Fuzzy Optimization and Decision Making, 3, 63–82, 2004

Robust Solution to Fuzzy Identification Problem with Uncertain Data by Regularization

Fuzzy Approximation to Physical Fitness with Real World Medical Data: An Application

MOHIT KUMAR[1] mohit.kumar@etechnik.uni-rostock.de
Institute of Occupational and Social Medicine, Faculty of Medicine, University of Rostock, D-18055 Rostock, Germany

REGINA STOLL Regina.Stoll@med.uni-rostock.de
Institute of Occupational and Social Medicine, Faculty of Medicine, University of Rostock, D-18055 Rostock, Germany

NORBERT STOLL norbert.stoll@etechnik.uni-rostock.de
Institute of Automation, Department of Electrical Engineering and Information Technology, University of Rostock, D-18119 Rostock-Warnemünde, Germany

Abstract. This study considers the robust identification of the parameters describing a Sugeno type fuzzy inference system with uncertain data. The objective is to minimize the worst-case residual error using a numerically efficient algorithm. The Sugeno type fuzzy systems are linear in consequent parameters but nonlinear in antecedent parameters. The robust consequent parameters identification problem can be formulated as second-order cone programming problem. The optimal solution of this second-order cone problem can be interpreted as solution of a Tikhonov regularization problem with a special choice of regularization parameter which is optimal for robustness (Ghaoui and Lebret (1997). SAIM Journal of Matrix Analysis and Applications 18, 1035–1064). The final regularized nonlinear optimization problem allowing simultaneous identification of antecedent and consequent parameters is solved iteratively using a generalized Gauss–Newton like method. To illustrate the approach, several simulation studies on numerical examples including the modelling of a spectral data function (one-dimensional benchmark example) is provided. The proposed robust fuzzy identification scheme has been applied to approximate the physical fitness of patients with a fuzzy expert system. The identified fuzzy expert system is shown to be capable of capturing the decisions (experiences) of a medical expert.

Keywords: fuzzy-modelling, least-squares problems, uncertainty, robustness, second-order cone programming, regularization, robust identification

1. Introduction

Modelling of dynamic processes is required for simulation, prediction, model based control and fault diagnosis. Fuzzy modelling or rule based modelling is considered better than conventional mathematical techniques based on nonlinear differential equations for dealing with ill-defined and uncertain processes. Fuzzy modelling is based on a set of fuzzy if-then rules derived from expert domain knowledge to handle uncertainty. However, in some situations this expert domain knowledge may not be sufficient to design fuzzy model due to lack of knowledge or problems due to different biases of human experts. So there is need of learning of fuzzy inference system from data. As data used in learning contains generally vagueness and ambiguity, so there is need of developing algorithms for learning from imprecise data. Fuzzy modelling and control has been widely applied for modelling complex processes and decision problems (Zadeh (1973)). The main advantage of fuzzy inference system over classical learning system and neural-network is its linguistic interpretability of its function. Automatic construction or tuning of these fuzzy systems from example data has been widely explored for Sugeno type fuzzy inference system (Babuska (2000), Bodenhofer and Bauer (2000), Espinosa and Vandewalle (2000), Setnes et al. (1998)), since they are supposed to ideally combine simplicity with good analytical properties Takagi and Sugeno (1985).

It is easy to state that solution of fuzzy identification problem may exhibit very sensitive behavior to perturbations in the data. Many regularization methods have been proposed to decrease sensitivity of identification algorithms. As pointed by Golub (1989), the choice of regularization parameter is usually not obvious and application dependent. Several criteria for optimizing the regularization parameters have been proposed by Hunt (1973), Elden (1977) and Furuya (1989). These criteria are chosen according to some additional a priori information, of deterministic or stochastic nature. The extensive survey by Nashed (1981), Demoment (1989) and Hanke and Hansen (1993) discuss these problems and some applications. In this paper we consider the subject of deterministic robustness of fuzzy identification problem in which the perturbations in the data are deterministic and unknown but bounded. The robust identification problem can be formulated as second-order cone programming (SOCP) problem. The optimal solution of second-order cone problem can be interpreted as solution of a Tikhonov regularization problem with a special choice of regularization parameter which is optimal for robustness (Ghaoui and Lebret 1997). A numerical method based on generalized Gauss–Newton is suggested for solving nonlinear regularized optimization problem, maintaining the interpretability (which is key feature of fuzzy inference system) of the fuzzy system. The proposed numerical method is shown to converge to the robust solution of identification problem.

The later part of the paper deals with the fuzzy modelling of the functional relationship between physical fitness and various measurable physiological parameters. The fuzzy expert system must be tuned by the experience of a medical expert to identify the complex physiological relations between the physiological parameters. However, different medical experts may have different opinions and different biases of their mind regarding the notion of physical fitness. So we assume a fixed level of uncertainty in the opinion of our medical expert. Hence the identification of fuzzy expert system must be robust against the uncertainty lying in the opinion, therefore encouraging the use of robust fuzzy identification scheme in the problem of physical fitness approximation.

2. Sugeno Fuzzy Inference System

Let us consider a Sugeno fuzzy inference system ($F_s : X \rightarrow Y$), mapping n-dimensional input space ($X = X_1 \times X_2 \times \cdots \times X_n$) to one-dimensional real line, consisting of K different rules. The ith rule is in the the form:

If x_1 is A_{i1} and x_2 is A_{i2} .... and x_n is A_{in} then $y = \alpha_i$; for all $i = 1, 2, \ldots, K$, where $A_{i1}, A_{i2}, \ldots, A_{in}$ are non-empty fuzzy subsets of $X_1, X_2, \ldots, X_n$ respectively such that membership function $\mu_{A_{ij}}$: $X_j \rightarrow [0, 1]$ fulfill $\sum_{i=1}^{K} \prod_{j=1}^{n} \mu_{A_{ij}}(x_j) > 0$ for all $x_j \in X_j$. The values $\alpha_1, \alpha_2, \ldots, \alpha_K$ are real numbers. So we have

$$F_s(x_1, x_2, \ldots, x_n) = \frac{\sum_{i=1}^{K} \alpha_i \prod_{j=1}^{n} \mu_{A_{ij}}(x_j)}{\sum_{i=1}^{K} \prod_{j=1}^{n} \mu_{A_{ij}}(x_j)}. \tag{1}$$

We assume that x_j belongs to a non-empty real intervals, i.e. $x_j \in [a_j, b_j]$ for all $j = 1, \ldots, n$.

The various classes of membership functions can be formulated in a common mathematical frame by considering a knot vector θ consisting of real elements, such that shape of membership functions depends on the elements of vector θ which partition the universe of each input variable ($x_i \in [a_i, b_i]$) into P_i linguistic terms. To elaborate the construction of membership functions based on knot vector (θ), two examples one for trapezoidal and other for gaussian membership functions are provided below.

