*Production Systems and Information Engineering*
Volume 5 (2009), pp. 177-188.

# GRADIENT-BASED CONSEQUENT OPTIMIZATION OF A FRI RULE BASE

Zoltán Krizsán
University of Miskolc, Hungary
Department of Information Technology
krizsan@iit.uni-miskolc.hu

Szilveszter Kovács
University of Miskolc, Hungary
Department of Information Technology,
Technical University of Kosice, Slovakia
Department of Cybernetics and AI
szkovacs@iit.uni-miskolc.hu

[Received January 2009 and accepted June 2009]

**Abstract.** The main contribution of this paper is the extension of an existing Fuzzy Rule Interpolation (FRI) method by gradient-based consequent optimization. The targeted FRI method is an application oriented approach, called *FIVE* (Fuzzy Rule Interpolation based on the Vague Environment of the Fuzzy Rule Base [1]). The goal of the consequent optimization is the rule base parameter optimization based on input-output sample data of the modelled system.

*Keywords*: Fuzzy Rule Interpolation, FRI, FIVE, gradient-based rule optimization

## 1. Introduction

There are more and more practical applications of Fuzzy Rule Interpolation (FRI) methods appearing in recent literature. Their popularity is based on their ability to handle incomplete fuzzy knowledge representation i.e. 'sparse' fuzzy rule bases. A 'sparse' rule base in this case means a fuzzy rule base, which does not have rules for all the possible observations, in other words, at least one observation may exist which does not lead to an interpretable conclusion applying classical fuzzy reasoning methods (like Zadeh, Mamdani, Larsen, or Takagi-Sugeno). Numerous FRI methods can be found in the literature, and every method has its own advantage. Some of them are very precise, some of them are less precise but their computational complexity is better. In the last few years FRI based systems have been applied successfully for several fuzzy modeling and control tasks.

For example Johanyák, Parhiban and Sekaran [2] developed fuzzy models for an anaerobic tapered fluidized bed reactor, Johanyák and Szabó [3] used a FRI based fuzzy model for tool life prediction depending on cutting parameters in the case of machining operations.

An application oriented aspect of FRI emerges in the concept of FIVE. The fuzzy reasoning method FIVE (originally introduced in [4] and described in [5], [6] and [1]) was developed to fit the speed requirements of direct fuzzy control, where the conclusions of the fuzzy controller are applied directly as control actions in a real-time system (see e.g. a downloadable and executable code of a real-time vehicle path tracking and collision avoidance control at [7]).

Automatic rule base generation is a hard task and most of the available methods are based on a gradient-free approach. The main problem of these methods is the slow convergence compared to gradient-based methods. On the other hand, for gradient-based parameter optimization there has to be a performance measure which is at least partially derivable with the optimizable (tunable) parameters.

In the following, first the FRI method FIVE will be introduced in more detail with the applied Shepard interpolation, then the paper will suggest a gradient-based optimization method (steepest descent) for the automatic consequent optimization of the FIVE rule base. Finally, we illustrate the method with two sample data sets:

1. the training sample is a simple input-output set of randomly selected data from $y = sin(x)/x$,
2. the training data set contains measured porosity values in a real environment corresponding to specified input values (gamma ray, deep induction resistance and sonic travel time).

## 2. The Concept of FIVE

The FIVE FRI method is based on the concept of vague environment [8]. Applying the idea of vague environment, the linguistic terms of fuzzy partitions can be described by scaling functions [8] and the fuzzy reasoning itself can be replaced by classical interpolation. The concept of vague environment is based on the similarity or indistinguishability of the elements considered. Two values in a vague environment are $\epsilon$-distinguishable if their distance is greater than $\epsilon$. The distances in a vague environment are weighted distances. The weighting factor or function is called scaling function (factor) [8]. Two values in the vague environment $X$ are $\epsilon$-indistinguishable if

$$\epsilon \geq \delta_s(x_1, x_2) = \left| \int_{x_2}^{x_1} s(x)dx \right|, \tag{2.1}$$

where $\delta_s(x_1, x_2)$ is the scaled distance of the values $x_1$, $x_2$ and $s(x)$ is the scaling function on $X$.

