Tuning of Fuzzy Systems as an Ill-Posed Problem

Martin Burger[1], Josef Haslinger[2], and Ulrich Bodenhofer[2]

[1] SFB F 013 Numerical and Symbolic Scientific Computing and Industrial Mathematics Institute, Johannes Kepler Universität Linz, Altenbergerstr. 69, A-4040 Linz, Austria
[2] Software Competence Center Hagenberg GmbH, Hauptstr. 99, A-4232 Hagenberg, Austria

Abstract. This paper is concerned with the data-driven construction of fuzzy systems from the viewpoint of regularization and approximation theory, where we consider the important subclass of Sugeno controllers. Generally, we obtain a nonlinear constrained least squares approximation problem which is *ill-posed*. Therefore, nonlinear regularization theory has to be employed. We analyze a smoothing method, which is common in spline approximation, as well as Tikhonov regularization, along with rules how to choose the regularization parameters based on nonlinear regularization theory considering the error due to noisy data.For solving the regularized nonlinear least squares problem, we use a generalized Gauss-Newton like method. A typical numerical example shows that the regularized problem is not only more robust, but also favors solutions that are easily interpretable, which is an important quality criterion for fuzzy systems.

1 Introduction

Fundamentally, the idea of fuzzy sets and systems, dated back to Zadeh [9], is to provide a mathematical model that can present and process human imprecise knowledge. Beside the top-down construction of fuzzy rule systems from expert knowledge, the (semi)automatic construction from example data has become a major issue. In the field of continuous model identification, so-called Sugeno controllers [8] have emerged as appropriate models, since they ideally combine simplicity with good analytical properties.

Definition 1. Let X be a non-empty set (the input space), let $A_1, A_2, \dots, A_n$ be non-empty fuzzy subsets of X with membership functions $\mu_{A_i} : X \to [0,1]$, $\sum \mu_{A_i}(x) > 0$ for all $x \in X$, and $\alpha_1, \alpha_2, \dots, \alpha_n$ be elements of $\mathbb{R}$, and consider the rulebase $(i = 1, 2, \dots, n)$

$$\textsf{if } x \textsf{ is } A_i \textsf{ then } y = \alpha_i.$$

Then the Sugeno controller defines the following input-output function $F_s : X \to \mathbb{R}$

$$F_s(x) = \frac{\sum \mu_{A_i}(x)\, \alpha_i}{\sum \mu_{A_i}(x)}. \tag{1}$$

As obvious from (1), we have to find appropriate fuzzy sets A_i and consequent parameters α_i. From the practical point of view, it is reasonable and necessary to consider only families of fuzzy sets which can be described by finite sets of parameters. In this paper, we concentrate on chains of trapezoidal membership functions and B-splines [10]. Beside simplicity and good analytical properties, both models provide excellent circumstances also to allow qualitative insight into the numerical relationships [7,10].

We restrict ourselves to the one-dimensional case, i.e. a single input-single output controller. Equation (1) together with the properties of trapezoidal or B-spline membership functions ($\sum_{i=1}^{n} \mu_{A_i}(x) = 1, \forall x \in X$) reduces tuning of a Sugeno controller to fitting a set of measured data $\{(x_i, y_i)\}_{i=1,\ldots,m}$ by a linear combination of these membership functions in the least squares sense, i.e. seeking a solution of the minimization problem

$$\sum_{i=1}^{m} \Big(y_i - \sum_{j=1}^{n} \alpha_j b_j(x_i; \boldsymbol{t}) \Big)^2 = \min_{((\alpha_1, \alpha_2, \ldots, \alpha_n), \boldsymbol{t}) \in \mathbb{R}^n \times [a,b]^\ell}, \tag{2}$$

where b_j represents the j-th trapezoidal or B-spline membership function and $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \ldots, \alpha_n)^T$ the consequent values. The concrete shape of the membership functions depends on an ascending ℓ-dimensional knot sequence $\boldsymbol{t}$, which has to be included in the optimization procedure. Therefore, the minimization problem (2) is constrained and nonlinear.

As will be shown in Sect. 2, the nonlinear least squares problem is ill-posed in the sense that arbitrarily small errors in the measured data possibly lead to arbitrarily large errors in the solution of (2), which necessitates nonlinear regularization theory [1,4]. In this paper, we adapt the existing theory to the given approximation problem.

