2=a1, 3=a2, 4=b1, 5=c2)
# p ... intersection of the branches
# a1,a2 ... absolute terms of the branches
# b1,b2 ... linear terms of the branches
# c2  ... quadratic terms of the branches
```

```
    if(linear == "left") {
        p <- r[1]; a1 <- r[2]; a2 <- r[3]; b1 <- r[4]; c2 <- r[5];
        b2 <- b1 + (a1 - a2)/p - c2 * p
        z <- 1:length(x)
        for(i in 1:length(x)) {
            if(x[i] < p)
                z[i] <- a1 + b1 * x[i]
            else z[i] <- a2 + b2 * x[i] + c2 * x[i]^2
        } # END FOR
    } # END IF
    z
} # END PROC



fit.two.branches <- function(data, alpha = 0.05, linr = NULL, graf = T)
{
# data ... matrix [N x 2]
# alpha ... confidence level
# lin ... both, left, right, none ... which branches are linear?
    dimnames(data) <- list(NULL, c("x", "y"))
    N <- dim(data)[1]
    tt <- as.data.frame(data)
    equiv <- median(data[, 1])    # first estimate of the end-point (intersection)

    if(mode(linr) == "NULL")
    {
    pp <- c(equiv, 0, 1, 1, 0.1, 0.5)
    r <- nls(y ~ two.branches(x, pp, linear = "none"), tt, list(pp = pp)) # Regression
    c1_abs(summary(r)$parameters[5, 1])
```

```
    c2_abs(summary(r)$parameters[6, 1])
    equiv_abs(summary(r)$parameters[1, 1])
    sigm1 <- summary(r)$parameters[5, 2]   # Standard deviation of 1st quadratic term
    sigm2 <- summary(r)$parameters[6, 2]   # Standard deviation of 2nd quadratic term
    tvalue <- qt(1 - alpha/2, N - 1)
    if ((sigm1*tvalue>c1)&&(sigm2*tvalue>c2)) linr_"both"
    if ((sigm1*tvalue>c1)&&(sigm2*tvalue<c2)) linr_"left"
    if ((sigm1*tvalue<c1)&&(sigm2*tvalue>c2)) linr_"right"
    if ((sigm1*tvalue<c1)&&(sigm2*tvalue<c2)) linr_"none"
    }

cat("\n")
cat("Linear Branches     :", linr, "\n")
cat("Equivalence         :", equiv, "\n")

    if(linr == "both")
{ pp <- c(equiv, 0, 0.1, 1)
r <- nls(y ~ two.branches(x, pp, linear = "both"), tt, list(pp = pp))
}
    if(linr == "none")
{ pp <- c(equiv, 0, 1, 1, 0.1, 0.5)
r <- nls(y ~ two.branches(x, pp, linear = "none"), tt, list(pp = pp))
}
    if(linr == "left")
{ pp <- c(equiv, 0, 1, 1, 0.1)
r <- nls(y ~ two.branches(x, pp, linear = "left"), tt, list(pp = pp))
}