For finding connections between fuzzy sets and a vague environment the membership function $\mu_A(x)$ can be introduced as an indicating level of similarity of $x$ to a specific element $a$ that is a representative or prototypical element of the fuzzy set $\mu_A(x)$, or equivalently, as the degree to which $x$ is indistinguishable from $a$ (2.2) [8]. The $\alpha$-cuts of the fuzzy set $A$ are the sets that contain the elements that are $(1-\alpha)$-indistinguishable from $a$ (see Fig. 1):

$$\begin{aligned} 1 - \alpha \geq \delta_s(a, b), \mu_A(x) &= 1 - min\{\delta_s(a, b), 1\} \\ &= 1 - min\left\{ \left| \int_b^a s(x)dx \right|, 1 \right\}. \end{aligned} \tag{2.2}$$

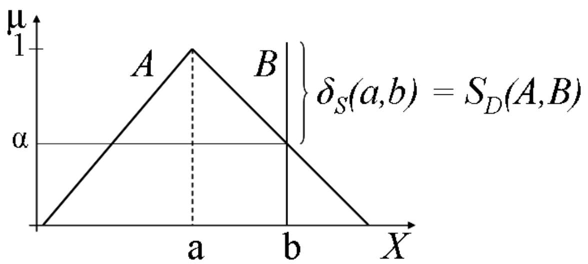


**Figure 1.** The $\alpha$-cuts of $(x)$ contain the elements that are $(1-\alpha)$ indistinguishable from $a$

In this case (see Fig. 1), the scaled distance of points $a$ and $b$ ( $\delta_s(a, b)$ ) is the *Disconsistency Measure (SD)* (mentioned and studied among other distance measures in [9] by Turksen et al.) of fuzzy sets $A$ and $B$ (where $B$ is a singleton).

$$S_D(A, B) = 1 - \sup_{x \in X} \mu_{A \cap B}(x) = \delta_s(a, b) \; if \; \delta_s(a, b) \in [0, 1], \tag{2.3}$$

where $A \cap B$ denotes the min t-norm: $\mu_{A \cap B}(x) = min[\mu_A(x), \mu_B(x)] \forall x \in X$.

Taking into account the most common way of building a traditional fuzzy logic controller where the first step is defining the fuzzy partitions on the antecedent and consequent universes by setting up the linguistic terms and then based on these terms building up the fuzzy rule base, the concept of vague environment [8] is straightforward. The goal of the fuzzy partitions is to define indistinguishability, or vagueness in the different regions of the input-output universes. The vague environment is characterized by its scaling function. For generating a vague environment of a fuzzy partition an appropriate scaling function is needed, which describes the shapes of all the terms in the fuzzy partition. Generally, a fuzzy partition cannot be characterized by a single scaling factor, so the question is how to describe all fuzzy sets of the fuzzy partition with one universal scaling function. For this task, the concept of an approximate scaling function is proposed in [4], [5], [6] as an approximation of the scaling functions describing the terms of the fuzzy partition separately.

## 3. Shepard Interpolation for FIVE

The main idea of the FRI method FIVE can be summarized in the followings:

1. If the vague environment of a fuzzy partition (the scaling function or at least the approximate scaling function) exists, the member sets of the fuzzy partition can be characterized by points in that vague environment. (These points indicate the positions of the fuzzy terms, while the membership functions are described by the scaling function itself.)
2. If all the vague environments of the antecedent and consequent universes of the fuzzy rule base exist, all the primary fuzzy sets (linguistic terms) compounding the fuzzy rule base can be characterised by points in their vague environment. Therefore the fuzzy rules (built-up from the primary fuzzy sets) can also be characterized by points in the vague environment of the fuzzy rule base. In this case, approximate fuzzy reasoning can be handled as a classical interpolation task.
3. Applying the concept of vague environments (the distances of points are weighted distances), any crisp interpolation, extrapolation, or regression method can be adapted very simply for approximate fuzzy reasoning [4], [5] and [6].