Trapezoidal membership function: Let $\theta = (t_1^1, \ldots, t_1^{2P1-2}, t_2^1, \ldots, t_2^{2P2-2}, \ldots, t_n^1, \ldots, t_n^{2Pn-2}) \in R^L$, such that for ith input, $a_i \leq t_i^1 \leq \cdots \leq t_i^{2Pi-2} \leq b_i$ holds for all $i = 1, \ldots, n$. So P_i trapezoidal membership functions for ith input ($A_{1i}, A_{2i}, \ldots, A_{P_i i}$) can be defined as

$$A_{1i}(x_i, \theta) = \begin{cases} 1, & \text{if } x_i \in [a_i, t_i^1], \\ \frac{-x_i + t_i^2}{t_i^2 - t_i^1}, & \text{if } x_i \in [t_i^1, t_i^2], \\ 0, & \text{otherwise.} \end{cases}$$

$$A_{ji}(x_i, \theta) = \begin{cases} \frac{x_i - t_i^{2j-3}}{t_i^{2j-2} - t_i^{2j-3}}, & \text{if } x_i \in [t_i^{2j-3}, t_i^{2j-2}], \\ 1, & \text{if } x_i \in [t_i^{2j-2}, t_i^{2j-1}], \\ \frac{-x_i + t_i^{2j}}{t_i^{2j} - t_i^{2j-1}}, & \text{if } x_i \in [t_i^{2j-1}, t_i^{2j}], \\ 0, & \text{otherwise.} \end{cases}$$

$$A_{P_i i}(x_i, \theta) = \begin{cases} \frac{x_i - t_i^{2P_i-3}}{t_i^{2P_i-2} - t_i^{2P_i-3}}, & \text{if } x_i \in [t_i^{2P_i-3}, t_i^{2P_i-2}], \\ 1, & \text{if } x_i \in [t_i^{2P_i-2}, b_i], \\ 0, & \text{otherwise.} \end{cases}$$

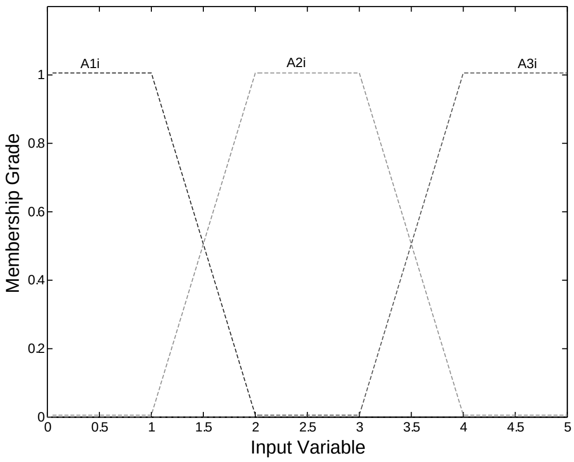


Figure 1. Trapezoidal membership functions.

Figure 1 shows an example with the choice of antecedent parameters as $a_i = 0$, $t_i^1 = 1$, $t_i^2 = 2$, $t_i^3 = 3$, $t_i^4 = 4$ and $b_i = 5$.

Gaussian membership functions with unit dispersion: Let $\theta = (t_1^1, \ldots, t_1^{P1-2}, t_2^1, \ldots, t_2^{P2-2}, \ldots, t_n^1, \ldots, t_n^{Pn-2}) \epsilon R^L$, such that for ith input, $a_i \leq t_i^1 \leq \cdots \leq t_i^{Pi-2} \leq b_i$ holds for all $i = 1, \ldots, n$. So P_i gaussian membership functions assuming unit dispersion for ith input $(A_{1i}, A_{2i}, \ldots, A_{P_i i})$ can be defined as

$$A_{1i}(x_i) = \mathrm{e}^{-(x_i - a_i)^2},$$

$$A_{ji}(x_i, \theta) = \mathrm{e}^{-(x_i - t_i^j)^2},$$

$$A_{P_i i}(x_i) = \mathrm{e}^{-(x_i - b_i)^2}.$$

Figure 2 shows an example with the choice of antecedent parameters as $a_i = 0$, $t_i^1 = 2.5$ and $b_i = 5$.

Total number of possible K rules depends on the number of membership functions for each input, i.e. $K = \Pi_{i=1}^n P_i$, where P_i is the number of membership functions defined over ith input. Depending upon the choice of membership functions Eq. (1) can be rewritten as function of θ.

$$F_s(x_1, x_2, \ldots, x_n) = \sum_{i=1}^{K} \alpha_i G_i(x_1, x_2, \ldots, x_n, \theta), \tag{2}$$

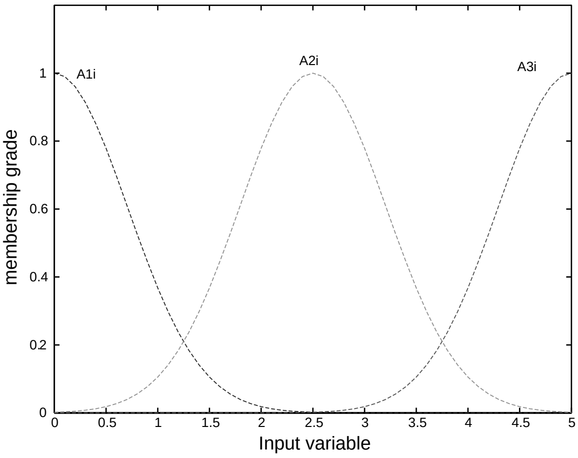


Figure 2. Gaussian membership functions.

where

$$G_i(x_1, x_2, \ldots, x_n, \theta) = \frac{\prod_{j=1}^{n} \mu_{A_{ij}}(x_j)}{\sum_{i=1}^{K} \prod_{j=1}^{n} \mu_{A_{ij}}(x_j)}.$$

Let us introduce the following notation i.e.

$$\alpha = [\alpha_i]_{i=1,\ldots,K} \in R^K, \quad x = [x_i]_{i=1,\ldots,n} \in R^n, \quad G = [G_i(x, \theta)]_{i=1,\ldots,K} \in R^K.$$

Now above equation can be written as

$$F_s(x) = G^{\mathrm{T}}(x, \theta)\alpha. \tag{3}$$

3. Problem Formulation

We consider the fuzzy approximation problem of the process described by following unknown equation at kth instant of time:

$$y(k) = f(x(k) + \delta x) + v(k),$$

where δx is the uncertainty present in vector $x(k) \in R^n$ and $v(k)$ is the bounded noise present in measurements. The aim is to postulate a fuzzy model, unfalsified by noisy measurements $[x(k), y(k)]_{k=1,\ldots,m}$. So we are required to estimate the fuzzy parameters $[\alpha, \theta]$ satisfying the following equation.

$$y(k) = (G(x(k), \theta) + \delta G)^{\mathrm{T}}\alpha + v(k),$$

where δG is the uncertainty in regression vector due to noise δx in $x(k)$.

