Proceeding of the Korean Nuclear Society Spring Meeting
Kwangju, Korea, May 1997

Automatic Correlation Generation using the Alternating Conditional Expectation Algorithm

Han Gon Kim, Byong Sup Kim, and Sung Jae Cho

Korea Electric Power Research Institute
Center for Advanced Reactors Development
103-16, Munji-Dong, Yusong-Gu,
Taejon, Korea 305-380

Abstract

An alternating conditional expectation (ACE) algorithm, a kind of non-parametric regression method, is proposed to generate empirical correlations automatically. The ACE algorithm yields an optimal relationship between a dependent variable and multiple independent variables without any preprocessing and initial assumption on the functional forms. This algorithm is applied to a collection of 12,879 CHF data points for forced convective boiling in vertical tubes to develop a new critical heat flux (CHF) correlation. The mean, root mean square, and maximum errors of our new correlation are -0.558%, 12.5%, and 122.6%, respectively. Our CHF correlation represents the entire set of CHF data with an overall accuracy equivalent to or better than that of three existing correlations.

1. Introduction

In general, standard non-linear regression analysis is used to develop an empirical correlation from experimental data set. Traditionally this analysis should be performed according to following steps; the initial assumption of functional form, bounding the range of coefficients, and the determination of the fitting coefficient using statistical tools. When we develop the correlation for bivariate problem composed of one indenpendent and one dependent variables, this process is fairly simple. Most of the engineering problems, however, are multi-variate problem composed of one dependent and multi independent variables. In this case, the determination of initial function form is very difficult and sophisticated problem. The correlation developer should have a deep physical insight on the phenomena.

From 1970s, statisticians have investigated on the non-parametric regression analysis which does not require the assumption of initial functional form. The alternating conditional expectation (ACE) algorithm [1, 2] is the most fruitful result among this kind of researches.

The ACE method is a generalized regression algorithm that yields an optimal relationship between a dependent variable y and multiple independent variables $\{x_n, n=1,\cdots,p\}$. The objective of the ACE algorithm is to find optimal transformations $\theta(y)$ and $\{\phi_n(x_n),\ n=1,\cdots,p\}$ which maximize the statistical correlation between $\theta(y)$ and $\sum\phi_n(x_n)$ without a priori estimate of the functional forms $\theta(y)$ and $\phi_n(x_n)$. Once the optimal transformations are obtained, simple regression analysis is performed to determine the functional forms for the transformed dependent and independent variables.

Thus, the ACE algorithm offers considerable advantages over traditional nonlinear regression techniques, which require an initial selection of the functional forms and often iterative modifications of the functional relationships as well. In addition, nonlinear regression analysis requires in general sufficiently accurate estimates of the fitting coefficients to arrive at a converged solution. In contrast, the ACE algorithm guarantees the convergence of the transformations, without the need to provide initial estimates for the transformations, and, once the ACE iteration is converged, simple regression analysis usually suffices to generate actual analytical functional forms for the transformed variables.

In this paper, the ACE algorithm is applied to develop critical heat flux (CHF) correlation because development of CHF correlations is the typical example of non-linear regression analysis. It typically involves iterative estimates of the functional form of the correlation coupled with multivariate nonlinear regression analysis. Even with reasonable understanding of the physical phenomena involved, such correlation development process is tedious and time-consuming, because the functional form has to be properly selected before sufficiently accurate correlations may be attained.

2. The ACE Algorithm

The ACE algorithm was formally derived [1] through a functional analysis approach. Instead of repeating the formal derivation, we begin with a physical justification of the ACE algorithm, which serves as a heuristic derivation of the basic algorithm. The approach we take is based on a physical interpretation of the conditional expectation for a set of discrete data points. We use a bivariate formulation to illustrate the concept and make the necessary extension to multivariate regression problems.

For a bivariate regression problem with a set of N experimental data points $\{ (x_i, y_i),\ i = 1,\cdots, N \}$, we wish to find a transformation $\theta(y)$ of the dependent variable y and a functional fit $\phi(x)$ such that the square error in the regression of $\theta(y_i)$ and $\phi(x_i)$

$$e^2 = \frac{1}{N}\sum_{i=1}^{N}\left[\theta(y_i) - \phi(x_i)\right]^2 \tag{1}$$

is minimized. We assume that the optimal transformations, $\theta(y)$ and $\phi(x)$ exist. And we also assume, without loss of generality, that the optimal transformations, $\theta(y)$ and $\phi(x)$, minimizing Eq. (1) are properly normalized such that $E[\theta(y)] = E[\phi(x)] = 0$ and $E^2[\theta(y)] = 1$.