Existing methods for determining the appropriate parameters of a Sugeno controller from data have either concentrated on genetic algorithms or artificial neural networks. Scarcely, also simple numerical methods have been considered. All these methods have in common that they are slow, inaccurate, and unreliable, due to the notorious ignorance of the inherent ill-posedness of the problem. Based on the theoretical investigations, a method for solving the regularized problem is proposed which allows stable and accurate solutions.

2 Regularized Tuning of Sugeno controllers

First, we show, that the nonlinear least-squares problem is indeed ill-posed, and then we consider two methods for regularization, along with rules how to choose the regularization parameters depending on the given noise level. Proofs and further details can be found in [2].

2.1 The Ill-Posedness of the Least-squares Problem

Assume $(\boldsymbol{x}, \boldsymbol{y})$ is a set of training data, where $\boldsymbol{x} = (x_1, x_2, \ldots, x_m)^T$ is the training data vector, and $\boldsymbol{y} = (y_1, y_2, \ldots, y_m)^T$ the desired output for $\boldsymbol{x}$.

Whenever data is measured, we have to consider data errors in $\boldsymbol{x}$ and $\boldsymbol{y}$. In this paper, we assume the error bounds

$$\|\boldsymbol{x} - \boldsymbol{x}^\gamma\|_{\ell^2} \leq \gamma \text{ and } \|\boldsymbol{y} - \boldsymbol{y}^\delta\|_{\ell^2} \leq \delta, \tag{3}$$

where $\|\boldsymbol{x}\|_{\ell^2} := \sqrt{\sum_{i=1}^m x_i^2}$ denotes the usual ℓ_2 norm.

The following example shows that the problem of finding a minimum for (2) is ill-posed, even if we have complete information about the function f, from which the samples $\boldsymbol{y}$ are taken.

Example 1. Let $n = 2$, $k \in \mathbb{N}$, $k \geq 2$, $a = t_1^k = 0$, $t_2^k = k^{-3}$ and $t_3^k = 2k^{-3}$, $t_4^k = b = 1$, and choose $\alpha_1^k = k$, $\alpha_2^k = 0$. The fuzzy membership functions b_1 and b_2 are defined by

$$b_1(x; \boldsymbol{t}) = \begin{cases} 1 & \text{if } x \leq t_2 \\ \frac{t_3 - x}{t_3 - t_2} & \text{if } t_2 < x < t_3 \\ 0 & \text{if } t_3 \leq x \end{cases} \tag{4}$$

$$b_2(x; \boldsymbol{t}) = 1 - b_1(x; \boldsymbol{t}). \tag{5}$$

Then $f^k = \alpha_1^k b_1(x; \boldsymbol{t}^k) + \alpha_2^k b_2(x; \boldsymbol{t}^k)$ converges to zero in $L_2([0,1])$, but $\boldsymbol{\alpha}^k$ has no bounded subsequence. Hence, the optimization problem is unstable with respect to perturbations in the data.

2.2 Smoothing

First, we consider "smoothing", a common method for spline approximation (e.g. see [3,6]), where an additional term in the optimization functional characterizes the smoothness of the spline, i.e. (2) is replaced by

$$\sum_{i=1}^m \Big(y_i^\delta - \sum_{j=1}^n \alpha_j b_j(x_i^\gamma; \boldsymbol{t}) \Big)^2 + \beta \Big| \sum_{j=1}^n \alpha_j b_j(.; \boldsymbol{t}) \Big|_{H^k(\Omega)}^2 = \min_{(\boldsymbol{\alpha}, \boldsymbol{t})} \tag{6}$$

where $|.|_{H^k(\Omega)}$ denotes the norm or seminorm in the Sobolev space $H^k(\Omega)$. In addition we impose the constraints

$$t_{j+1} - t_j \geq \varepsilon, \qquad j = 1, \ldots, \ell - 1, \tag{7}$$

on the knot vector $\boldsymbol{t}$ in order to prevent possible instabilities caused by two equal or almost equal knots.

The proof that (6) subject to (7) is indeed a well-posed problem is based on a result adapted from stability estimates in finite element theory. The solution will converge to a minimizer of the original problem with the additional constraint (7) for fixed ε and appropriately chosen $\beta \to 0$ as the noise level $\gamma, \delta \to 0$. However, we cannot show convergence with $\varepsilon \to 0$.

