RBF kernel method and its applications to clinical data

Emma Perracchione[a] · Ilaria Stura[b]

Abstract

In this paper, basing our considerations on kernel-based approaches, we propose a new strategy allowing to approximate the prostate cancer dynamics. In particular, starting from several measurements of a specific biomarker, we estimate the tumor growth rate. To achieve this aim, we pre-process data via Radial Basis Function (RBF) interpolation. A careful choice of the basis function and of its shape parameter enables us to obtain reliable approximations of the cancer evolution. Numerical evidence supports our findings.

1 Introduction

Over the last years, the topic of numerical approximation of multivariate data has gained popularity in various disciplines, such as numerical solution of PDEs, image registration, neural networks, optimization, statistics, finance and modeling 3D objects (see e.g. [1, 3, 19, 20, 24, 27]).

In this paper, we consider a clinical application. Specifically, we deal with the problem of approximating the growth rate of the prostate cancer, one of the most common tumors in men. Since it is characterized by a slow growth, it is usually detected in the early stage of the disease. This enables clinicians to promptly intervene with radiotherapy or surgery, i.e. Radical Prostatectomy (RP).

However, a relapse after a RP occurs in the $25-30\%$ of cases. The recurrence can be seen as a new tumor and, depending on the patient, different growth characteristics are observed. The relapse can be diagnosed by measuring the Prostate Specific Antigen (PSA) value with a blood exam. Moreover, thanks to this biomarker, we are able to estimate the growth rate of the tumor. To reach this aim, at first we model the cancer dynamics with a Gompertzian function. Then, we validate such model with challenging clinical real data taken from *Eureka1 data set* [12], a collection of data (e.g. PSA, tumor stage, surgical methods) about prostatectomized patients of Piemonte Region (Italy).

Summarizing, given a PSA series, to estimate the growth rate of the tumor, we develop a method consisting of two steps:

1. reconstruct the whole PSA curve with a RBF-based method [10, 23];

2. use the so constructed curve to estimate the growth rate of the prostate cancer via a stochastic optimization technique [15, 16].

In particular, here we focus on the first item. The choice of RBF-based methods follows from the fact that, since they are based only on a set of independent data, their implementation turns out to be easy and cheap from a computational point of view. Indeed, the cost of a mesh generation is eliminated.

It is well-known that the value of the shape parameter of the considered RBF affects the accuracy of the fit. In fact, *good* values of the shape parameter depend on the specific PSA series, i.e. they can truly vary from one patient to another one. Therefore, we take into account the problem of suitably selecting *safe* shape parameters. Specifically, they are selected by minimizing an error bound depending on the so-called *power function* [26].

The guidelines of the paper are as follows. In Section 2 the scattered data interpolation problem is presented. Section 3 is devoted to the presentation of both the clinical scenario and the mathematical model. Finally, Sections 4 and 5 deal with results and conclusions, respectively.

2 The scattered data interpolation problem

In order to model the cancer dynamics, we need to consider a procedure enabling us to fit a function given few samples. More formally, given $X_N = \{x_i, i = 1, \ldots, N\} \subseteq \Omega$ a set of N distinct data points or nodes, arbitrarily distributed in a domain $\Omega \subseteq \mathbb{R}^d$, with an associated set $F_N = \{f_i = f(x_i), i = 1, \ldots, N\} \subseteq \mathbb{R}$ of data values or function values, the scattered data interpolation problem consists in finding a function $R : \Omega \longrightarrow \mathbb{R}$ satisfying

$$R(x_i) = f_i, \quad i = 1, \ldots, N. \tag{1}$$

To solve such problem we consider kernel-based interpolation methods [2, 8, 9, 10, 23].

[a]Department of Mathematics "G. Peano", University of Torino
[b]Department of Neuroscience "R. Levi Montalcini", University of Torino

2.1 Kernel-based interpolation methods

In order to obtain effective and reliable approximations, we need to focus on positive definite and symmetric kernels $\Phi : \Omega \times \Omega \longrightarrow \mathbb{R}$. Moreover, we also require that there exist:

    i. a positive shape parameter ε;

    ii. a function $\phi : \mathbb{R}^+ \longrightarrow \mathbb{R}$ such that

$$\Phi(\boldsymbol{x}, \boldsymbol{y}) = \phi_\varepsilon(||\boldsymbol{x} - \boldsymbol{y}||_2) = \phi(\varepsilon||\boldsymbol{x} - \boldsymbol{y}||_2), \quad \text{for all} \quad \boldsymbol{x}, \boldsymbol{y} \in \Omega,$$

    where $|| \cdot ||_2$ represents the Euclidean norm.