With a judicious selection of the transformation $\theta(y)$, the error in Eq. (1) could vanish, in principle, if $\theta(y_i)$ equals $\phi(x_i)$ for every point. In practice, however, this idealized situation will not materialize because the experimental data contain random noises and so do $\theta(y_i)$ and $\phi(x_i)$. Thus, a smooth functional representation $\theta(y)$ cannot be equated exactly to $\phi(x)$ at every data point. Instead, $\theta(y_i)$ is considered, in the ACE algorithm, the expectation of several realizations of $\phi(x)$ for the i^{th} point, rather than a single unique realization $\phi(x_i)$ as in conventional regression analysis. Thus, we interpret $\theta(y_i)$ as a conditional expectation $E[\phi(x)|y = y_i]$ to minimize Eq. (1). In most regression problems, in practice, there is usually only one value y_i, and hence one value $\phi(x_i)$, for the i^{th} data point, and the conditional expectation $\theta(y_i)$ has to be evaluated with the neighboring values $\{ \phi(x_j),\ j = i - M,\cdots, i + M \}$, for some M. In the simplest approach, $\theta(y_i)$ could be determined as an arithmetic average of the neighboring data. In general, some kind of weighted average over the neighboring data may be taken as the conditional expectation.

With this smoothing concept, the transformation $\theta(y)$ at the i^{th} point is obtained as $S[\phi(x)|y = y_i]$ instead of exact conditional expectation. We may equivalently consider $\phi(x_j)$ as the conditional expectation $E[\theta(y)|x = x_j]$ = $S[\theta(y)|x = x_j]$. Therefore the optimal transformations, $\theta(y)$ and $\phi(x)$, may be defined as

$$\theta(y) = \frac{S[\phi(x)|y]}{\left\|S[\phi(x)|y]\right\|}, \quad \phi(x) = S[\theta(y)|x]. \tag{2}$$

In practical problem, the smoothing operator, $S[\theta|y]$ can be defined as weighted sum of θ for a given interval of y. The ACE algorithm consists of an iterative use of the two smoothing operations of Eq. (2) in alternating directions.

Based on the bivariate derivation of the ACE algorithm, we can generalize Eq. (2) for a multivariate problem, given a set of experimental data $\{ (y_i, x_{1i}, \cdots, x_{pi}),\ i = 1, \cdots, N \}$:

$$\theta(y) = \frac{S\left(\sum \phi_n(x_n) \middle| y\right)}{\left\| S\left(\sum \phi_n(x_n) \middle| y\right) \right\|}, \quad \phi_n(x_n) = S\left(\theta(y) - \sum_{l \neq n}^{p} \phi_l(x_l) \middle| y \right) \tag{3}$$

The optimal transformations, θ and $\phi_1, \cdots, \phi_p$ cannot be obtained directly because they are coupled to each other through Eqs. (3). Thus, the ACE algorithm requires the following iterative procedures:

1. *Initialization.* $\theta^0(y) = y / \|y\|$ and $\phi_1^0(x_1) = \cdots \phi_p^0(x_p) = 0$.
2. *Inner iteration.* Sort $\theta(y)$ and $\{ \phi_l(x_l),\ l = 1, \cdots, p$ and $l \neq n \}$ in an ascending order of $\phi_n(x_n)$ and evaluate $\phi_n(x_n)$ for iteration step t using the second equation of Eq. (3). And then iterate until squared error fails to decrease.
3. *Outer iteration.* Sort $\{ \phi_n(x_n),\ n = 1, \cdots, p \}$ in an ascending order of $\theta(y)$ and calculate $\theta(y)$ using the first equation of Eq. (3). Continue with step 2 until squared error between $\theta(y)$ and $\phi_n(x_n)$ fails to decrease.

When convergence is attained, the data in each transformed variable are usually smooth and slowly varying. Selecting simple functional forms for the transformations, we perform standard regression analysis for each transformation and finally obtain the functional form of y versus $x_1, \cdots, x_p$ if $\theta(y)$ has an inverse function:

To illustrate the basic concepts of the ACE algorithm, we borrow a simple example from Ref. 1. In this bivariate example, we generate 200 simulated data points from the model

$$y_i = \exp\left(\sin 2\pi x_i + \varepsilon_i / 2\right), \quad i = 1, \cdots, 200, \tag{4}$$

where x_i is sampled uniformly over the interval [0, 1] and the noise ε_i is drawn independently of x_i from a normal distribution N[0,1]. Subject to noise fluctuations, the plot of y versus x in Fig. 1(a) does not render a unique or ready estimate for the functional form.