Owning to its simple multidimensional applicability, this paper suggests the adaptation of the Shepard operator based interpolation (first introduced in [10]) for interpolation based fuzzy reasoning. The Shepard interpolation method for arbitrarily placed bivariate data was introduced as follows [10]:

$$f = g(x,y) = \begin{cases} f_k & if\ (x,y) = (x_k, y_k) \text{ for some } k \\ \left(\sum_{k=0}^{n} f(x_k,y_k)/d_k^{\lambda}\right) / \left(\sum_{k=0}^{n} 1/d_k^{\lambda}\right) & \text{otherwise,} \end{cases} \tag{3.1}$$

where the measurement points $x_k$, $y_k$ ( $k \in [0,n]$ ) are irregularly spaced in the domain of $f \in R^2 \rightarrow R, \lambda > 0$ , and $d_k = \left[(x-x_k)^2 + (y-y_k)^2\right]^{1/2}$ . This function can be used typically when a surface model is required to interpolate scattered spatial measurements.

The adaptation of the Shepard interpolation method for interpolation based fuzzy reasoning in the vague environment of the fuzzy rule base is straightforward by substituting the Euclidean distances $d_k$ by the scaled distances $\delta_{s,k}$ :

$$\delta_{s,k} = \delta_s(a_k, x) = \left[\sum_{i=1}^{m} \left(\int_{a_{k,i}}^{x_i} S_{X_i}(X_i)dX_i\right)^2\right]^{1/2}, \tag{3.2}$$

where $S_{X_i}$ is the $i^{th}$ scaling function of the $m$ dimensional antecedent universe, $x$ is the $m$ dimensional crisp observation and $a_k$ is the abscissa of the prototype point of the $k^{th}$ fuzzy set in the $i^{th}$ antecedent dimension.

Thus, in the case of singleton rule consequents ($c_k$) the fuzzy rule $R_k$ has the following form:

$$If\ x_1 = A_{k,1}\ and\ x_2 = A_{k,2}\ and \ldots and\ x_m = A_{k,m}\ then\ y = c_k. \tag{3.3}$$

By substituting (3.2) into (3.1) the conclusion of interpolative fuzzy reasoning can be obtained as:

$$y_{(x)} = \begin{cases} c_k & \text{if } x = a_k \text{ for some } k \\ \left(\sum_{k=1}^{r} c_k/d_{s,k}^{\lambda}\right) / \left(\sum_{k=1}^{r} 1/d_{s,k}^{\lambda}\right) & \text{otherwise.} \end{cases} \tag{3.4}$$

## 4. Gradient-Based Consequent Optimization

The main contribution of this paper is the suggestion of a gradient-based optimization method (steepest descent) for the consequent optimization of the FIVE rule base.
If the performance function is derivable, we can apply the gradient method. Consequent Optimization is based on a set of sample (training) data. The goal of the optimization method is to minimize the squared error $E$ of the fuzzy model.

$$E = \sum_{k=1}^{N} (y_d(x_k) - y(x_k))^2, \tag{4.1}$$

where $y_d(x_k)$ is the desired output of the $k^{th}$ training data and $y(x_k)$ is the output of the fuzzy model applying FIVE (as interference technique), $N$ is the number of the training data points.

The applied steepest descent parameter optimization method modifies the rule consequents based on their partial derivatives to the squared error function $E$ (4.1) in the following manner:

$$g(c_k) = \frac{\partial E(c_k)}{\partial c_k} = \frac{\partial E(c_k)}{\partial y(x)} \frac{\partial y(x)}{\partial c_k} \tag{4.2}$$

$$c_{k_{next}} = c_k - \tau g(c_k), \tag{4.3}$$

where $\tau$ is the step size of the iteration and $c_{k_{next}}$ is the next iteration of the $k^{th}$ conclusion $c_k$.

