



# Rational function ridge regression in kinetic modeling: A case study<sup>☆</sup>

Veli-Matti Taavitsainen

Helsinki Metropolia University of Applied Sciences, PL 4000, 00079 Metropolia, Finland

## ARTICLE INFO

### Article history:

Received 24 May 2012

Received in revised form 13 October 2012

Accepted 3 November 2012

Available online 12 November 2012

### Keywords:

Rational functions

Ridge regression

Kinetic modeling

Esterification

Parameter estimation

## ABSTRACT

Rational functions appear in many theoretical formulae in chemistry, chemical engineering, and biochemistry. However, rational functions have not been used much in empirical multivariate modeling. This study is a continuation to developments in [1] and [3]. In this study, the focus is on empirical modeling using rational function in a case study of estimating the kinetics of esterification of ethanol with acetic acid. The results are compared with those obtained by traditional mechanistic modeling using nonlinear parameter estimation. The motivation for the approach is to find good models with less effort. Currently, developing and testing different mechanistic kinetic models are time consuming, and in spite of the effort, the performance of the resulting candidate models can be virtually the same. An alternative approach is to use empirical generic models. In this study, this approach is shown to require less effort than the mechanistic modeling approach.

The proposed method is based on modeling the reaction rate by an empirical second order rational function. The rational function is first transformed into a linear form which is used to estimate the unknown model parameters using ridge regression. After this, the linear model is back-transformed into the original rational form which is then used for calculating both the fitted values and the predicted values. Finally, the estimated rational function for the rate expression is used to solve numerically the system of differential equations corresponding to the reaction kinetics.

The data is obtained from [4], and it consists of several batch reactions between ethanol and acetic acid with different initial concentrations. Six of the eight batches are used for model estimation, and two of these are omitted for model validation (prediction).

The results are very promising, and both the fitted and the predicted values are comparable to those obtained using traditional mechanistic modeling with nonlinear parameter estimation. It also seems that the method is able to take into account the non-ideal behavior of the reactants. The proposed method offers a flexible and fast method for developing kinetic models for reactor design and control, even in cases where the kinetic rate expression is not known by theory.

© 2012 Elsevier B.V. All rights reserved.

## 1. Introduction

Almost all theoretically known dependencies in physics, chemistry and biology are nonlinear. Although linear models are often useful local approximations, there are cases where taking nonlinearity into account is unavoidable. The most common forms of nonlinearity in chemistry, or in biological systems, are exponential, logarithmic, logistic and rational functional forms. Many of them can be put into linear regression form using some linearizing transformation. For rational functions, the 'natural' linearizing transformation is to multiply both sides by the denominator, and rearrange the terms. However, this form is inapplicable for prediction because the new predictor terms

contain the dependent variable. In principle, this problem is easily overcome by carrying out the regression using the linearized form, using the original rational form for predictions. Unfortunately, this well-known trick works poorly in cases that contain several independent (explanatory) variables because of extreme collinearity. A natural solution to the collinearity problem is to use ridge regression, or some other constrained regression. This approach is called rational function ridge regression (RRR).

Rational type of nonlinearities is very common in chemistry, especially in catalytic reactions, and rational function ridge regression has already been shown to be a potential method for nonlinear empirical (soft) modeling, demonstrated in [1–3]. The applications tested in [1–3] include multivariate calibration, QSAR (quantitative structure–activity relationship) studies, calibration of sensory data and RSM (response surface modeling). These studies also include comparisons to other nonlinear empirical methods.

The objectives of this study are: 1) to continue testing the applicability of RRR in applications between hard and soft modeling, and 2) to develop a fast method for kinetic modeling.

<sup>☆</sup> Paper presented at the Eighth Winter Symposium on Chemometrics (WSC8), Drakino, Russia, 27th February to 2nd March, 2012.

E-mail address: [Veli-Matti.Taavitsainen@metropolia.fi](mailto:Veli-Matti.Taavitsainen@metropolia.fi).

## 2. Theory

A rational function has the form

$$y = \frac{P_m(x, \beta_p)}{Q_n(x, \beta_Q)}. \quad (1)$$

$P_m(x, \beta_p)$  and  $Q_n(x, \beta_Q)$  are polynomials of the independent variables given in the vector  $\mathbf{x}$ . The vectors  $\beta_p$  and  $\beta_Q$  contain the coefficients of the polynomials  $P$  and  $Q$  of orders  $m$  and  $n$ , typically  $m \leq 2$  and  $n \leq 2$ .

Direct estimation of the parameters of rational functions requires iterative nonlinear estimation algorithms, where finding good initial estimates may be problematic. Our approach is based on a linearizing transformation and ridge regression (or some other constrained regression), called RRR.

Because any fraction can be reduced by canceling out any constant, we can assume that the intercept term in the denominator  $Q_n(x, \beta_Q)$  is 1. Let us denote such a polynomial by

$$Q_n = 1 + \tilde{Q}_n(x, \beta_Q), \quad (2)$$

i.e. the intercept  $\beta_{Q0}$  in  $\tilde{Q}_n(x, \beta_Q)$  is zero. Multiplying both sides of Eq. (1) by the denominator gives

$$y = P_m(x, \beta_p) - y\tilde{Q}_n(x, \beta_Q). \quad (3)$$

This is a linear model with respect to the polynomial coefficients  $\beta_p$  and  $\beta_Q$ , and the regressor terms are the powers in  $P$  and the powers of  $Q$  multiplied by  $-y$ . The common factor  $y$  in  $Q$  terms makes the problem strongly, and unavoidably collinear and inapplicable for prediction of new observations. Thus, the estimation must be carried out using some other method than OLS (ordinary least squares) regression, and predictions have to be calculated using the back-transformed model, i.e. the original rational function model.