We make a simple application of the ACE algorithm and obtain converged transformations in 7 iterations with a convergence criterion of 1.0×10^{-5}. Based on the plots of the converged data in the transform domains in Fig. 1(b) and (c), it is logical to assume the functional form

$$\theta(y) = \ln y, \ \phi(x) = \sin 2\pi x, \tag{5}$$

which allows us to retrieve the original functional form.

3. The Generation of CHF Correlation using the ACE Algorithm

We now apply the ACE algorithm to a collection of experimental CHF data so that we obtain a new CHF correlation that is applicable to a broad range of physical parameters. For this purpose, we use the KAIST CHF data bank [3, 4]. 12,879 CHF data are selected to develop CHF correlation for upward water flow in vertical round tube.

The ACE algorithm is a modern data regression tool using various statistical estimators such as mean, standard deviation, and covariance. Thus, it is desirable that input data are statistically well behaved. Some physical parameters in our problem, however, show undesirable statistical behavior. In particular, many data points were obtained at some specific pressure, e.g., near the atmospheric pressure. For this reason, we use non-dimensional parameters instead of physical parameters to arrive at the CHF correlation as following [3]:

$$(x_1, x_2, x_3, x_4, y) = \left(\ln\left(\frac{\sigma \rho_f}{G^2 L}\right), \frac{\Delta h_i}{h_{fg}}, \frac{L}{D}, \ln\left(\frac{\rho_g}{\rho_f}\right), \frac{q_{CHF}}{G h_{fg}} \right) \tag{6}$$

In Eq. (6), first and third variables are taken logarithmic transformation because their parameter ranges are much broader than other variables.

We achieved the converged transformations, with a convergence criterion of 1.0×10^{-5} both for inner and outer iterations, after 9 outer iterations and in approximately 600 minutes of CPU time on a

HP9000/735 machine. We present the transformations, $\phi_n(x_n)$ versus x_n and $\theta(y)$ versus y, obtained through the ACE algorithm in Fig. 2. It is clear from Fig. 2 that the ACE algorithm generates simple functional forms both for the dependent and independent variables. Using the plots of Fig. 2 and a standard regression tool, we can obtain simple analytic functions, $\theta(y)$ and $\{ \phi_n(x_n),\ n = 1,\cdots,4 \}$. For example, $\theta(y)$ can be represented as $a + b\ln y$. For the fitting of other functions, we used piecewise linear functions.

After individual analytic functions of $\theta(y)$ and $\{ \phi_n(x_n),\ n = 1,\cdots,4 \}$ are obtained, we can get the final form of a new CHF correlation through a simple inversion process and a few manipulations as followings:

$$q_{CHF} = A_1 A_2 \left(\frac{\sigma \rho_f}{G^2 L} \right)^{A_3} \left(0.22 + \left(\frac{\Delta h_i}{h_{fg}} \right)^{1.1} \right)^{0.76} \exp\left(0.12 (L / D)^{0.44} \right) \left(\frac{\rho_g}{\rho_f} \right)^{A_4} G h_{fg}, \tag{7}$$

where the coefficients, A_1 through A_4, in Eq. (7) are given as

$$(A_1, A_3) = \begin{cases} 6.47 \times 10^{-3}, 5.68 \times 10^{-2} \ \mathit{for} \ \frac{\sigma \rho_f}{G^2 L} \leq 4.45 \times 10^{-7} \\ 9.41 \times 10^{-2}, 0.24 \ \mathit{for} \ \frac{\sigma \rho_f}{G^2 L} \leq 2.85 \times 10^{-5} \\ 3.33 \times 10^{-2}, 0.14 \ \mathit{else} \end{cases} \qquad (A_2, A_4) = \begin{cases} 0.96, 3.77 \times 10^{-2} \ \mathit{for} \ \frac{\rho_g}{\rho_f} \leq 4.71 \times 10^{-3} \\ 1.45, 0.12 \ \mathit{for} \ \frac{\rho_g}{\rho_f} \leq 3.85 \times 10^{-2} \\ 1.0, 0.0 \ \mathit{for} \ \frac{\rho_g}{\rho_f} \leq 1.26 \times 10^{-1} \\ 2.3, 0.4 \ \mathit{else} \end{cases}$$

The detail process to get Eq. (7) from the results of ACE algorithm shown in Fig. 2 is given in Ref. [5].