According to (4.1), (4.2) can be rewritten in the following form:

$$g(c_k) = -2(y_d(x_k) - y(x_k)) \frac{\partial y(x)}{\partial c_k}. \tag{4.4}$$

Applying the Shepard interpolation formula of FIVE (3.4), for the partial derivatives we get the following formulas:

$$\frac{\partial y(x)}{\partial c_k} = \begin{cases} 1 & if\ x = a_k \text{ for some } k \\ \left(1/d_{s,k}^{\lambda}\right) / \left(\sum_{k=1}^{r} 1/d_{s,k}^{\lambda}\right) & \text{otherwise.} \end{cases} \tag{4.5}$$

According to (4.3), (4.4) and (4.5) the next iteration of the $k^{th}$ conclusion $c_k$ can be calculated.

## 5. Test and Benchmark

### 5.1. Application Example

The training data of the application example are a simple input-output set of randomly selected data from the y=sin(x)/x function in the domain of [-20, 20]. For demonstration purposes this domain is covered by 13 single input, single output fuzzy rules in the following form (for the $k^{th}$ rule of the rule base):

$$If\ x = A_k\ then\ y = c_k. \tag{5.1}$$

For the initial state of the experiment all the consequents of the fuzzy rules are set to 1 ($c_k = 1, k \in [1, 13]$).

The antecedents ($A_k$) of the fuzzy rules are fixed and more or less evenly distributed in the domain according to 5.1.

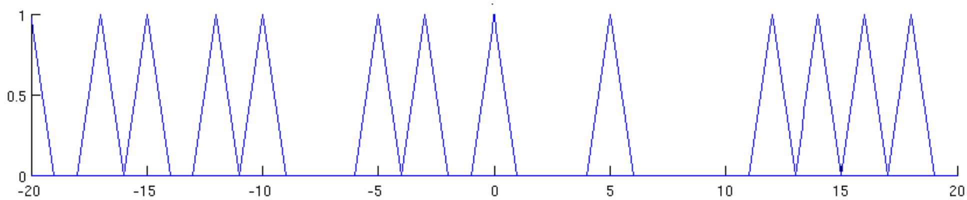


**Figure 2.** Fixed antecedents ($A_k$) of the 13 fuzzy rules

Fig. 3 introduces the values of the training data, the conclusions of the initial, and the parameter-optimized fuzzy models. The change of the squared error of the training data and the fuzzy model (4.1) in the function of the iteration steps is illustrated in Fig. 4.

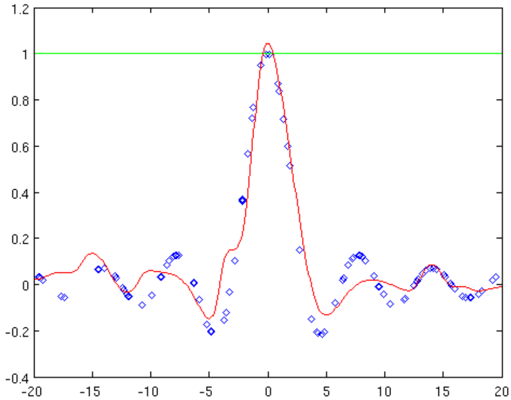


**Figure 3.** Training data (circles), conclusions of the initial (horizontal line), and the parameter-optimized fuzzy model (curve)

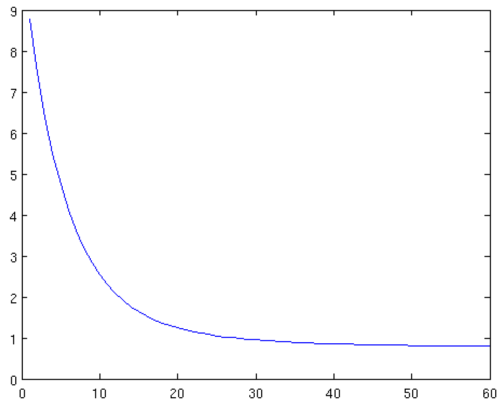


**Figure 4.** Change of the squared error (4.1) against the iteration steps

## 5.2. Petrophysical Properties Benchmark

In order to prove the practical applicability of our technique we compared the performance of a fuzzy model generated with the method presented above to some previously published results obtained using other methods for a real world problem taken from the field of petrophysical properties analysis.

The main goal of this example is to compare the optimized FIVE system to the system generated by the RBE-DSS (introduced by Johanyák in [11]) method.