Theorem 1 (Convergence under Constraints). [2] *Let $\varepsilon > 0$ be fixed, let (γ^k, δ^k) be a monotone sequence converging to $(0,0)$ and let $(\boldsymbol{x}^{\gamma^k}, \boldsymbol{y}^{\delta^k})$ be a corresponding data sequence satisfying (3) with $(\gamma, \delta) = (\gamma^k, \delta^k)$. Moreover, let the regularization parameter β^k be chosen such that $\beta^k \to 0$ and*

$$\frac{\max\{\gamma^k, \delta^k\}}{\beta^k} \to 0. \tag{8}$$

If a minimizer of (2) with exact data exists, then each sequence of minimizers $(\boldsymbol{\alpha}^k, \boldsymbol{t}^k)$ of (6), (7) with noisy data $(\boldsymbol{x}^{\gamma^k}, \boldsymbol{y}^{\delta^k})$ and $\beta = \beta^k$ has a convergent subsequence and the limit of each convergent subsequence is a minimizer of the least squares problem (2) subject to (7).

2.3 Tikhonov Regularization

The second approach under investigation is classical Tikhonov regularization in the parameter space $\mathbb{R}^n \times \mathbb{R}^\ell$. It consists of minimizing the functional

$$\sum_{i=1}^{m} \Big(y_i^\delta - \sum_{j=1}^{n} \alpha_j b_j(x_i^\gamma; \boldsymbol{t}) \Big)^2 + \beta_1 \sum_{j=1}^{n} \alpha_j^2 + \beta_2 \sum_{j=1}^{\ell} (t_j - t_j^*)^2 = \min_{(\boldsymbol{\alpha}, \boldsymbol{t})} \tag{9}$$

for appropriately chosen β_1 and β_2 (depending on δ and $\boldsymbol{y}^\delta$), where $\boldsymbol{t}^*$ is a prior for $\boldsymbol{t}$, e.g. uniform grid points. In this case, we can show convergence for appropriate choice of $\beta_1 \to 0$ as the noise level tends to zero even for $\beta_2 = 0$.

Theorem 2 (Convergence). [2] *Assume that a minimizer of problem (9) exists. Moreover, let (γ^k, δ^k) be a sequence converging to $(0,0)$ and denote by $(\boldsymbol{\alpha}^k, \boldsymbol{t}^k)$ the according sequence of minimizers of (9) with data $(\boldsymbol{x}^\gamma, \boldsymbol{y}^\delta)$, satisfying (3). Then $(\boldsymbol{\alpha}^k, \boldsymbol{t}^k)$ has a convergent subsequence and the limit of every convergent subsequence is a minimizer of (9) with exact data $(\boldsymbol{x}, \boldsymbol{y})$ if the regularization parameters satisfy $\beta_1^k \to 0$, $\beta_2^k \to 0$, and*

$$\frac{\max\{\gamma^k, \delta^k\}}{\beta_1^k} \to 0 \tag{10}$$

$$\exists \kappa > 0 : \frac{\beta_1^k}{\beta_2^k} \geq \kappa. \tag{11}$$

3 Numerical Example – Spectral Data

Details of the optimization algorithm for solving the constrained least squares problem – a generalized Gauss-Newton like method – can be found in [5,6].

Similar to [7], we want to construct a transparent rule-based model from noisy data measurements considering the spectral data function

$$f(x) := 12\, \mathrm{e}^{\frac{-(x-4.8)(x-5.8)}{0.7}} - 12\mathrm{e}^{-(x+3.5)^2} + 0.8x \qquad x \in [-10, 10].$$

By using inputs x uniformly distributed in $[-10, 10]$, 50 samples of $f(x)$ were obtained and then disturbed with uniformly distributed noise within a noise level of 10% (δ =9.5804, maximal error =2.0398).

In this example, the input space is split into eight triangular fuzzy sets (B-splines of order 2) which can be interpreted as "negative big", "negative medium", "negative small", "negative very small", "positive very small", "positive small", "positive medium", and "positive big".

Figure 1 shows the results for approximation without any regularization, while Fig. 2 shows the results for smoothing and Tikhonov regularization. Although the residuum is smaller without regularization, only Tikhonov regularization succeeds in constructing an interpretable fuzzy controller, since knots are separated appropriately. In approximation and smoothing, knots of the optimized sequence nearly coincide leading to questionable and difficultly interpretable membership functions. For Tikhonov regularization, $\boldsymbol{t}^*$ is chosen to be equidistant within $[-10, 10]$. The linguistic interpretation of the fuzzy model constructed by Tikhonov regularization is given in Table 1.