4. Testing of the CHF Correlation

In order to test the accuracy of the new CHF correlation represented by Eq. (7), we have performed simulation of the CHF data used in our ACE application. We have also compared our ACE-based correlation with a number of existing CHF correlations.

Our first test consists of simulating the entire set of 12,879 CHF data points. The mean, rms, and maximum errors are -0.558%, 12.5%, and 122.6%, respectively. There are 86.8% of data points or a total 11,174 points, and 96.0% of data points or a total of 12,366 points, within ±20% and ±30% rms errors, respectively. Furthermore, 99.9% of data points or a total of 12,868 points lie within ±80% rms errors although the maximum prediction error is 122.6%.

To assess the overall accuracy of the correlation represented by Eq. (7), we compare our results with the Katto, Bowring, and Biasi correlations which are inlet condition type correlations. Table I shows the prediction errors of the four correlations applied over the whole CHF data points. As shown in the table Eq. (7) shows the best prediction results with respect to rms error. Biasi correlation shows extremly large error because many data are beyond its valid parameter ranges. We can not say that Eq. (7) is better than other correlations because every correlation should be tested within their valid parameter ranges. However, this table provide a clue that the development of CHF correlation using ACE algorithm can be applied in real engineering problem successfully.

Table I. The Overall Comparison of CHF Prediction Errors

Correlation	Mean Error	RMS Error	Max. Error
Katto	6.9	18.7	283.4
Biasi	32.8	-	-
Bowring	-1.8	14.2	143.8
Eq. (7)	-0.6	12.5	122.6

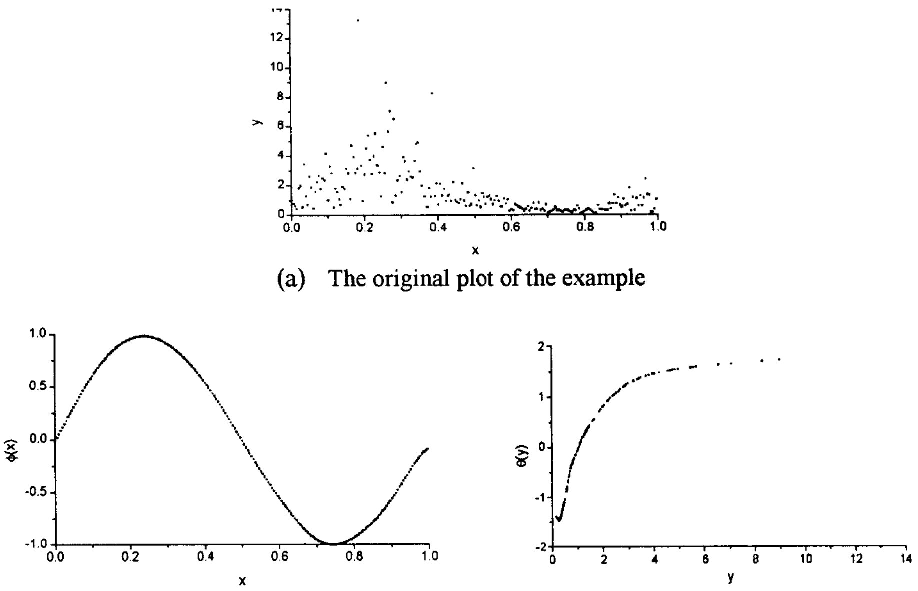
5. Summary and Conclusions

In this paper, we have proposed the new regression method, ACE algorithm and we modified this algorithm so that it can be used large engineering problem. And we have used the ACE algorithm to perform regression analysis of a large set of CHF data and developed a new CHF correlation to verify the performance of the algorithm. With the CHF data cast in terms of five dimensionless variables, a simple application of the ACE algorithm automatically yields optimal transformations, $\theta(y)$ and $\phi_n(x_n)$, $n=1,\cdots,4$. These transformations are obtained without the need to specify initial estimates for the functional forms or the fitting coefficients whatsoever. The new CHF correlation is compared with three well-known CHF correlations reveals that the new correlation represents the entire set of CHF data points with an overall accuracy equivalent to or better than the existing correlations.