Let us introduce the following notation:

$$Y = [y(k)]_{k=1,\dots,m} \in R^m,$$
$$B(\theta) = [G_i(x(k), \theta)]_{k=1,\dots,m;\ i=1,\dots,K} \in R^{m\times K}.$$

Now, our problem consists of finding the solution (α, θ) to an over determined set of equations $B(\theta)\alpha \approx Y$, where the matrix B and vector Y are subject to bounded perturbations. First we assume that the given model is not a single pair (B, Y) but a family of matrices $(B + \Delta B, Y + \Delta Y)$, where perturbation matrix $[\Delta B\, \Delta Y]$ is an unknown but bounded matrix, i.e. $\|\Delta B\, \Delta Y\|_F < \rho$, where $\|\cdot\|_F$ denotes the Frobenius norm and $\rho > 0$ is given. For a fixed (α, θ), we define the worst-case residual as

$$\mathrm{e}(B, Y, \rho, \alpha, \theta) = \max_{\|\Delta B\, \Delta Y\|_F \le \rho} \|(B(\theta) + \Delta B)\alpha - (Y + \Delta Y)\|,$$

where $\|\cdot\|$ denotes the usual Euclidean norm. We say that (α, θ) is a robust solution if it minimizes the worst-case residual. To find the optimal parameters (α^*, θ^*), we need to compute following problem

$$\Phi(B, Y, \rho) = \min_{(\alpha,\theta)} \max_{\|\Delta B\, \Delta Y\|_F \le \rho} \|(B(\theta) + \Delta B)\alpha - (Y + \Delta Y)\|.$$

Initially we keep a fixed value of variable θ and perform the minimization over the variable α only. Then the minimization problem reduces to

$$\Phi(B, Y, \rho, \theta) = \min_{\alpha} \max_{\|\Delta B\, \Delta Y\|_F \le \rho} \|(B + \Delta B)\alpha - (Y + \Delta Y)\|. \tag{4}$$

As stated by Ghaoui and Lebret (1997), this problem can be formulated as SOCP, however for the sake of completion we write the various steps involved. Using the triangular inequality, we have

$$\|B\alpha - Y\| + (\|\alpha\|^2 + 1)^{1/2} \ge \max_{\|\Delta B\, \Delta Y\|_F \le 1} \|(B + \Delta B)\alpha - (Y + \Delta Y)\|.$$

Choose perturbation matrices as

$$[\Delta B\, \Delta Y] = \frac{(B\alpha - Y)}{\|B\alpha - Y\|(\|\alpha\|^2 + 1)^{1/2}} [\alpha^{\mathrm{T}}\ -1].$$

With the above choice of perturbation matrices we have

$$\|(B+\Delta B)\alpha-(Y+\Delta Y)\| = \|B\alpha - Y\| + (\|\alpha\|^2+1)^{1/2}.$$

This implies that when $\rho = 1$, the worst-case residual is given by (assuming fixed θ)

$$\mathrm{e}(B, Y, 1, \alpha, \theta) = \|B\alpha - Y\| + (\|\alpha\|^2+1)^{1/2}.$$

The problem of minimizing $e(B, Y, 1, \alpha)$ over α has unique solution α^*, because of strict convexity of the worst-case residual. This problem can be formulated as the SOCP as follows:

$$\text{minimize}\, \lambda \text{ subject to } \|B\alpha - Y\| \le \lambda - \tau, \quad \|[\alpha^{\mathrm{T}}\; 1]^{\mathrm{T}}\| \le \tau.$$

Using duality results for SOCP, we have dual problem to above problem as

$$\text{maximize } Y^{\mathrm{T}}z - v \text{ subject to } B^{\mathrm{T}}z + u = 0, \quad \|z\| \le 1, \quad \|[u^{\mathrm{T}}\; v^{\mathrm{T}}]^{\mathrm{T}}\| \le 1.$$

Since both primal and dual problems are strictly feasible, there exists optimal points for both of them. If $\lambda = \tau$ at optimum, then $B\alpha = Y$ and $\lambda = \tau = (\|\alpha\|^2+1)^{1/2}$. In this case, the optimal α^* is the minimum-norm solution to $B\alpha = Y$. So, $\alpha^* = B^{\dagger}Y$, where $B^{\dagger}$ is the pseudoinverse of B. If $\lambda > \tau$ then again, both primal and dual problems are strictly feasible, therefore primal and dual-optimal objectives are equal. So,

$$\|B\alpha - Y\| + \|[\alpha^{\mathrm{T}}\; 1]\| = \lambda = Y^{\mathrm{T}}z - v = -(B\alpha - Y)^{\mathrm{T}}z + \alpha^{\mathrm{T}}B^{\mathrm{T}}z - v$$

Using $\|z\| \le 1$, $\|[u^{\mathrm{T}}\; v]^{\mathrm{T}}\| \le 1$, $u = -B^{\mathrm{T}}z$, we get

$$z = -\frac{B\alpha - Y}{\|B\alpha - Y\|},$$

$$[u^{\mathrm{T}}\; v] = -\frac{[\alpha^{\mathrm{T}}\; 1]}{(\|\alpha\|^2+1)^{1/2}}.$$

Replacing these values in $B^{\mathrm{T}}z + u = 0$, we get optimal value α^* as

$$\alpha^* = (B^{\mathrm{T}}B + \mu I)^{-1}B^{\mathrm{T}}Y, \text{ with } \mu = \frac{\|B\alpha - Y\|}{(\|\alpha\|^2+1)^{1/2}}.$$

The optimal solution α^* of robust identification problem can be interpreted as solution of a Tikhonov regularization problem with μ as regularization parameter. The corresponding regularized problem is

$$\text{minimize } \|B\alpha - Y\|^2 + \mu\|\alpha\|^2.$$

This implies that

$$\Phi(B, Y, 1, \theta) = \min_{\alpha} \|B\alpha - Y\|^2 + \mu\|\alpha\|^2, \quad \mu = \frac{\|B\alpha - Y\|}{(\|\alpha\|^2 + 1)^{1/2}}.$$

For any $\rho > 0$, we have $\Phi(B, Y, \rho) = \rho\Phi(B/\rho, Y/\rho, 1)$. It is easy to see that

$$\Phi(B, Y, \rho, \theta) = \min_{\alpha} \|(B/\rho^{1/2})\alpha - (Y/\rho^{1/2})\|^2 + \mu\|\alpha\|^2.$$

At this stage, we see the problem from regularization point of view. Instead of robust solution, we try to rather find equivalent regularized solution. If we also try to tune the parameter θ, the corresponding regularized minimization problem is

$$\Phi(B, Y, \rho) = \min_{(\alpha,\theta)} \|(B(\theta)/\rho^{1/2})\alpha - (Y/\rho^{1/2})\|^2 + \mu\|\alpha\|^2.$$

If we also provide regularization to the antecedent parameters, then the minimization problem becomes

$$\Phi(B, Y, \rho) = \min_{(\alpha,\theta)} \|(B(\theta)/\rho^{1/2})\alpha - (Y/\rho^{1/2})\|^2 + \mu\|\alpha\|^2 + \mu\|\theta - \theta_0\|^2$$

where θ_0 is initial guess about shape of membership functions and μ is the regularization parameter. Let

$$\Psi(\alpha, \theta) = \|(B(\theta)/\rho^{1/2})\alpha - (Y/\rho^{1/2})\|^2 + \mu\|\alpha\|^2 + \mu\|\theta - \theta_0\|^2.$$

So, we have

$$\Phi(B, Y, \rho) = \min_{(\alpha,\theta)} \Psi(\alpha, \theta).$$

Moreover, we also want to preserve the interpretability of fuzzy system during learning. So the membership functions can be prevented from overlapping by imposing some constraints on the position of knots, for instance, in case of trapezoidal membership functions the constraints can be formulated, i.e. for all $i = 1, \ldots, n$

$$\begin{aligned}
&t_i^1 - a_i \geq \epsilon_i \\
&t_i^{j+1} - t_i^j \geq \epsilon_i \quad \text{for all } j = 1, 2, \ldots, (2P_i - 3) \\
&b_i - t_i^{2P_i-2} \geq \epsilon_i
\end{aligned}$$