Acknowledgements

The work of Martin Burger is supported by the *Austrian National Science Foundation FWF* under grant SFB F013/08. Josef Haslinger and Ulrich Bodenhofer are working in the framework of the *Kplus Competence Center Program* which is funded by the Austrian Government, the Province of Upper Austria, and the Chamber of Commerce of Upper Austria.

We thank Prof. Heinz W. Engl (University of Linz) for the idea and motivation of this cooperation and for stimulating and fruitful discussions.

References

1. A. Binder, H.W. Engl, C.W. Groetsch, A. Neubauer, and O. Scherzer. Weakly closed nonlinear operators and parameter identification in parabolic equations by Tikhonov regularization. *Appl. Anal.*, 55:215–234, 1994.
2. M. Burger, J. Haslinger, and U. Bodenhofer. Regularized optimization of fuzzy controllers. Technical Report SCCH-TR-0056, Software Competence Center Hagenberg, 2000.
3. C. de Boor. *A practical guide to splines*. Springer-Verlag, New York, 1978.
4. H.W. Engl, M. Hanke, and A. Neubauer. *Regularization of Inverse Problems*. Kluwer, Dordrecht, 1996.
5. T. Schütze. *Diskrete Quadratmittelapproximation durch Splines mit freien Knoten*. PhD thesis, Technical University Dresden, Department of Mathematics, 1997.
6. H. Schwetlick and T. Schütze. Least squares approximation by splines with free knots. *BIT*, 35(3):361–384, 1995.
7. M. Setnes, R. Babuška, and H.B. Verbruggen. Rule-based modeling: Precision and transparency. *IEEE Trans. Syst. Man Cybern., Part C: Applications & Reviews*, 28(1):165–169, 1998.

8. T. Takagi and M. Sugeno. Fuzzy identification of systems and its applications to modeling and control. *IEEE Trans. Syst. Man Cybern.*, 15(1):116–132, 1985.
9. L.A. Zadeh. Outline of a new approach to the analysis of complex systems and decision processes. *IEEE Trans. Syst. Man Cybern.*, 3(1):28–44, 1973.
10. J. Zhang and A. Knoll. Constructing fuzzy controllers with B-spline models. In *Proc. FUZZ-IEEE'96*, volume I, pages 416–421, 1996.

Rule: Antecedent		Consequent singleton	Consequent label
R1 : If x is *Negative Big*	then	y= -7.605	*Negative Medium*
R2 : If x is *Negative Medium*	then	y= -5.025	*Negative Medium*
R3 : If x is *Negative Small*	then	y=-11.063	*Negative Big*
R4 : If x is *Negative very Small*	then	y= -0.460	*Negative very Small*
R5 : If x is *Positive very Small*	then	y= 1.367	*Positive very Small*
R6 : If x is *Positive Small*	then	y= 15.095	*Positive Big*
R7 : If x is *Positive Medium*	then	y= 4.968	*Positive Medium*
R8 : If x is *Positive Big*	then	y= 7.682	*Positive Medium*

Table 1. Sugeno controller identified from noisy data

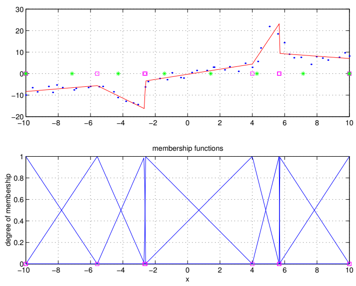

Fig. 1. Approximation with 8 triangular membership functions ($\varepsilon = 0.01$) without regularization

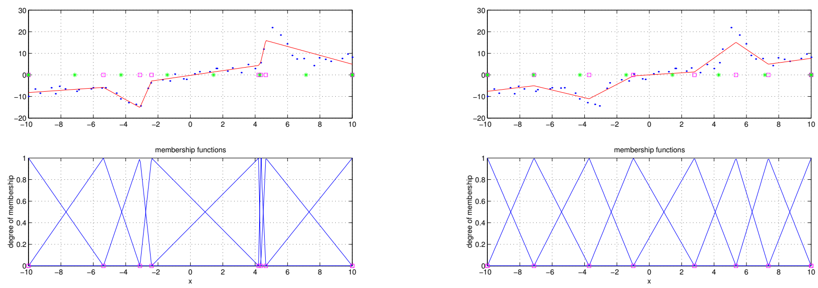

Fig. 2. Approximation with 8 triangular membership functions with smoothing (left; $k = 1, \beta = 0.01$, $\varepsilon = 0.01$) and Tikhonov regularization (right; $\beta_1 = \beta_2 = 0.5$)