A kernel Φ satisfying the two above mentioned properties is called a radial kernel. In this context, the RBF approximation takes place. Indeed, given a positive definite radial kernel Φ, we construct the following data-dependent basis

$$\mathcal{T}_{X_N} = \{\Phi(\cdot, \boldsymbol{x}_1), \ldots, \Phi(\cdot, \boldsymbol{x}_N)\}. \tag{2}$$

Using (2), the interpolant R can be expressed as

$$R(\boldsymbol{x}) = \sum_{k=1}^{N} \lambda_k \Phi(\boldsymbol{x}, \boldsymbol{x}_k), \quad \boldsymbol{x} \in \Omega. \tag{3}$$

The approximant (3) is constructed by imposing the interpolation conditions (1). As a consequence, it is computed by solving the following linear system

$$\begin{pmatrix} \Phi(\boldsymbol{x}_1, \boldsymbol{x}_1) & \cdots & \Phi(\boldsymbol{x}_1, \boldsymbol{x}_N) \\ \vdots & \ddots & \vdots \\ \Phi(\boldsymbol{x}_N, \boldsymbol{x}_1) & \cdots & \Phi(\boldsymbol{x}_N, \boldsymbol{x}_N) \end{pmatrix} \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_N \end{pmatrix} = \begin{pmatrix} f_1 \\ \vdots \\ f_N \end{pmatrix}, \tag{4}$$

or simply

$$K\boldsymbol{\lambda} = \boldsymbol{f},$$

where K is the so-called *kernel matrix*. The solution reconstructed in this way is a function of the native space

$$\mathcal{N}_\Phi(\Omega) = \text{span}\{\Phi(\cdot, \boldsymbol{x}), \boldsymbol{x} \in \Omega\}.$$

We remark that the system (4) admits a unique solution if and only if Φ is a strictly positive definite kernel [9].

In order to give error bounds, we need to write the interpolant R in *Lagrange form*, i.e. using the so-called *cardinal basis functions* [26].

To this aim, we recall the following theorem [9].

Theorem 2.1. *Suppose Φ is a strictly positive definite kernel. Then, for any set $X_N = \{\boldsymbol{x}_i, i = 1, \ldots, N\} \subseteq \Omega$ of distinct data points, there exist cardinal functions $u_k^* \in \mathcal{T}_{X_N}$ such that $u_k^*(\boldsymbol{x}_i) = \delta_{ik}$, i.e.*

$$u_k^*(\boldsymbol{x}_i) = \begin{cases} 1 & \text{if } i = k, \\ 0 & \text{if } i \neq k. \end{cases}$$

From Theorem 2.1, we can express the interpolant in cardinal form, i.e.

$$R(\boldsymbol{x}) = \sum_{k=1}^{N} f(\boldsymbol{x}_k) u_k^*(\boldsymbol{x}), \quad \boldsymbol{x} \in \Omega.$$

2.2 Error bounds

After expressing the interpolant by means of the cardinal functions, for any strictly positive definite kernel $\Phi \in C(\Omega \times \Omega)$, any set of distinct points $X_N = \{\boldsymbol{x}_i, i = 1, \ldots, N\} \subseteq \Omega$, and any vector $\boldsymbol{u} \in \mathbb{R}^N$, we define the quadratic form [9]

$$Q(\boldsymbol{u}) = \Phi(\boldsymbol{x}, \boldsymbol{x}) - 2\sum_{k=1}^{N} u_k \Phi(\boldsymbol{x}, \boldsymbol{x}_k) + \sum_{i=1}^{N} \sum_{k=1}^{N} u_i u_k \Phi(\boldsymbol{x}_i, \boldsymbol{x}_k).$$

Definition 2.1. Suppose $\Omega \subseteq \mathbb{R}^d$ and $\Phi \in C(\Omega \times \Omega)$ is strictly positive definite. For any distinct points $X_N = \{\boldsymbol{x}_i, i = 1, \ldots, N\} \subseteq \Omega$ the power function is defined as

$$\left[P_{\Phi, X_N}(\boldsymbol{x})\right]^2 = Q(\boldsymbol{u}^*(\boldsymbol{x})),$$

where $\boldsymbol{u}^*$ is the vector of cardinal functions from Theorem 2.1.