These constraints can be formulated in term of a matrix inequality $C\theta \geq h$ (Burger et al. (2002)). Finally the constrained regularized optimization problem is

$$\min_{(\alpha,\theta)} \Psi(\alpha, \theta) = \|(B(\theta)/\rho^{1/2})\alpha - (Y/\rho^{1/2})\|^2 + \mu\|\alpha\|^2 + \mu\|\theta - \theta_0\|^2,$$

where

$$\mu = \frac{\|B(\theta)\alpha - Y\|}{(\|\alpha\|^2 + 1)^{1/2}} \quad \text{and} \quad C\theta \geq h.$$

4. Iterative Solution of Fuzzy Identification Problem

Let $[\alpha^*, \theta^*]$ be the solution of robust fuzzy identification problem, minimizing the functional $\Psi(\alpha, \theta)$. Therefore, $[\alpha^*, \theta^*]$ must satisfy following equation.

$$\alpha^* = \left(\frac{B^{\mathrm{T}}(\theta^*)B(\theta^*)}{\rho} + \mu^* I\right)^{-1} \frac{B^{\mathrm{T}}(\theta^*)Y}{\rho} \quad \text{with} \quad \mu^* = \frac{\|B(\theta^*)\alpha^* - Y\|}{(\|\alpha^*\|^2 + 1)^{1/2}}. \tag{5}$$

Consider the singular value decomposition (SVD) of $B(\theta)$:

$$B(\theta) = U(\theta)\begin{bmatrix} \Sigma(\theta) \\ 0 \end{bmatrix} V^{\mathrm{T}}(\theta),$$

where $U(\theta) \in R^{m\times m}$, $V(\theta) \in R^{K\times K}$ are orthogonal and $\Sigma(\theta)$ is a diagonal matrix with diagonal entries $\sigma_1, \dots, \sigma_K$ such that

$$\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_K \geq 0$$

are the singular values of B (assuming $B(\theta)$ a full rank matrix). We define $b_1(\theta) \in R^K$ and $b_2(\theta) \in R^{m-K}$ as

$$\begin{bmatrix} b_1(\theta) \\ b_2(\theta) \end{bmatrix} = U^{\mathrm{T}}(\theta)Y.$$

The expression (5) for α^* can be rewritten as

$$\alpha^* = V(\theta^*)(\Sigma^2(\theta^*) + \mu^*\rho I)^{-1}\Sigma(\theta^*)b_1(\theta^*),$$

and therefore, $\|\alpha^*\| = \|\Sigma(\theta^*)(\Sigma^2(\theta^*) + \mu^*\rho I)^{-1}b_1(\theta^*)\|$. Also,

$$\begin{aligned} Y - B(\theta^*)\alpha^* &= U(\theta^*)U^{\mathrm{T}}(\theta^*)Y - U(\theta^*)\begin{bmatrix} \Sigma(\theta^*) \\ 0 \end{bmatrix} V^{\mathrm{T}}(\theta^*)\alpha^*, \\ &= U(\theta^*)\begin{bmatrix} b_1(\theta^*) - \Sigma^2(\theta^*)(\Sigma^2(\theta^*) + \mu^*\rho I)^{-1} b_1(\theta^*) \\ b_2(\theta^*) \end{bmatrix}, \\ &= U(\theta^*)\begin{bmatrix} \mu\rho(\Sigma^2(\theta^*) + \mu^*\rho I)^{-1} b_1(\theta^*) \\ b_2(\theta^*) \end{bmatrix}. \end{aligned}$$

Hence,

$$\| Y - B(\theta^*)\alpha^* \| = \sqrt{\| b_2(\theta^*)\|^2 + (\mu^*)^2 \rho^2 \| (\Sigma^2(\theta^*) + \mu^*\rho I)^{-1} b_1(\theta^*)\|^2}.$$

Now, the expression for μ^* reduces to

$$\mu^* = \left(\frac{\| b_2(\theta^*)\|^2 + (\mu^*)^2 \rho^2 \| (\Sigma^2(\theta^*) + \mu^*\rho I)^{-1} b_1(\theta^*)\|^2}{1 + \| \Sigma(\theta^*)(\Sigma^2(\theta^*) + \mu^*\rho I)^{-1} b_1(\theta^*)\|^2} \right)^{1/2}.$$

The above expression implies

$$\begin{aligned} &(\mu^*)^2 (1 + \| \Sigma(\theta^*)(\Sigma^2(\theta^*) + \mu^*\rho I)^{-1} b_1(\theta^*)\|^2) - \| b_2(\theta^*)\|^2 \\ &\quad - (\mu^*)^2 \rho^2 \| (\Sigma^2(\theta^*) + \mu^*\rho I)^{-1} b_1(\theta^*)\|^2 = 0, \end{aligned}$$

and can be further rewritten as

$$1 + b_1^{\mathrm{T}}(\theta^*)(\Sigma^2(\theta^*) - \rho^2 I)(\Sigma^2(\theta^*) + \mu^*\rho I)^{-2} b_1(\theta^*) - \frac{\| b_2(\theta^*)\|^2}{(\mu^*)^2} = 0.$$

Note that, $\theta^* = \arg\min_\theta \Psi(\alpha^*, \theta)$, with $C\theta^* \geq h$. Assume that

$$Z(\theta, \mu) = \begin{bmatrix} Y/\rho^{1/2} \\ (0)_{K\times 1} \\ (\mu)^{1/2}(\theta - \theta_0) \end{bmatrix}, \quad H(\theta, \mu) = \begin{bmatrix} B(\theta)/\rho^{1/2} \\ (\mu)^{1/2}(I)_K \\ (0)_{L\times K} \end{bmatrix},$$

and $r(\theta, \mu) = Z(\theta, \mu) - H(\theta, \mu)\alpha^*$. Now, θ^* can be formulated as the solution of a nonlinear least squares constrained problem:

$$\theta^* = \arg\min_\theta \| r(\theta, \mu^*)\|^2, \quad C\theta \geq h. \tag{6}$$

We introduce two nonlinear functions:

$$R(\theta,\mu)=\left(\frac{\|b_2(\theta)\|^2+\mu^2\rho^2\|(\Sigma^2(\theta)+\mu\rho I)^{-1}b_1(\theta)\|^2}{1+\|\Sigma(\theta)(\Sigma^2(\theta)+\mu\rho I)^{-1}b_1(\theta)\|^2}\right)^{1/2},$$

and secular function $\mathscr{G}(\theta,\mu)$

$$\mathscr{G}(\theta,\mu)=1+b_1^{\mathrm{T}}(\theta)(\Sigma^2(\theta)-\rho^2 I)(\Sigma^2(\theta)+\mu\rho I)^{-2}b_1(\theta)-\frac{\|b_2(\theta)\|^2}{\mu^2}.$$

Therefore, $\mathscr{G}(\theta^*,\mu^*)=0$.