    if(linr == "right")
{ pp <- c(equiv, 0, 1, 1, 0.1)
```

```
r <- nls(y ~ two.branches(x, pp, linear = "right"), tt, list(pp = pp))
}

    equiv <- r$parameters[1]          # Computed point estimate of the end-point
    sigm <- summary(r)$parameters[1, 2]       # and its standard deviation
    tvalue <- qt(1 - alpha/2, N - 1)
    LL <- equiv - sigm * tvalue
    UU <- equiv + sigm * tvalue
    p <- r$parameters
    if(graf) {                    #plotting graph
        plot(data)
        xx <- seq(min(data[, 1]), max(data[, 1]), length = 200)
        lines(xx, two.branches(xx, r$parameters,linr))
        abline(v = equiv, col = 2)
        abline(v = c(LL, UU), col = 3, lty = 2)
    }
    cat("\nEnd-point      :", equiv, "\nConfidence Interval",
    100 * (1 - alpha),"%:", LL, UU, "\n")
    cat("Linear Branches    :", linr, "\n")
}
fit.two.branches(a)
```

References

[1] M. Meloun, J. Militký, M. Forina, Chemometrics for Analytical Chemistry, Vol. 2, PC-Aided Regression and Related Methods, Ellis Horwood, Chichester, 1994.

[2] P. Jandera, S. Kolda, S. Kotrlý, End-point evaluation in instrumental titrimetry, II. Confidence intervals in extrapolation of linear titration curves, Talanta 17 (1970) 443–454.

[3] J.M. Chambers, T.J. Hastie, Statistical Models, Chapman & Hall, New York, 1993 (S-Plus™ is a trademark of Mathsoft Ltd.).

[4] V. Jehlička, V. Mach, Determination of estimates of parameters of calibration straight line with objective elimination of remote measurements, Collect. Czech. Chem. Commun. 60 (1995) 2064–2073.

[5] C. Liteanu, E. Hopirtean, Studia Univ. Babes-Bolyai, Ser. Chem. 11 (1) (1966) 135.

[6] Li. Heng, Improvement of Gran's method in standard addition and subtraction methods by a new plot method, Anal. Lett. 24 (1991) 473.

[7] N. Akimoto, H. Hanakuma, K. Hozumi, Errors in acid–base titration using Gran's plot method, Anal. Sci. 3 (1987) 515.

[8] E. Still, Determination of the equivalence-point in potentiometric titrations with Gran's first method used to test the electrode response, Anal. Chim. Acta 107 (1979) 377.

[9] C. Liteanu, I. Rica, V. Liteanu, Confidence interval of equivalence point in linear titrations, Talanta 25 (1978) 593.

[10] G. Gran, Analyst 77 (1952) 66.

[11] S.R. Goode, Computerized curve-fitting to determine equivalence point in spectrophotometric titrations, Anal. Chem. 49 (1977) 1408.

[12] M.C. Ortiz-Fernández, A. Herrero-Gutiérrez, Regression by least median of squares, a methodological contribution to titration analysis, Chemometr. Intell. Lab. Syst. 27 (1995) 231.

[13] N.B. Milic, Z.M. Durisic, A computer program GEZ for determination of the equivalence point of the acid–base titration, and E^0 of the glass electrode, Anal. Chim. Acta 331 (1996) 23.
[14] L.M. Schwartz, Uncertainty of a titration equivalence point — a graphical method using spreadsheet to predict values and detect systematic errors, J. Chem. Educ. 69 (1992) 879.
[15] R. Delevie, Explicit expressions of the general-form of the titration curve in terms of concentrations — writing a single closed-form expression for the titration curve for a variety of titrations without using approximations or segmentation, J. Chem. Educ. 70 (1993) 209.
[16] T. Moisio, M. Heikonen, Expressions of the general-form of the acid–base titration curve, Fresenius' J. Anal. Chem. 356 (1996) 461.
[17] D. Ceaucescu, E.V. Ceaucescu, Equivalence point and confidence-interval of linear titration curves in outlook of normal bidimensional distribution of branches, Revue Roum. Chimie 22 (1977) 563.
[18] M. Meloun, J. Militký, M. Forina, Chemometrics for Analytical Chemistry, Vol. 1, PC-Aided Statistical Data Analysis, Ellis Horwood, Chichester, 1992.
[19] TriloByte, QC-Expert 2.1^{TM}, User Manual, TriloByte Statistical Software Ltd, Pardubice, Czech Republic, 1999 (QC-Expert 2.1^{TM} is a trademark of TriloByte Statistical Software Ltd., Pardubice, http://www.trilobyte.cz).
[20] ADSTAT 2.0^{TM} User Manual, TriloByte Statistical Software Ltd., Pardubice, Czech Republic, 1992 (ADSTAT 2.0^{TM} is a trademark of TriloByte Statistical Software Ltd., Pardubice).