Following the guidelines of [9], the power function can be computed as

$$P_{\Phi, X_N}(\boldsymbol{x}) = \sqrt{\Phi(\boldsymbol{x}, \boldsymbol{x}) - (b(\boldsymbol{x}))^T K^{-1} b(\boldsymbol{x})}, \tag{5}$$

where $b = (\Phi(\cdot, x_1), \ldots, \Phi(\cdot, x_N))^T$.

We are now able to state the following theorem.

Theorem 2.2. *Let $\Omega \subseteq \mathbb{R}^d$ and $\Phi \in C(\Omega \times \Omega)$ be a strictly positive definite kernel. Suppose that the points $X_N = \{\boldsymbol{x}_i, i = 1, \ldots, N\} \subseteq \Omega$ are distinct. Then*

$$|f(\boldsymbol{x}) - R(\boldsymbol{x})| \le P_{\Phi, X_N}(\boldsymbol{x}) ||f||_{\mathcal{N}_\Phi(\Omega)}, \quad \boldsymbol{x} \in \Omega, \tag{6}$$

where $f \in \mathcal{N}_\Phi(\Omega)$.

For further refinements of the error bound stated in Theorem 2.2, refer to [9, 23].

Note that Theorem 2.2 provides a criterion for choosing good values of the shape parameter ε. This turns out to be the key step for an accurate approximation. Precisely, following Theorem 2.2, the error is decomposed into two components: a component independent of the data function f and one depending on f. Therefore, we can minimize the power function with respect to the shape parameter. Such approach has the advantage of being independent of the knowledge of f.

With this strategy we are able to select a suitable shape parameter for the approximation process. Anyway, since the second component of the error is not minimized, this is not an optimal approach. For further developments about safe choices of the shape parameter or for stable computation of the interpolant in the flat limit $\varepsilon \longrightarrow 0$ refer to [5, 7, 10, 11].

In the next section we focus on an application to clinical data. Precisely, we consider several samples of PSA values.

3   The clinical scenario

This section is devoted to present the clinical background. In the following subsection we give a brief outline about the prostate cancer scenario, while the mathematical model used to simulate the tumor growth is shown in Subsection 3.2.

3.1   The prostate cancer

The prostate cancer has a very slow growth, in both primary and secondary tumor. It can be early diagnosed with various techniques, such as the PSA and the digital rectal exams. While before the surgery an increase of the PSA does not necessarily imply that the patient has a tumor, the recurrence in prostatectomized patients can be monitored by dosing it [6]. In fact, the PSA is an antigen produced only by prostate cells, i.e. its value should be zero after a RP; if it is greater than $0.2\ ng/mL$, a biochemical recurrence occurs. This value can be easily found by means of a cheap blood exam without side effects. Thus, during the follow-up, a patient should test his PSA value each 3-6 months.

These assumptions enable us to simulate the cancer dynamics with our proposed method. In fact, in case of more aggressive tumors, only few measurements of the volume are available, because it can be estimated only via medical images, which are very expensive and can have side effects on the patient. Moreover, no free tumor growth is observed in humans. This is due to the fact that the mass is reduced with therapies as soon as it is discovered.

3.2   The mathematical model

This section is devoted to present the mathematical model used to simulate the cancer growth. We focus our attention on the Phenomenological Universalities (PUN) law [4]. It includes different kinds of growth, namely the Malthus, Gompertz and von Bertalanffy laws and can be applied in a wide variety of fields, such as in biology and medicine, see [13, 14].

After preliminary studies [22], the best candidate to model prostate cancer turns out to be the Gompertzian function

$$f(x) = f_0 \exp\left[\log\left(f_\infty/f_0\right)(\exp(-\alpha x) - 1)\right], \tag{7}$$

where α is the growth rate, f_0 is the initial PSA value and f_∞ is the carrying capacity, i.e. the upper bound that can be reached by the tumor for $x \to \infty$. In other words, it is characterized by an initial exponential growth, a progressive velocity decrease and finally the achievement of a carrying capacity due to physical barriers or lack of nutrients.