One of the key tasks in the course of the analysis of petroleum oil well data is the prediction of petrophysical properties corresponding to specific input data, i.e. depth values that are different from the original ones used by the experiments. Such properties are porosity, permeability and the volume of clay [12]. The expensive and time-consuming character of data collection from boreholes increases the significance of the prediction. The predicted values help making decisions on the rentability of the exploration of a specific region. The research task of Johanyák was to create a fuzzy model with low complexity that is applicable for the prediction of porosity (PHI) based on well log data described by three input variables. These are the gamma ray (GR), deep induction resistance (ILD), and sonic travel time (DT).

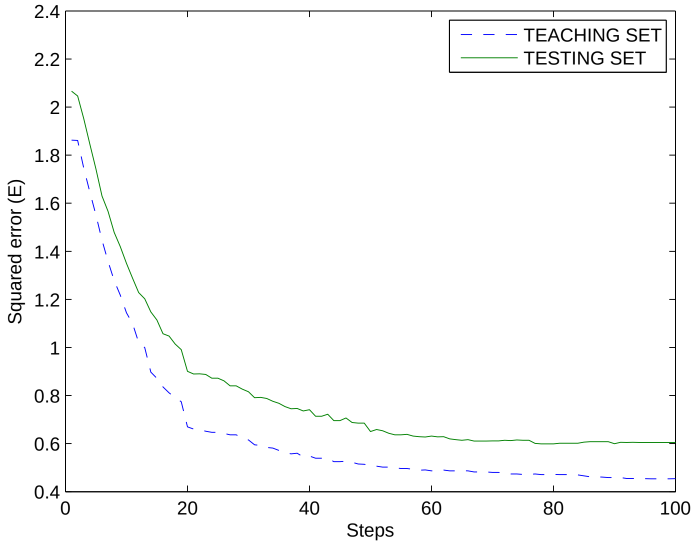


**Figure 5.** Change of the squared error (4.1) against the iteration steps

The reference system used in the course of the benchmark was developed by Johanyák [11] using RBE-DSS as the model identification technique and LESFRI as the FRI method.

The initial rule base was generated by the RBE-DSS method which is sub-optimal in the case of method FIVE, hence the beginning correlation is not a notable value. With the help of gradient-based optimization (introduced in Section 4) after 100 steps the correlation became good as can be seen in Tab. 1 and Fig. 6, and the performance function value (squared error) decreased as Fig. 5 shows. The optimization results are worse than the results obtained with RBE-DSS and LESFRI because our method modifies only the consequent, and disregards the antecedent part of the rules.

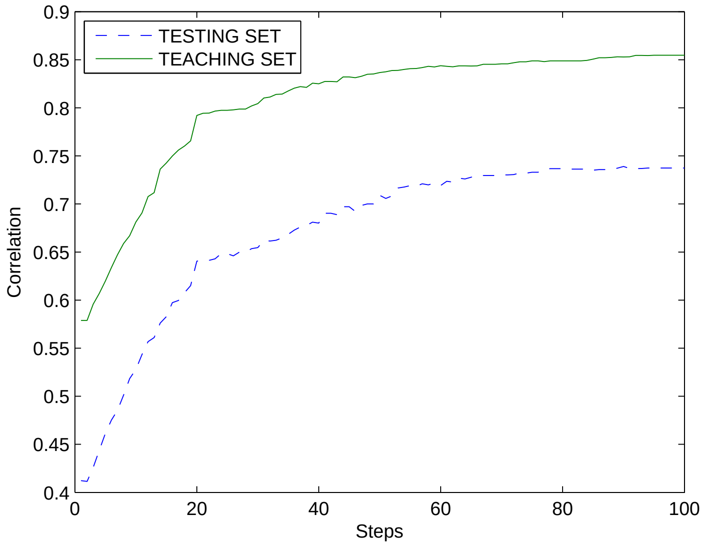


**Figure 6.** Change of the correlation against the iteration steps

**Table 1.** Changing of the correlation coefficient against the iteration steps

| | beginning | step 50 | step 100 | RBE-DSS |
|---|---|---|---|---|
| training | 0.5885 | 0.8343 | 0.8562 | 0.934 |
| test | 0.4056 | 0.6935 | 0.7281 | 0.890 |

## 6. Conclusion

As a first step of automatic rule base generation for FRI methods, this paper suggests a gradient-based optimization method (steepest descent) for the consequent optimization of the FIVE rule base. Consequent optimization is based on a set of sample (training) data. The goal of the optimization method is to minimize the squared error of the training data and the FRI fuzzy model.