For the computation of θ^*, we need to solve the constrained nonlinear optimization problem (6), using some numerical optimization technique. However, in some cases it is quite possible that chosen numerical method for solving the optimization problem does not converge or converges locally. Therefore, we define sub-optimal solution $(\bar{\alpha},\bar{\theta})$ such that $\mathscr{G}(\bar{\theta},\bar{\mu})=0$, where $\bar{\theta}$ may not be the global minimizer of $\|r(\theta,\mu^*)\|^2$. However, note that the sub-optimal solution $(\bar{\theta},\bar{\alpha})$ is robust in the sense that

$$\bar{\alpha}=\arg\min_{\alpha}\max_{\|\Delta B\,\Delta Y\|_F\le\rho}\|(B(\bar{\theta})+\Delta B)\alpha-(Y+\Delta Y)\|.$$

It can be seen that $\mathscr{G}(\theta,\mu)$ and $R(\theta,\mu)$ are related by

$$\mathscr{G}(\theta,\mu)=\left(1-\left(\frac{R(\theta,\mu)}{\mu}\right)^2\right)(1+\|\Sigma(\theta)(\Sigma^2(\theta)+\mu\rho I)^{-1}b_1(\theta)\|^2).$$

A solution of fuzzy identification problem, say $(\bar{\alpha},\bar{\theta})$, will be robust if it is the root of secular function, i.e. $\mathscr{G}(\bar{\theta},\bar{\mu})=0$. Also, note that $\bar{\alpha}=(H(\bar{\theta},\bar{\mu}))^{\dagger}Z(\bar{\theta},\bar{\mu})$ and $\bar{\mu}=R(\bar{\theta},\bar{\mu})$. Therefore,

$$\frac{\partial\mathscr{G}(\bar{\theta},\bar{\mu})}{\partial\mu}=\frac{2(1-\partial R(\bar{\theta},\bar{\mu})/\partial\mu)(1+\|\Sigma(\bar{\theta})(\Sigma^2(\bar{\theta})+\bar{\mu}\rho I)^{-1}b_1(\bar{\theta})\|^2)}{\bar{\mu}},$$

where $\frac{\partial\mathscr{G}(\bar{\theta},\bar{\mu})}{\partial\mu}$ is the partial derivative value of secular function $\mathscr{G}(\theta,\mu)$ with respect to μ at point $(\bar{\theta},\bar{\mu})$. It follows from above equation that

$$\frac{\partial R(\bar{\theta},\bar{\mu})}{\partial\mu}=1-\frac{\bar{\mu}(\partial\mathscr{G}(\bar{\theta},\bar{\mu})/\partial\mu)}{2(1+\|\Sigma(\bar{\theta})(\Sigma^2(\bar{\theta})+\bar{\mu}\rho I)^{-1}b_1(\bar{\theta})\|^2)}. \tag{7}$$

It was shown by Chandrasekaran (1998) that if $\bar{\mu}>0$ is the positive root of $g(\mu)=b_1^{\mathrm{T}}(\Sigma^2-\rho^2 I)(\Sigma^2+\mu I)^{-2}b_1-\frac{\rho^2}{\mu^2}\|b_2\|^2$, then $\frac{dg(\bar{\mu})}{d\mu}>0$. Now, making similar arguments it can be shown that $\partial\mathscr{G}(\bar{\theta},\bar{\mu})/\partial\mu > 0$. Therefore, it follows from Eq. (7) that

$$\frac{\partial R(\bar{\theta},\bar{\mu})}{\partial\mu} < 1.$$

Differentiating $\mathscr{G}(\theta,\mu)$, we have

$$\frac{\partial\mathscr{G}(\theta,\mu)}{\partial\mu} = \frac{2\|b_2(\theta)\|^2}{\mu^3} - 2\rho b_1^{\mathrm{T}}(\theta)(\Sigma^2(\theta) - \rho^2 I)(\Sigma^2(\theta) + \mu\rho I)^{-3} b_1(\theta)$$

and therefore

$$\bar{\mu}\frac{\partial\mathscr{G}(\bar{\theta},\bar{\mu})}{\partial\mu} = 2 + 2b_1^{\mathrm{T}}(\bar{\theta})\Sigma^2(\bar{\theta})(\Sigma^2(\bar{\theta}) - \rho^2 I)(\Sigma^2(\bar{\theta}) + \bar{\mu}\rho I)^{-3} b_1(\bar{\theta}),$$

using $\mathscr{G}(\bar{\theta},\bar{\mu}) = 0$. Now Eq. (7) becomes

$$\frac{\partial R(\bar{\theta},\bar{\mu})}{\partial\mu} = (\rho^2 + \bar{\mu}\rho)\frac{b_1^{\mathrm{T}}(\bar{\theta})\Sigma^2(\bar{\theta})(\Sigma^2(\bar{\theta}) + \bar{\mu}\rho I)^{-3} b_1(\bar{\theta})}{1 + \|\Sigma(\bar{\theta})(\Sigma^2(\bar{\theta}) + \bar{\mu}\rho I)^{-1} b_1(\bar{\theta})\|^2} > 0.$$

It follows that

$$0 < \frac{\partial R(\bar{\theta},\bar{\mu})}{\partial\mu} < 1. \tag{8}$$

Based on the above analysis (inequality (8) and Eq. (6)), following iterative scheme is suggested for the robust identification of fuzzy parameters (α,θ) in presence of uncertainty in the identification data.

1. Choose α_0, θ_0, iteration count $k = 0$ and a positive constant ϵ.
2. Compute $\mu_{k+1} = R(\theta_k,\mu_k)$.
3. Compute $\alpha_{k+1} = (H(\theta_k,\mu_{k+1}))^{\dagger} Z(\theta_k,\mu_{k+1})$.
4. If $|\mathscr{G}(\theta_k,\mu_{k+1})| > \epsilon$
   (a) Compute $Z(\theta_k,\mu_{k+1})$ and $H(\theta_k,\mu_{k+1})$.
   (b) Let $r(\theta_k) = Z(\theta_k,\mu_{k+1}) - H(\theta_k,\mu_{k+1})\alpha_{k+1}$.
   (c) Compute Jacobian $r'(\theta_k)$ by the method of finite differences, which is a full rank matrix (due to regularization).
   (d) Solve $s^* = \arg\min_s[\|r(\theta_k) + r'(\theta_k)s\|^2,\ Cs \geq h - C\theta_k]$, using the method suggested by Lawson and Hanson (1995).
   (e) $\theta_{k+1} = \theta_k + s^*$.
5. $k := k + 1$.

The above iterative allows the updating of antecedent parameters θ only if $|\mathscr{G}(\theta_k, \mu_{k+1})| > \epsilon$. Actually, ϵ is chosen a small positive constant. So, there is no updating of θ_k near the optimal point ($\mathscr{G}(\theta_k, \mu_{k+1}) \approx 0$). This facilitate the damping of the oscillations near optimal point and convergence of the proposed scheme to at least sub-optimal solution is ensured. Assume that there exists a positive integer N such that $\theta_{N-1} = \bar{\theta}$ and $|\mathscr{G}(\theta_{N-1}, \mu_N)| < \epsilon$. Now, noting that $0 < \partial R(\bar{\theta}, \bar{\mu})/\partial\mu < 1$, we say that the iterative scheme

$$\mu_{k+1} = R(\bar{\theta}, \mu_k), \quad k = N, N+1, \ldots$$

is locally convergent to $\bar{\mu}$.

5. Numerical examples

To illustrate the proposed method of robust fuzzy identification, a number of simulation studies for the reconstruction and rule-based modelling of a one-dimensional spectral data function are considered. The goal of task is to construct a transparent rule based model from noisy measurements with a known level of perturbation in the data. The spectral data function was simulated as

$$f(x) = 1.2\mathrm{e}^{-(x-4.8)(x-5.8)/0.7} - 1.2\mathrm{e}^{-(x+3.5)^2} + 0.08x,$$

for $x \in [-10, 10]$. The output of the function was added with random numbers, chosen from a uniform distribution on the interval (–0.1, 1.0). Initially, the input space is divided into 10 symmetrical trapezoidal membership functions distributed uniformly over the input space. We apply the proposed method for the robust identification of spectral data function with the perturbation value ($\rho = 0.4293$) and taking $\epsilon = 0.05$. Figure 3 shows the convergence behavior of the proposed method and Figure 5 shows the fitting of data with fuzzy model. The interpretability of fuzzy system remains preserved during and after identification, as a consequence of solving constrained optimization problem. Figure 4 shows the shape of membership functions after identification. Representing the 10 membership functions as $(A1, A2, \ldots, A10)$, the identified fuzzy rule base is stated in Table 1.