We point out that (7) can be seen as a solution of the following ODE

$$\begin{cases} \dfrac{df(x)}{dx} = C \exp(-\alpha x) f(x), \\ f(0) = f_0, \end{cases}$$

where C, which is *somehow* related to the carrying capacity, is given by

$$C = \alpha \log\left(\frac{f_\infty}{f_0}\right). \tag{8}$$

We can easily observe that the carrying capacity does not heavily depend on the single patient, but it is related to the kind of tumor. Precisely, in our case, we use as reference value $f_\infty = 200\ ng/mL$. To fix it, we use the information of the Eureka1 database. In fact, the maximum PSA value among 3538 patients is $196\ ng/mL$.

The proposed model is validated by fitting (7) with clinical data. In particular, we reconstruct the PSA curve via RBF interpolation and then we estimate the growth rate α by means of an optimization tool.

4 Fitting procedures and results

We remark that our aim consists in finding the *best* values of α for which (7) accurately fits the PSA data. Therefore, the objective function, which is minimized with a stochastic technique (namely Particle Swarm Optimization (PSO) [15, 16, 17, 21]) is

$$\sum_{i=1}^{N}\{f_i - f_0 \exp\left[\log\left(f_\infty/f_0\right)\left(\exp(-\alpha x_i) - 1\right)\right]\}^2, \tag{9}$$

where (x_i, f_i), $i = 1, \ldots, N$, are the real data. In our case we have really few data ($4 \leq N \leq 15$). Thus, in order to obtain reliable estimations via the PSO, we first reconstruct the curve defined by such few samples via RBF interpolation. In doing so, we use the Matérn C^4 kernel, defined as

$$\phi(\boldsymbol{x}, \boldsymbol{y}) = \exp\left(-||\boldsymbol{x} - \boldsymbol{y}||_2\right)(3 + 3||\boldsymbol{x} - \boldsymbol{y}||_2 + ||\boldsymbol{x} - \boldsymbol{y}||_2^2). \tag{10}$$

The proposed method has been tested with a group of prostatectomized patients who relapsed at different times. For instance, in Figure 1 we report four different patients. Specifically, after approximating the PSA curve (see the blue line in Figure 1), we estimate the growth parameter of the tumor by minimizing (9) with respect to α. The resulting Gompertzian function is plotted in red in Figure 1.

We also report in Table 1 the Root Mean Square Errors (RMSEs) of the fitted curves

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^{N} \left|f(x_i) - f_i\right|^2},$$

where (x_i, f_i), $i = 1, \ldots, N$, are the real data.

We expect a correlation between the value of the parameter α and the velocity of the tumor growth. Let us consider Figure 1: the estimated growth rates α are $1.45E-02$, $1.29E-02$, $7.51E-03$ and $9.40E-03$ for the case (a),(b),(c) and (d), respectively. This is consistent with our expectations. In fact, the two patients plotted at the top of Figure 1 are characterized by a faster growth and a larger value of α than the ones plotted at the bottom.

Therefore, we numerically observe that the Gompertzian function is a reliable choice for modeling the tumor dynamics and moreover it enables us to assess the risk of the disease by estimating α. This is also confirmed by the results reported in Table 1. In particular, we can note that the approximation turns out to be accurate. The worst error corresponds to the data set that shows a truly low *regularity* of samples (see Figure 1, case (b)).

	case (a)	case (b)	case (c)	case (d)
RMSE	$2.36E-02$	$3.51E-01$	$4.77E-02$	$4.18E-02$

Table 1: RMSEs obtained by using the Matérn C^4 as RBF interpolant for the RBF-PSO algorithm.

5 Conclusions and work in progress

The proposed technique, based on combining an optimization method with RBF interpolation, allows to effectively simulate the cancer dynamics by estimating its growth rate. The accuracy of the RBF fit is ensured by properly selecting a safe shape parameter. In particular, it is found out by minimizing the power function (see Section 2). Anyway, since the second component of the error bound in (6) also depends on the basis function via the native space norm, this is not an optimal approach. Therefore, work in progress consists in considering other strategies, such as the Leave One Out Cross Validation (LOOCV) scheme [10, 18]. In this way we could obtain an optimal shape parameter for each patient.

Finally, as evident from numerical simulations, the positivity of PSA samples is not always preserved throughout the interpolation process (see Figure 1). Thus, we need to consider techniques enabling us to fit positive data values with positive approximants [25, 28].