Several alternative methods can be used in estimation of the polynomial coefficients, including ridge regression, PLSR (partial least squares regression), LASSO, LAR (least angle regression) and ENET (elastic nets). LASSO and LAR are  $L_1$  norm minimization based methods, and ENET is a combination of  $L_1$  norm and  $L_2$  norm minimization. The last three alternatives in RRR have recently been studied in [2] in a case of multivariate calibration of spectral measurements. In contrast, the current study is restricted to ridge regression only.

After obtaining the estimates of  $\beta_p$  and  $\beta_Q$ , say  $\hat{\beta}_p$  and  $\hat{\beta}_Q$ , the linearized model is back-transformed into the original form

$$\hat{y} = \frac{P_m(x, \hat{\beta}_p)}{1 + \tilde{Q}_n(x, \hat{\beta}_Q)}. \quad (4)$$

The fitted values and future predictions are calculated using this formula. To be more explicit, let us consider the simplest case of first order polynomials for two variables:

$$y = \frac{b_0 + b_1x_1 + b_2x_2}{1 + c_1x_1 + c_2x_2}. \quad (5)$$

The corresponding linear model is

$$y = b_0 + b_1x_1 + b_2x_2 + d_1x_1y + d_2x_2y, \quad (6)$$

where  $d_i = -c_i, i = 1, 2$ . The predicted and fitted values can be calculated by

$$\hat{y} = \frac{\hat{b}_0 + \hat{b}_1x_1 + \hat{b}_2x_2}{1 - \hat{d}_1x_1 - \hat{d}_2x_2}, \quad (7)$$

where  $\hat{b}_i$  and  $\hat{d}_i$  are the estimates for the  $b$ s and  $d$ s after ridge regression. Consequently  $\hat{c}_i = -\hat{d}_i$  and the denominator of Eq. (7) could be expressed as  $1 + \hat{c}_1x_1 + \hat{c}_2x_2$  in concordance with Eq. (5).

Of course, with orders higher than one, one choice for the denominator could be a polynomial without linear terms, which is a common type of a model in heterogeneous catalysis, for example. In high dimensional problems, polynomials with interaction terms will typically introduce too many unknown parameters. Full quadratic polynomials are feasible when the number of independent variables is small enough, or the data has been compressed, or some variable selection scheme has been employed.

There are several alternatives in choosing the ridge penalty parameter. The chosen ones given below use either the coefficient of de-

termination  $R^2 = 1 - \frac{\sum_i (y_i - \hat{y}_i)^2}{\sum_i (y_i - \bar{y})^2}$ , where the  $\hat{y}_i$ s stand for fitted values

and  $\bar{y}$  stands for the mean values of the measured responses, or the cross-validated coefficient of determination which is defined similarly, but so that  $\hat{y}_i$ s stand for cross-validatory predicted values.

- 1) Choose the penalty parameter value that gives the smallest  $Q^2$ -value in the linear regression step ( $Q_L^2$ ), i.e. in estimating the parameters of the transformed model.
- 2) Choose the penalty parameter value that gives the largest  $R^2$ -value in using the back-transformed model ( $R_{BT}^2$ ).
- 3) Choose the penalty parameter value that gives the largest  $Q^2$ -value in using the back-transformed model ( $Q_{BT}^2$ ).

Alternatives 1 and 2 are fast if deleted residuals are used for leave-one-out cross-validation. The estimation is simply repeated using the chosen candidate values for the penalty parameter. Interestingly, the  $R_{BT}^2$  and  $Q_L^2$  often have maximal values with approximately the same penalty parameter values.

Since the linearized model cannot be used for true predictions, the advantage of the computational efficiency of the leave-one-out cross-validation cannot be applied in assessing the true predictive power of using the back-transformed model. Consequently,  $R^2$ - or  $Q^2$ -values are not the same for linearly and non-linearly (i.e. using the back-transformed model) estimated response values. For this reason, cross-validation should also be carried out for the nonlinearly calculated estimates. Nonlinear cross-validation, unlike linear leave-one-out cross-validation, cannot be completed without iteration, and consequently it is highly computer-intensive. However, experience has shown that the ridge penalty parameter value that maximizes the  $Q^2$ -value of the linear regression step is also quite often very close to the penalty parameter value of maximal  $Q^2$ -value in cross-validating the back-transformed model.

The ridge principle can also be used for variable selection by iteratively suppressing down estimates of coefficients of small absolute value. This principle of iterative suppression can be carried out in many ways, by using penalties that are inversely proportional to the absolute values of the regression coefficients, for example. In this approach, a preset tolerance for the absolute values of the regression coefficients is chosen. All coefficients whose absolute values are below this tolerance receive an iteratively increasing ridge penalty. Some preliminary results of applying this technique will be presented.

## 3. Experimental

As a case study, this investigation utilizes the same data set as in [4]. The data consists of 22 batches of esterification of ethanol and acetic acid at different temperatures, and with different initial concentrations and different amounts of the catalyst.

In [4], 18 batches were used to compare traditional mechanistic modeling to implicit calibration, i.e. for nonlinear kinetic parameter estimation using spectra without explicit calibration with measure

Download English Version:

<https://daneshyari.com/en/article/7563422>

Download Persian Version:

<https://daneshyari.com/article/7563422>

[Daneshyari.com](https://daneshyari.com)