Note that once the antecedent parameters (θ) are known, the consequent parameters can be alternatively estimated by solving the SOCP problem. Therefore, we also considered the estimation of consequent parameters by solving the SOCP problem using SEDUMI 1.02 (a matlab toolbox for optimization over symmetric cones). However, the results obtained are same as shown in Table 1.

We repeat the experiment for different levels of data uncertainty. Figure 6 shows the results obtained and Figure 7 shows the plot between optimal regularization parameter μ^* and uncertainty level ρ. The plot between μ^* and ρ helps us to choose

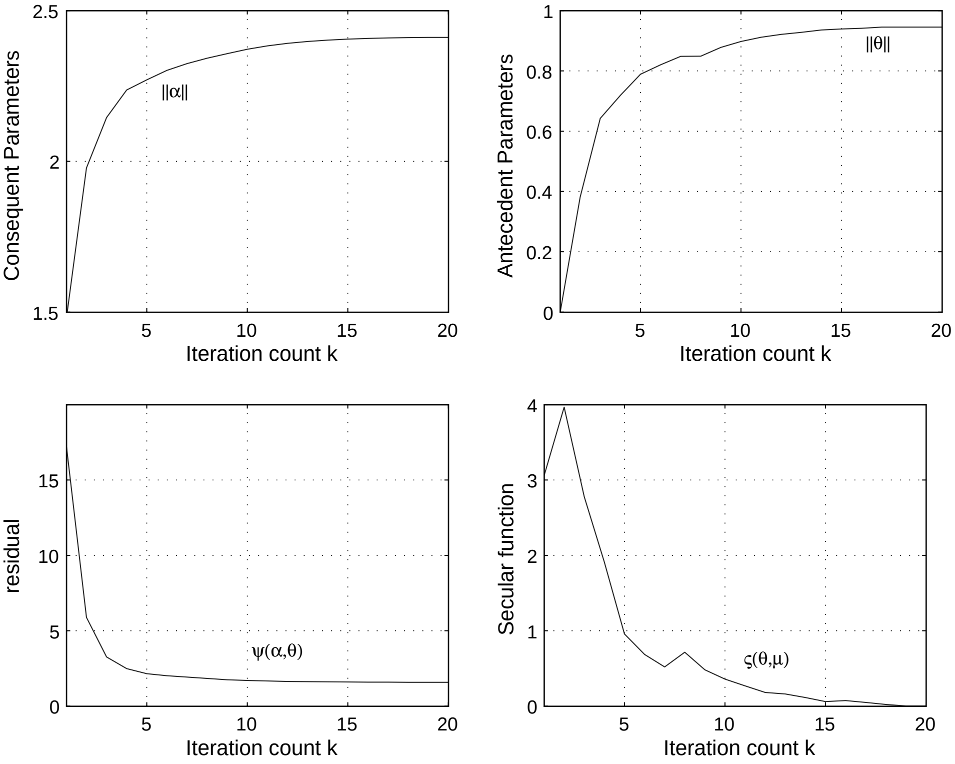


Figure 3. Convergence of the parameters.

the optimal value of regularization parameter for a given level of uncertainty in the identification data.

6. Fuzzy Approximation of Physical Fitness

The aim of problem is to approximate the physical fitness or to identify the functional relationship between physical fitness and other physiological parameters with a fuzzy inference system (Figure 8). This complex and uncertain functional relationship is not so easy to define even in linguistic terms due to the uncertainty, natural diversity and subjectivity in the opinion of individuals. So, there is a need of developing some robust techniques for the automatic construction of such an expert system from example data.

The proposed method was used to identify the fuzzy expert system. The real world clinical data consists of various physiological variables: body mass index (BMI),

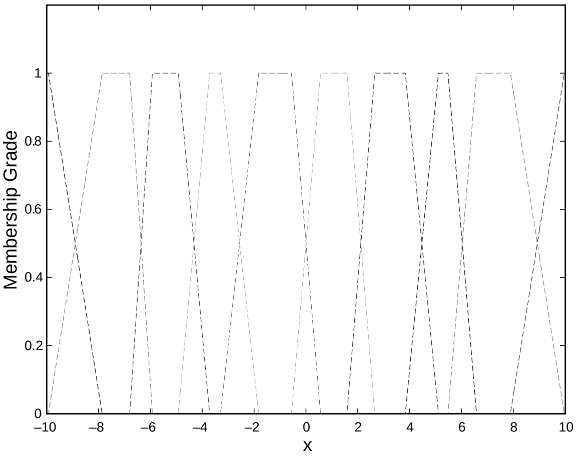


Figure 4. Various membership functions after identification.

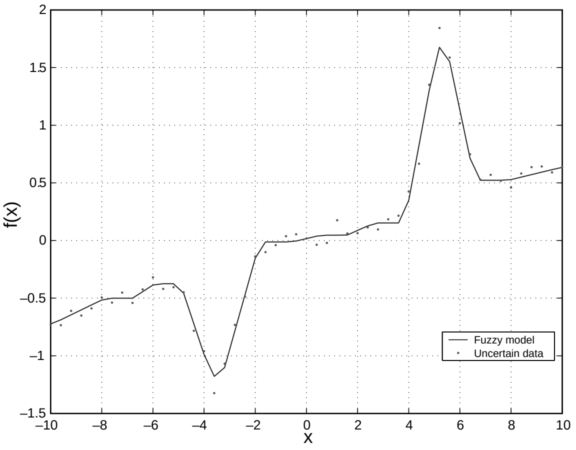
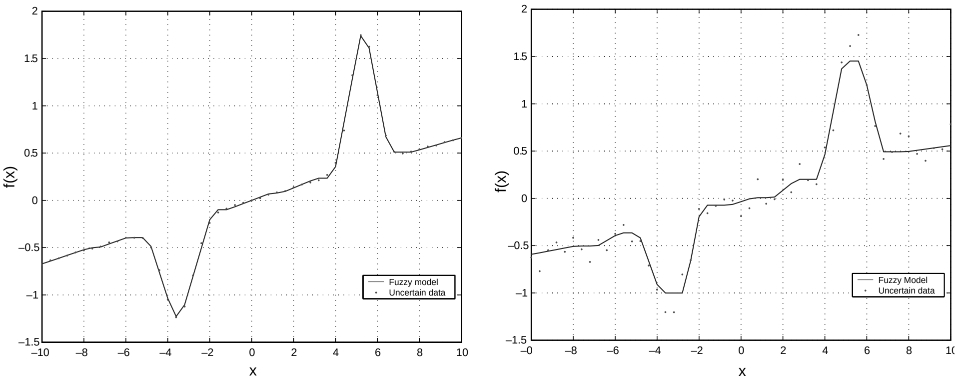


Figure 5. Regularized (optimal for robustness) fuzzy approximation.