The ACE algorithm is a prime example of modern statistical analysis tools which have been made feasible through the advance of powerful digital computers. With the success we have had in our realistic application of the ACE algorithm, we recommend further use of the modern regression tool in complex engineering problems, especially where the underlying physical relationships are not well known. This would, for example, be the case in obtaining parametric representations of the mixed or high-level nuclear waste which is poorly characterized.

References

1. L. Breiman and J. H. Friedman, "Estimating Optimal Transformations for Multiple Regression and Correlation," *Journal of the American Statistical Association, Theory and Method*, 80:391, 580-619 (1985).
2. R. Chen and R. S. Tsay, "Nonlinear Additive ARX Models," *Journal of the American Statistical Association, Theory and Methods*, 88:423, 955-967 (1993).
3. S. K. Moon and S. H. Chang, "Classification and Prediction of the Critical Heat Flux Using Fuzzy Theory and Artificial Neural Networks," *Nucl. Eng. Design*, 150, 151-161 (1994).
4. S.H. Chang et al., "The KAIST CHF Data Bank", *KAIST-NUSCOL-9401*, Korea Advanced Institute of Science and Technology (1994).
5. H. G. Kim and J. C. Lee, "The Development of a Generalized CHF Correlation through Alternating Conditional Expectation Algorithm," Nuclear Science and Engineering, Submitted for Publication.



(a) The original plot of the example

(b) Transformation of x

(c) Transformation of y

Fig. 1 The results of the ACE algorithm for the example problem

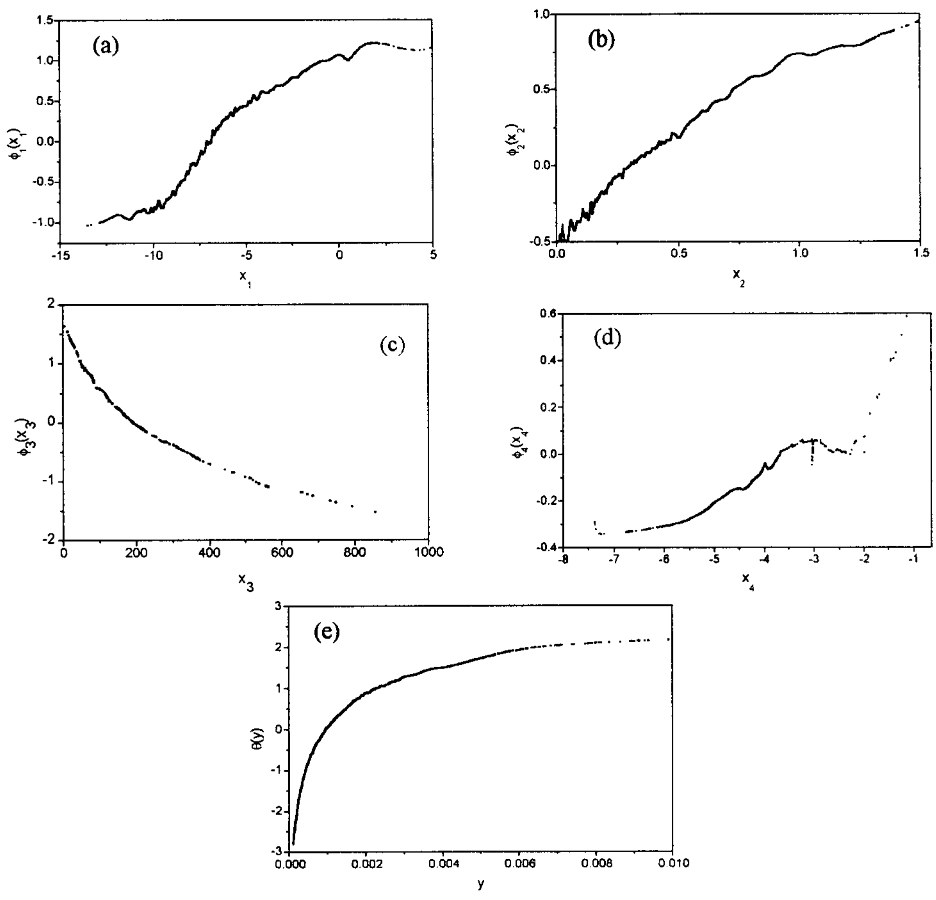


Fig. 2 The results of the ACE algorithm for CHF data