Based on the numerical example (introduced in Section 5.2) the correlation is increasing and the squared error is decreasing quickly and it is close to the result obtained with RBE-DSS and LESFRI. The optimization results are worse than the results of RBE-DSS because this method modifies only the consequent, and does not alter the antecedent part of the rules. The chosen

FRI method (FIVE) is rather simple but efficient, serving as a good basis for further improvement of fully automatic FRI FIVE rule base generation which optimizes the antecedent part as well.

## Acknowledgements

This research was partly supported by the Hungarian National Scientific Research Fund grant number OTKA K77809 and the Intelligent Integrated Systems Japanese-Hungarian Laboratory.

## REFERENCES

[1] KOVÁCS, S.: Extending the fuzzy rule interpolation "five" by fuzzy observation. In *Advances in Soft Computing, Computational Intelligence, Theory and Applications*, Springer, Germany, ISBN 3-540-34780-1, 2006, pp. 445–497.

[2] JOHANYÁK, Z., PARTHIBAN, R., and SEKARAN, G.: Fuzzy modeling for an anaerobic tapered fluidized bed reactor. In *Advances in Soft Computing, Computational Intelligence, Theory and Applications*, Timisoara, Romania, ISBN 3-540-34780-1, 2006, pp. 67–72.

[3] JOHANYÁK, Z. and SZABÓ, A.: Tool life modelling using RBE-DSS method and LESFRI inference mechanism. *A GAMF Közleményei*, pp. 18–27.

[4] KOVÁCS, S.: New aspects of interpolative reasoning. In *Proceedings of the 6th. International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems*, Granada, Spain, 1996, pp. 477–482.

[5] KOVÁCS, S. and KÓCZY, L. T.: Approximate fuzzy reasoning based on interpolation in the vague environment of the fuzzy rule base as a practical alternative of the classical CRI. In *Proceedings of the 7th International Fuzzy Systems Association World Congress*, Prague, Czech Republic, 1997, pp. 144–149.

[6] KOVÁCS, S. and KÓCZY, L. T.: The use of the concept of vague environment in approximate fuzzy reasoning. In *Fuzzy Set Theory and Applications, Tatra Mountains Mathematical Publications, Mathematical Institute Slovak Academy of Sciences*, vol. 12, Bratislava, Slovak Republic, 1997, pp. 169–181.

[7] KOVÁCS, S.: `http://www.iit.uni-miskolc.hu/~szkovacs`.

[8] FRANK, K.: Fuzzy sets and vague environments. *Fuzzy Sets and Systems*, **66**, (1994), 207–221.

[9] TURKSEN, I. and ZHONG, Z.: An approximate analogical reasoning scheme based on similarity measures and interval valued fuzzy sets. *Fuzzy Sets and Systems*, **34**, (1990), 323–346.

[10] SHEPARD, D.: A two dimensional interpolation function for irregularly spaced data. In *Proc. 23rd ACM Internat.*, 1968, pp. 517–524.

[11] Johanyák, Z.: *Fuzzy szabály-interpolációs módszerek és mintaadatok alapján történő automatikus rendszergenerálás.* Ph.D. thesis, Hatvany József Informatikai Tudományok Doktori Iskola, 2007.

[12] Wong, K. W. and Gedeon, T. D.: Petrophysical properties prediction using self-generating fuzzy rules inference system with modified alpha-cut based fuzzy interpolation. In M. Nikravesh, J. Kacprzyk, and L. A. Zadeh (eds.), *Proceedings of The Seventh International Conference of Neural Information Processing ICONIP*, 2000, pp. 1088–1092.