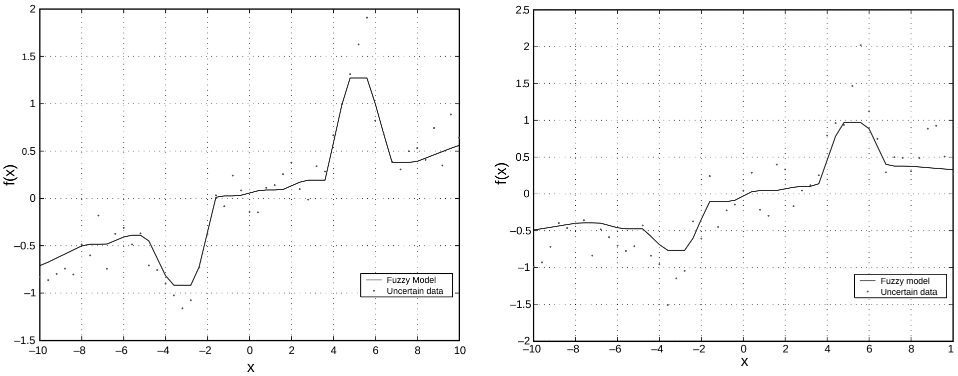
absolute $VO_{2\,max}$ and relative $VO_{2\,max}$. This set of three parameters are the inputs of the fuzzy system and the output of fuzzy system is the quantification of physical fitness on a scale ranging from 0 to 1. Our training data consists of 80 patients and a comment on the physical fitness of these patients was made quantitatively by a medical expert based on his experience. However, we assume an uncertainty level of 10% in the comment of medical expert.

Table 1. Fuzzy rules obtained from uncertain data.

Rule	Antecedent	Consequent	Proposed method	SOCP
R1	If x is $A1$	then y is equal to	−0.7235	−0.7235
R2	If x is $A2$	then y is equal to	−0.5009	−0.5009
R3	If x is $A3$	then y is equal to	−0.3748	−0.3748
R4	If x is $A4$	then y is equal to	−1.1778	−1.1778
R5	If x is $A5$	then y is equal to	−0.0130	−0.0130
R6	If x is $A6$	then y is equal to	0.0464	0.0464
R7	If x is $A7$	then y is equal to	0.1519	0.1519
R8	If x is $A8$	then y is equal to	1.6761	1.6761
R9	If x is $A9$	then y is equal to	0.5223	0.5223
R10	If x is $A10$	then y is equal to	0.6347	0.6347



(a) a. For $\rho = 0.0407$ (b) b. For $\rho = 0.7552$



(c) a. For $\rho = 1.2431$ (d) b. For $\rho = 1.7320$

Figure 6. Robust fuzzy identification for different levels of uncertainty.

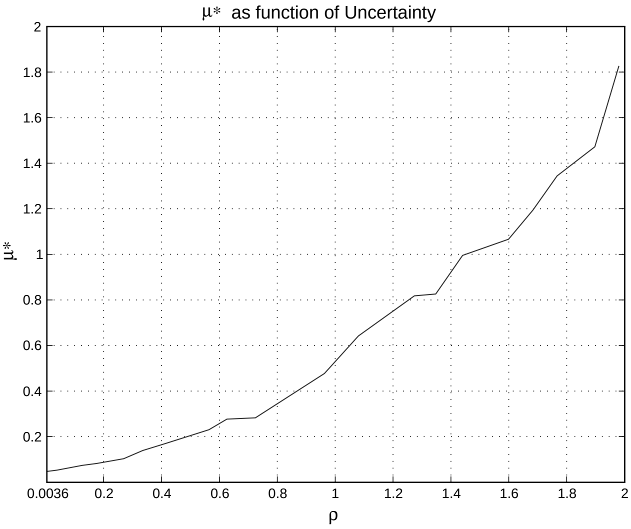


Figure 7. Optimal regularization parameter (μ^*) as function of ρ.

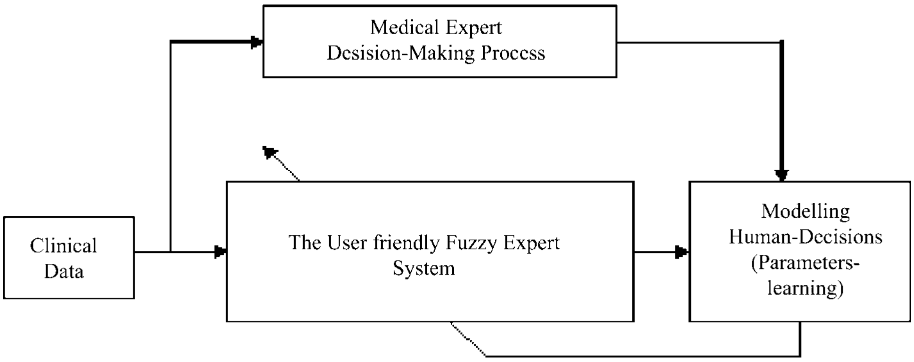


Figure 8. Fuzzy approximation of physical fitness.

The whole range of all input parameters has been divided into two membership functions with linguistic terms *Normal* and *Abnormal*. This results the fuzzy inference system to consist of eight rules. To preserve the interpretability of fuzzy system, the membership functions were prevented from overlapping by putting constraints on the position of knots. Figure 9 shows the shape of identified membership functions and Table 2 shows the identified interpretable fuzzy rule base approximating the

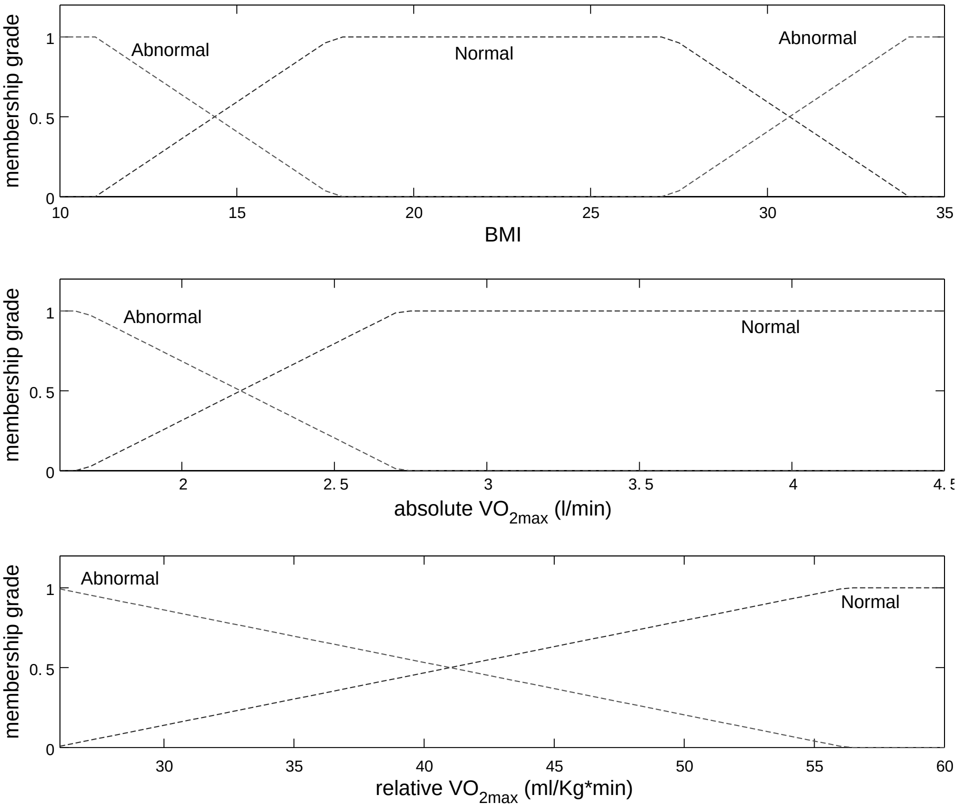


Figure 9. Shape of membership functions after identification.