6 Acknowledgements

References

[1] G. Allasia, R. Cavoretto, A. De Rossi. Local interpolation schemes for Landmark-based image registration: A comparison. *Math. Comput. Simul.* **106**:1–25, 2014.

[2] M.D. Buhmann. Radial Basis Functions: Theory and Implementation. Cambridge Monogr. Appl. Comput. Math., vol. 12, Cambridge Univ. Press, Cambridge, 2003.
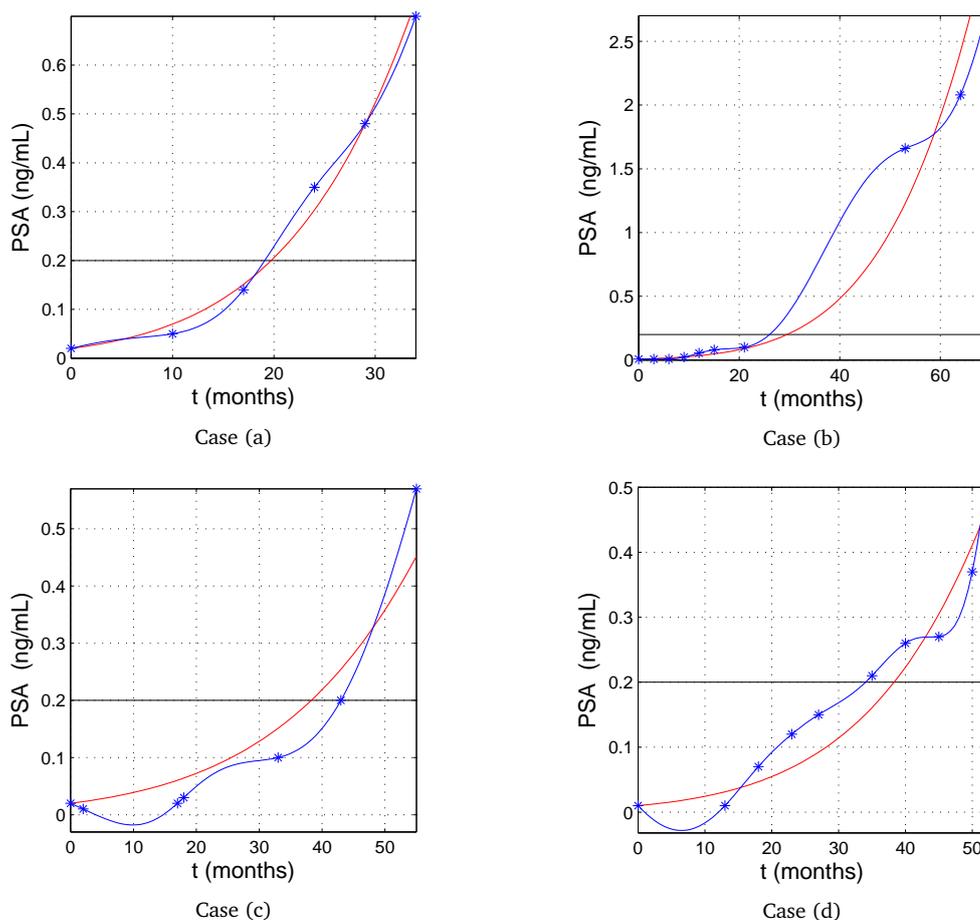
Figure 1: The predicted Gompertzian growth for four different patients. The blue and red lines indicate the reconstructed RBF interpolant and the approximated Gompertz, respectively; the blue dots represent the PSA samples and the black line the relapse threshold.

[3] J.C. Carr, W.R. Fright, R.K. Beatson. Surface interpolation with radial basis functions for medical imaging. *IEEE Trans. Med. Imaging* **16**:96–107, 1997.

[4] P. Castorina, P.P. Delsanto, C. Guiot. A classification scheme for phenomenological universalities in growth problems in physics and other sciences. *Phys. Rev. Lett.* **97**:1–4, 1996.

[5] S. De Marchi, G. Santin. Fast computation of orthonormal basis for RBF spaces through Krylov space methods. *BIT* **55**:949–966, 2015.

[6] G. Dimonte, E.J. Bergstralh, M. Bolander, R.J. Karnes, D.J. Tindall. Use of tumor dynamics to clarify the observed variability among biochemical recurrence nomograms for prostate cancer. *Prostate* **72**:280–290, 2012.