Table 2. Fuzzy rules identified for physical fitness approximation.

Rule	BMI	$\text{AbsVO}_{2\max}$	$\text{relVO}_{2\max}$	Physical fitness
R1	Normal	Abnormal	Abnormal	0.1526
R2	Normal	Abnormal	Normal	0.9850
R3	Normal	Normal	Abnormal	0.1649
R4	Normal	Normal	Normal	1.0027
R5	Abnormal	Abnormal	Abnormal	0
R6	Abnormal	Abnormal	Normal	0
R7	Abnormal	Normal	Abnormal	0.1621
R8	Abnormal	Normal	Normal	0.9911

functional relationship between physical fitness and other physiological parameters. The identified fuzzy expert system was tested on another set of 80 patients for the

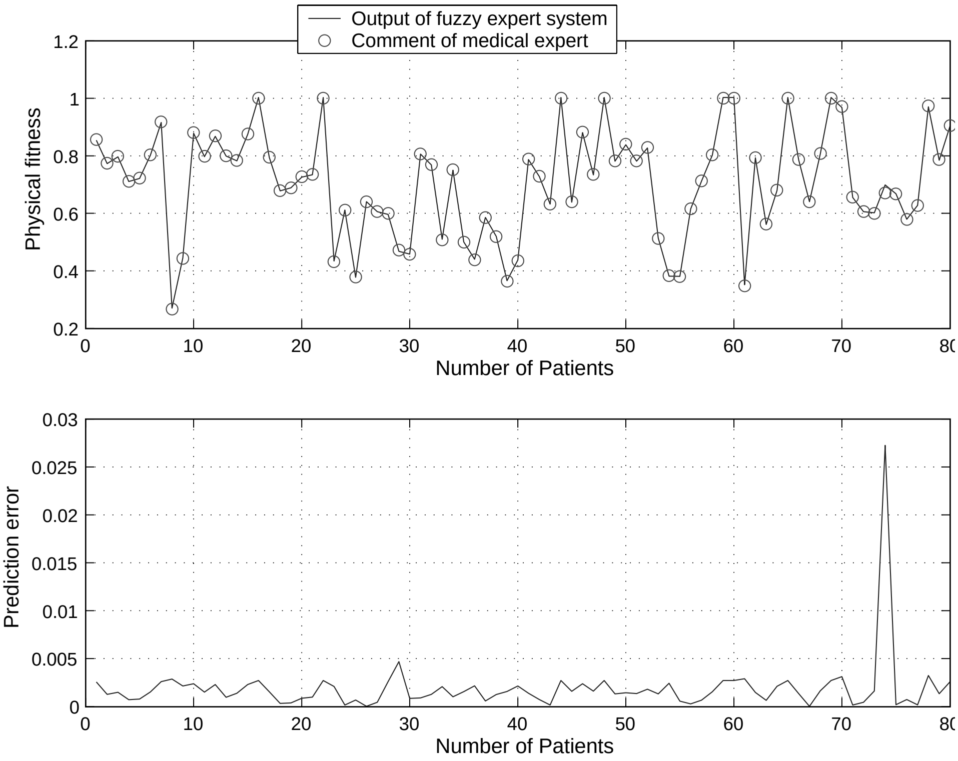


Figure 10. Testing of identified fuzzy expert system on 80 patients.

validation of our model. Figure 10 shows the remarkable results obtained, in predicting the physical fitness of patients using the identified fuzzy model.

7. Conclusion

This paper has outlined the robust (min–max) identification of fuzzy parameters in presence of uncertainty in the identification data by solving an equivalent regularized identification problem. An iterative method to solve the regularized identification problem was described and was shown to provide good results in application to the problems of modelling and approximation.

The proposed method was used to develop a fuzzy expert system for the approximation of physical fitness functional relationship with some of the physiological parameters. The developed fuzzy expert system proved successful in the modelling of human-experiences and handling the uncertainty lying in the human decisions.

Notes

1. Corresponding author.

References

Babuska, R. (2000). "Construction of Fuzzy Systems-interplay between Precision and Transparency," In *Proc. ESIT 2000*, Aachen, pp. 445–452.

Bodenhofer, U. and P. Bauer. (2000). "Towards an Axiomatic Treatment of Interpretability," In *Proc. IIZUKA2000*, Iizuka, pp. 334–339.

Burger, M., H. W. Engl, J. Haslinger, and U. Bodenhofer. (2002). "Regularized Data-driven Construction of Fuzzy Controllers," *Journal of Inverse and Ill-posed Problems* 10, 319–344.

Chandrasekaran, S., G. H. Golub, M. Gu, and A. H. Sayed. (1998). "Parameter Estimation in the Presence of Bounded Data Uncertainties," SIMAX, 19, 235–252.

Demoment, G. (1989). "Image Reconstruction and Restoration: Overview of Common Estimation Problems," *IEEE Transaction on Acoustic Speech and Signal Processing* 37, 3667–3672.

Elden, L. (1977). "Algorithm for Regularization of Ill-conditioned Least Squares Problems," *BIT* 17, 134–145.

Espinosa, J. and J. Vandewalle. (2000). "Constructing Fuzzy Models with Linguistic Integrity from Numerical Data-AFRELI Algorithm," *IEEE Transactions on Fuzzy Systems* 8(5), 591–600.

Furuya, M. (1989). "Optimization of Weighting Constant for Regularization in Least Squares System Identification," *Transactions on IEICE A J72A*, 1012–1015.

Ghaoui, L. El. and H. Lebret. (1997). "Robust Solutions to Least-Squares Problems with Uncertain Data," *SIAM Journal of Matrix Analysis and Applications* 18, 1035–1064.

Golub, G. H. and C. F. Van Loan. (1989). "*Matrix Computations*", 2nd Edition "Baltimore, MD: Johns Hopkins University Press.

Hanke M. and P. C. Hansen. Regularization Methods for Large-scale Problems. *Surveys on Mathematics for Industry* 3, 253–315.

Hunt, B. R. (1973). "The Application of Constrained Least Square Estimation to Image Restoration by Digital Computer," *IEEE Transactions on Computers* C-22, 805–812.

Lawson, C. L. and R. J. Hanson. (1995). "*Solving Least Squares Problems*", Philadelphia: SIAM Publications.

Nashed, M. Z. (1981). "Operator-theoretic and Computational Approaches to Ill-posed Problems with Applications to Antenna Theory," *IEEE Transactions on Antenna and Propagation* 29, 220–231.

Setnes, M., R. Babuka, and H. B. Verbruggen. (1998). "Rule-based Modeling: Precision, and Transparency," *IEEE Transactions on Systems Man and Cybernetics Part C: Applications and Reviews*, 28, 165–169.

Takagi, T. and M. Sugeno. (1985). "Fuzzy Identification of Systems and its Applications to Modeling and Control," *IEEE Transactions on Systems Man and Cybernetics* 15, 116–132.

Zadeh, L. A. (1973). "Outline of a new Approach to the Analysis of Complex Systems and Decision Processes," *IEEE Transactions on Systems Man and Cybernetics* 3(1), 28–44.