[7] T.A. Driscoll, B. Fornberg. Interpolation in the limit of increasingly flat radial basis functions. *Comput. Math. Appl.* **43**:413–422, 2002.

[8] G.E. Fasshauer. Positive definite kernels: Past, present and future. *Dolomites Res. Notes Approx.* **4**:21–63, 2011.

[9] G.E. Fasshauer. Meshfree Approximations Methods with MATLAB. World Scientific, Singapore, 2007.

[10] G.E. Fasshauer, M.J. McCourt. Kernel-based Approximation Methods using MATLAB. World Scientific, Singapore, 2015.

[11] B. Fornberg, E. Larsson, N. Flyer. Stable computations with Gaussian radial basis functions. *SIAM J. Sci. Comput.* **33**:869–892, 2011.

[12] D. Gabriele, F. Porpiglia, G. Muto, P. Gontero, C. Terrone, S. Annoscia, D. Randone, S. Benvenuti, G. Arena, I. Stura, C. Guiot. Eureka-1 database: An epidemiogical analysis. *Minerva Urol. Nefrol.* **1**:9–15, 2015.

[13] A.S. Gliozzi, C. Guiot, P.P. Delsanto, D.A. Iordache. A novel approach to the analysis of human growth. *Theor. Biol. Med. Model* **9**:1–16, 2012.

[14] C. Guiot, P.G. Degiorgis, P.P. Delsanto, P. Gabriele, T.S. Deisboeck. Does tumor growth follow a 'universal law'?. *J. Theor. Biol.* **225**:147–151, 2003.

[15] J. Kennedy, R.C. Eberhart. Particle swarm optimization. In: Proc. of 1995 IEEE Int. Conf. Neural Networks, IEEE, Piscataway, NJ, 1995, pp. 1942–1948.

[16] K.E. Parsopoulos, M.N. Vrahatis. Particle swarm optimization method for constrained optimization problems. *Frontiers Artificial Intelligence Appl.* **76**:214–220, 2002.

[17] S.N. Qasem, S.M. Shamsuddin. Radial basis function network based on time variant multi-objective particle swarm optimization for medical diseases diagnosis. *Appl. Soft. Comput.* **11**:1427–1438, 2011.

[18] S. Rippa. An algorithm for selecting a good value for the parameter c in radial basis function interpolation. *Adv. Comput. Math.* **11**:193–210, 1999.

[19] A. Safdari-Vaighani, A. Heryudono, E. Larsson. A radial basis function partition of unity collocation method for convection-diffusion equations arising in financial applications. *J. Sci. Comput.* **64**:341–367, 2015.

[20] V. Shcherbakov, E. Larsson. Radial basis function partition of unity methods for pricing vanilla basket options. *Comput. Math. Appl.* **71**:185–200, 2016.

[21] Y. Shi, R.C. Eberhart. A modified particle swarm optimizer. In: Proc. of 1998 IEEE Int. Conf. on Evolutionary Computation, IEEE, Anchorage, AK, USA, 1998, pp. 69–73.

[22] I. Stura, E. Venturino, C. Guiot. A two-clones tumor model: Spontaneous growth and response to treatment. *Math. Biosci.* **10**:19–28, 2015.

[23] H. Wendland. Scattered Data Approximation. Cambridge Monogr. Appl. Comput. Math., vol. 17, Cambridge Univ. Press, Cambridge, 2005.

[24] H. Wendland. Surface reconstruction from unorganized points.
`http://people.maths.ox.ac.uk/wendland/research/old/reconhtml/reconhtml.html`, 2002.

[25] J. Wu, Y. Lai, X. Zhang. Radial basis functions for shape preserving planar interpolating curves. *J. Comput. Inf. Sci.* **7**:1453–1458, 2010.

[26] Z. Wu, R. Schaback. Local error estimates for radial basis function interpolation of scattered data. *IMA J. Numer. Anal.* **13**:13–27, 1993.

[27] D. Wu, K. Warwick, Z. Ma, M.N. Gasson, J.G. Burgess, S. Pan, T.Z. Aziz. Prediction of parkinson's disease tremor onset using a radial basis function neural network based on particle swarm optimization. *Int. J. Neur. Syst.* **20**:109–116, 2010.

[28] J. Wu, X. Zhang, L. Peng. Positive approximation and interpolation using compactly supported radial basis functions. *Math. Probl. Eng.* **2010**:1–10, 2010.